การพยากรณ์ปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือ โดยการใช้เทคนิคการเรียนรู้ของเครื่อง

Forecasting sugarcane production in the Northeastern region of Thailand Using machine learning

กนกวรรณ ประยูรชาญ[1] , พงศธร ตันตระบัณฑิตย์[2] และอนุพงศ์ สุขประเสริฐ[3*]

[1,2,3*]คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

Kanokvan Prayunchan[1], Pongsatorn Tantrabundit[2] and Anupong Sukprasert[3*]

[1,2,3*]Mahasarakham Business School, Mahasarakham University

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบและทดสอบประสิทธิภาพของตัวแบบสำหรับพยากรณ์ปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือ โดยใช้ข้อมูลระหว่างปี พ.ศ. 2561 ถึงปี พ.ศ. 2565 ซึ่งเป็นปัจจัยที่ส่งผลต่อปริมาณผลผลิตอ้อย จำนวน 9 ปัจจัย มาวิเคราะห์ตามกระบวนวิเคราะห์ข้อมูล ด้วยเทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค ได้แก่ เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคการถดถอยเชิงเส้น (Linear Regression) และเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ผลการวิจัยพบว่า เทคนิคที่ให้ค่าประสิทธิภาพที่ดีที่สุดคือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) รองลงมาคือเทคนิคการถดถอยเชิงเส้น (Linear Regression), และเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) โดยมีค่าสัมประสิทธิ์การตัดสินใจพหุคูณ เท่ากับ 93.5 %, 93.10% และ 46.50% ตามลำดับ ดังนั้นเทคนิคเทคนิคโครงข่ายประสาทเทียมจึงมีความเหมาะสมที่จะนำไปสร้างตัวแบบสำหรับการปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือต่อไป

คำสำคัญ : 1) การพยากรณ์ 2) ผลผลิตอ้อย 3) โครงข่ายปราสาทเทียม 4) สมการถดถอยเชิงเส้น 5) เพื่อนบ้านใกล้ที่สุด

Abstract

The purpose of this study is to construct and assess the performance of machine learning models for estimating sugarcane output quantities in Thailand's northeastern region, using data from 2018 to 2022. These are the nine influencing elements on sugarcane production that were examined utilizing the Data Analytics Process, Four machine learning techniques Neural Network, Linear Regression, and K-Nearest Neighbor are used in this study. The study findings indicate that Neural Network outperforms Linear Regression and K-Nearest Neighbor, achieving coefficients of determination of 93.5%, 93.10%, and 46.50% respectively. Consequently, the Neural Network approach is deemed suitable for constructing models to forecast sugarcane production levels in the Northeastern region in the future.

Keywords: 1) Forecasting 2) Sugarcane 3) Neural network 4) Linear regression 5) K-Nearest Neighbor

[1]นิสิตปริญญาโท หลักสูตรบัญชีมหาบัณฑิต (Graduate Student, Master of Accountancy Program)

[2,3*] อาจารย์ประจำคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม E-mail: anupong.s@acc.msu.ac.th

บทนำ

อุตสาหกรรมอ้อยและน้ำตาลทรายมีความสำคัญอย่างมากต่อระบบเศรษฐกิจไทยเนื่องจากมีมูลค่าการส่งออกสูงและประเทศไทยเป็นผู้ส่งออกอันดับสองของโลก รองจากประเทศบราซิลที่เป็นผู้ผลิตและส่งออกอันดับหนึ่งของอาเซียนและของเอเชีย โดยจากตัวเลขด้านการบริโภคพบว่า มีการบริโภคน้ำตาลทรายในอาเซียนปริมาณ 14.34 ล้านตัน และมีการส่งออกน้ำตาลทรายทั้งหมด 10.23 ล้านตันในปี พ.ศ.2561 เป็นการส่งออกน้ำตาลทรายของไทยถึงร้อยละ 85.92 โดยตลาดนำเข้าน้ำตาลที่สำคัญในอาเซียนได้แก่ อินโดนีเซีย มาเลเซีย และสิงคโปร์ ทั้งนี้เพราะอินโดนีเซียมีประชากรมากและมีการผลิตที่ไม่เพียงพอต่อการบริโภคภายในประเทศ ส่วนมาเลเซียและสิงคโปร์เป็นผู้นำเข้าเพราะขาดแหล่งเพาะปลูกอ้อยในประเทศ ภัคจิรา ทวีกาญจน์ (2562) อ้อยเป็นวัตถุดิบในการผลิตน้ำตาล อีกทั้งเป็นพืชเศรษฐกิจที่สำคัญชนิดหนึ่งของประเทศ สำนักงานคณะกรรมการอ้อยและน้ำตาลทรายได้กำหนดนโยบายในการสนับสนุนการผลิตเพื่อเพิ่มปริมาณการส่งออกในอุตสาหกรรมอ้อยและน้ำตาลทรายของไทย เป็นการเพิ่มระดับการแข่งขันในตลาดโลก โรงงานน้ำตาลต่างๆ จึงพยายามหาแหล่งอ้อยเพื่อผลิตน้ำตาลทรายในปริมาณที่ตลาดต้องการหรือตามนโยบาย อย่างไรก็ตามปริมาณอ้อยที่เกษตรกรสามารถผลิตเพื่อใช้ในโรงงานมีความผันผวนเนื่องจากปัจจัยต่างๆ เช่น สภาพอากาศ, ปริมาณน้ำ, ศัตรูพืช, พื้นที่เพาะปลูกที่ไม่แน่นอนในแต่ละปี และสภาวะอื่นๆ ที่มีผลกระทบต่อผลผลิตอ้อย ณัฐนนท์ ลีลาตระกูล และคณะ (2559) นอกจากนี้ความแห้งแล้ง ความชื้นในอากาศ ความชื้นในดิน ฝนทิ้งช่วงเป็นเวลานาน กระจายไปทั่วพื้นที่ของประเทศไทย ศรุดา สมพอง และคณะ (2564) โดยเฉพาะภาคตะวันออกเฉียงเหนือที่เป็นพื้นที่ศึกษา ปัจจัยเหล่านี้ส่งผลให้ผลผลิตอ้อยมีความผันผวน โรงงานต้องเผชิญกับความเสี่ยงในการจัดหาวัตถุดิบให้เพียงพอสำหรับการผลิตน้ำตาล และเพียงพอกับความต้องการของตลาด

การพยากรณ์ปริมาณอ้อยมีความสำคัญในการวางแผนการผลิตทั้งในระบบในประเทศและสำหรับการส่งออกในอุตสาหกรรมอ้อยและน้ำตาลทรายของไทย ข้อมูลการพยากรณ์นี้ถูกนำมาใช้ในการบริหารจัดการกระบวนการการเก็บเกี่ยวอ้อยที่โรงงานน้ำตาลเพื่อเตรียมความพร้อมและควบคุมกระบวนการการเก็บเกี่ยวในแต่ละปีให้มีปัญหาน้อยที่สุด การพยากรณ์ปริมาณอ้อยช่วยในการวางแผนการผลิตอ้อยให้มีประสิทธิภาพมากขึ้น รวมถึงการจัดการทรัพยากรและควบคุมค่าใช้จ่ายที่เกิดขึ้นในกระบวนการผลิตน้ำตาล ข้อมูลปริมาณอ้อยที่คาดการณ์จากการเก็บเกี่ยวมีผลต่อการวางแผนการผลิตและการลงทุนในอุตสาหกรรม นอกจากนี้การพยากรณ์ปริมาณอ้อยยังช่วยให้เกษตรกรวางแผนการเก็บเกี่ยวผลผลิตในอนาคตรวมถึงการจัดการทรัพยากรที่มีอยู่เพื่อให้สามารถปรับตัวตามตลาดและสถานการณ์ที่เกิดขึ้นในอนาคต

ภาคตะวันออกเฉียงเหนือของไทยเป็นพื้นที่ปลูกอ้อยที่สำคัญเนื่องจากมีจำนวนพื้นที่ปลูกอ้อยมากเป็นอันดับ 1 ของประเทศ ในปี พ.ศ. 2565 มีพื้นที่ปลูกอ้อย 4,950,053 ไร่ เพิ่มขึ้นจากปี พ.ศ. 2564 จำนวน 288,258 ไร่ หรือคิดเป็นร้อยละ 6.18 และในขณะเดียวกันภาคตะวันออกเฉียงเหนือยังเป็นพื้นที่ที่มีโรงงานน้ำตาลมากเป็นอันดับหนึ่งของไทย คือ 21 แห่ง สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย (2566) ดังนั้นผู้วิจัยจึงสนใจศึกษาการพยากรณ์ปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือเพื่อสร้างตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อย เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ และเกษตรกรผู้ปลูกอ้อยและกลุ่มอุตสาหกรรมอ้อยและน้ำตาลนำข้อมูลการพยากรณ์ไปวางแผนจัดการผลผลิตในอนาคตต่อไป

การทบทวนวรรณกรรม

โครงข่ายประสาทเทียม (Artificial Neural Network) (Patterson Dan W, 1930) เป็นโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทคล้ายกับการทำงานในสมองมนุษย์ แบ่งเป็น 4 แบบ คือ โครงข่ายการส่งข้อมูลแบบไม่ย้อนกลับ (Feed Forward Network) โครงข่ายการส่งข้อมูลแบบย้อนกลับ (Feedback Network) โครงข่ายประสาทเทียมอย่างง่าย (Perceptron) และ โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back

Propagation Algorithm) โดยงานวิจัยนี้ได้เลือกใช้อัลกอริทึมของ โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับในการสร้าง แบบจำลอง เนื่องจากข้อมูลที่ใช้สำหรับเป็นปัจจัยนำเข้าเป็นข้อมูลย้อนหลังที่ต้องอาศัยการเรียนรู้หลาย ๆ รอบเพื่อให้ได้ข้อมูลผลลัพธ์ที่มีความน่าเชื่อถือ สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับประกอบด้วยขั้นของข้อมูลเบื้องต้น 3 ชั้นได้แก่ 1.ชั้นอินพุต (Input layer) 2. ชั้นซ่อน (hidden layer) และ 3 ชั้นเอาท์พุต (output layer)

สมการถดถอยเชิงเส้น (Linear Regression) เป็นกระบวนการทางสถิติที่ใช้ในการวิเคราะห์และประมาณค่าความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variables) และตัวแปรตาม (Dependent Variable) โดยทำการหาสมการเชิงเส้นที่เหมาะสมที่สุดสำหรับข้อมูลที่มีอยู่กระบวนการนี้สามารถเรียงลำดับได้ด้วยขั้นตอนต่าง ๆ ดังนี้ : 1. เลือกข้อมูล 2. กำหนดโมเดล 3. ปรับค่าความชันและตัดแกน y 4. การประมาณค่า 5. การประเมินความแม่นยำ การประมาณค่าด้วย Linear Regression เป็นวิธีที่มีความง่ายและที่นิยมใช้ในการทำนายผลลัพธ์หรือประมาณค่าในหลายสาขาวิชา เช่น วิทยาการคอมพิวเตอร์ วิทยาศาสตร์ข้อมูล วิศวกรรม การเงิน และอื่นๆ

เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) เป็นเทคนิคการเรียนรู้เครื่องจักรที่ใช้ในการจัดหมวดหมู่หรือทำนายข้อมูลโดยการค้นหาตัวอย่างที่อยู่ใกล้ที่สุดจากชุดข้อมูลอบรมของข้อมูลที่ต้องการจัดหมวดหมู่หรือทำนาย การค้นหา K ตัวอย่างที่อยู่ใกล้ที่สุดจากข้อมูลอบรม เมื่อต้องการจัดหมวดหมู่หรือทำนายข้อมูลใหม่ K คือจำนวนของตัวอย่างที่จะถูกใช้ในการตัดสินใจ ข้อมูลจัดหมวดหมู่หรือทำนายโดยการคำนวณค่าเฉลี่ยหรือค่าความห่างจาก K ตัวอย่างใกล้ที่สุดเพื่อหาค่าที่คาดหวังหรือค่ามากที่เหมาะสม สามารถใช้ในหลายสาขาเช่น การจัดหมวดหมู่ข้อมูล, การแบ่งกลุ่ม, การคำนวณค่าคาดหวัง, และอื่น ๆ ค่า K สำคัญในการเลือกและกำหนดคุณสมบัติของแบบจำลอง K-NN และอาจมีผลต่อประสิทธิภาพและการทำนายของระบบ เทคนิค K-NN สามารถใช้ในการแก้ปัญหาการจัดหมวดหมู่, การคาดคะเน, การคำนวณค่าทางสถิติ, การแก้ปัญหาความคลาดเคลื่อนข้อมูล, และหลายงานอื่น ๆ

เอกวิจัย เมยไธสง และคณะ (2565) ศึกษาการพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษา สาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง ได้แก่ โครงข่ายประสาทเทียม การเรียนรู้แบบต้นไม้ตัดสินใจ การค้นหาเพื่อนบ้านใกล้ที่สุด การเรียนรู้เบย์อย่างง่าย และการสุ่มป่าไม้ พบว่า วิธีโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นให้ค่าความถูกต้องสูงที่สุด ศรีรักษ์ ศรีทองชัย (2566) การเปรียบเทียบวิธีการประมาณค่าสูญหายในการพยากรณ์ความเข้มข้นของ PM2.5 ด้วยโครงข่ายประสาทเทียม LSTM ผลการศึกษาพบว่าวิธี วิธีเพื่อนบ้านใกล้สุด (KNN) และ วิธีการแทนค่าแบบพหุด้วยสมการลูกโซ่ (MICE) ให้ค่าความผิดพลาดน้อยที่สุดจึงเป็นวิธีการประมาณค่าสูญหายที่ดีที่สุด จตุภัทร เมฆพายัพ และกิดาการ สายธนู (2560) การพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกของประเทศไทยด้วยเทคนิคข่ายงานระบบประสาทเทียม (ANN) ผลการศึกษาแสดงให้เห็นว่าตัวแบบข่ายงานระบบประสาทเทียมที่มีสมรรถนะดีที่สุดในการพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกของประเทศไทยคือตัวแบบข่ายงานระบบประสาทเทียม วรินทรา แซ่โล้ และคณะ (2566) พบว่าการใช้แบบจำลองโครงข่ายประสาทเทียมร่วมกับการประยุกต์ค่าเวลาการเกิดน้ำท่าสูงสุด (TC) แบบจำลองนี้สามารถพยากรณ์อัตราการไหลล่วงหน้าได้ทั้งค่าปริมาณอัตราการไหลและรูปร่างของกราฟน้ำหลาก ค่าประสิทธิภาพของแบบจำลองมีค่าลดลงตามลำดับเมื่อชั่วโมงการพยากรณ์เพิ่มมากขึ้นในระยะเวลาที่ย้อนหลัง ธนะรัตน์ รัตนกูล และกันต์ธมน สุขกระจ่าง (2565) ศึกษาการทำนายปริมาณการใช้เชื้อเพลิงขนส่งด้วยโครงข่ายประสาทเทียมกรณีศึกษา รถบรรทุกดินถมในพื้นที่จังหวัดชายแดนใต้ประเทศไทย พบว่า โครงข่ายประสาทเทียมให้ค่าความคลาดเคลื่อนในการพยากรณ์ต่ำที่สุด จากผลการทดลองนี้สามารถไปใช้ไปเป็นข้อเสนอแนะในให้บริษัทกรณีศึกษาใช้ในการวางแผนใช้เชื้อเพลิงอย่างมีประสิทธิภาพ พงศ์ หลวงมูล และถาวร อ่อนประไพ (2563) ศึกษาแบบจำลองในรูปแบบสมการถดถอยเชิงเส้นอย่างง่ายและพหุคูณเพื่อประมาณการผลผลิตข้าวในแปลงทดลองร่วมกับค่าผลผลิตข้าวที่ได้จากการตั้งแปลงเก็บเกี่ยวผลการศึกษา พบว่าแบบจำลองที่น่าเชื่อถือมากที่สุด

จากงานวิจัยที่เกี่ยวข้องข้างต้นพบว่าเทคนิคการเรียนรู้ของเครื่องได้ถูกนำมาประยุกต์ใช้กันอย่างแพร่หลาย ล้วนเป็นเทคนิคที่มีประสิทธิภาพสูง ผู้วิจัยจึงได้สนใจศึกษาเทคนิคการเรียนรู้ของเครื่องดังกล่าวในการพยากรณ์ปริมาณผลผลิตอ้อย โดยทำการเก็บข้อมูลปัจจัยต่างๆ จากแหล่งข้อมูลตั้งแต่ปี พ.ศ. 2561 - 2565 และนำข้อมูลปัจจัย มาวิเคราะห์เปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ เพื่อการบริหารจัดการผลผลิตอ้อยในฤดูเก็บเกี่ยว เป็นการเตรียมความพร้อมและควบคุมกระบวนการผลิตน้ำตาลในแต่ละปีให้มีปัญหาน้อยที่สุด

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ
2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ

ขอบเขตของการวิจัย

การวิจัยครั้งนี้เป็นการพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกเฉียงเหนือของประเทศไทย โดยเก็บรวบรวมข้อมูลตัวแปรแต่ละตัวของจังหวัดที่ปลูกอ้อยในภาคตะวันออกเฉียงเหนือ 20 จังหวัด ตั้งแต่ปี พ.ศ. 2561 ถึงปี พ.ศ.2565 โดยทำการเก็บรวบรวมจาก 3 แหล่งข้อมูล ดังนี้

แหล่งที่ 1: ข้อมูลพื้นที่ปลูกอ้อย (ไร่) ผลผลิตอ้อยเฉลี่ยต่อไร่ (ตัน) ในบริเวณภาคตะวันออกเฉียงเหนือ เก็บข้อมูล ปี พ.ศ. 2561-2565 จากสำนักงานคณะกรรมการอ้อยและน้ำตาลกระทรวงอุตสาหกรรม

แหล่งที่ 2: ข้อมูลปริมาณน้ำฝนเฉลี่ย (มิลลิเมตร) อุณหภูมิสูงสุดเฉลี่ย (°c) โดยรวมในแต่ละพื้นที่จังหวัดที่มีการปลูกอ้อยในภาคตะวันออกเฉียงเหนือ เก็บข้อมูล ปี พ.ศ. 2561-2565 ซึ่งรวบรวมโดยสำนักงานสถิติแห่งชาติ จากกรมอุตุนิยมวิทยา

แหล่งที่ 3: ข้อมูลราคาอ้อยเฉลี่ย, ราคามันสำปะหลังเฉลี่ย, ราคาข้าวโพดเลี้ยงสัตว์เฉลี่ย (บาท/ตัน) ซึ่งเป็นข้อมูลปี พ.ศ. 2561-2565 เก็บรวบรวมโดยสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์

วิธีดำเนินการ

ผู้วิจัยได้วิเคราะห์ข้อมูลตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) โดยแบ่งออกเป็น 6 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การทำความเข้าใจปัญหา (Research Understanding) ปริมาณอ้อยที่เกษตรกรสามารถผลิตเพื่อใช้ในโรงงานไม่แน่นอนเนื่องจากมีปัจจัยหลายราย เช่น สภาพอากาศ, ปริมาณน้ำ, ศัตรูพืช, พื้นที่เพาะปลูกที่ไม่แน่นอนในแต่ละปี และสภาวะอื่น ๆ ที่มีผลกระทบต่อผลผลิตอ้อย ปัจจัยเหล่านี้ทำให้โรงงานเผชิญกับความเสี่ยงในการจัดหาวัตถุดิบสำหรับการผลิตน้ำตาล

ดังนั้น ผู้วิจัยจึงได้ศึกษาเรื่องการพยากรณ์ปริมาณอ้อยมีเพื่อนำข้อมูลมาใช้ในการวางแผนการผลิตทั้งในระบบในประเทศและสำหรับการส่งออกในอุตสาหกรรมอ้อยและน้ำตาลทรายของไทย ข้อมูลการพยากรณ์นี้ถูกนำมาใช้ในการบริหารจัดการกระบวนการการเก็บเกี่ยวอ้อยที่โรงงานน้ำตาลเพื่อเตรียมความพร้อมและควบคุมกระบวนการการเก็บเกี่ยวในแต่ละปีให้มีปัญหาน้อยที่สุด รวมถึงการจัดการทรัพยากรและควบคุมค่าใช้จ่ายที่เกิดขึ้นในกระบวนการผลิตน้ำตาล ข้อมูลปริมาณอ้อยที่คาดการณ์จะเก็บเกี่ยวมีผลต่อการวางแผนการผลิตและการลงทุนในอุตสาหกรรม นอกจากนี้การพยากรณ์ปริมาณอ้อยยังช่วยให้เกษตรกรวางแผนการเก็บเกี่ยวผลผลิตในอนาคต การจัดการทรัพยากรสำคัญเพื่อให้สามารถปรับตัวตามตลาดและสถานการณ์ที่เกิดขึ้นในอนาคตอย่างมีประสิทธิภาพ

ขั้นตอนที่ 2 การทำความเข้าใจเข้าใจข้อมูล (Data Understanding) ข้อมูลที่นำมาใช้สำหรับการสร้างตัวแบบในการพยากรณ์ริมาณผลผลิตอ้อยของภาคะวันออกเฉียงเหนือ ในการศึกษาครั้งนี้ผู้วิจัยใช้ข้อมูลระหว่างปี

พ.ศ. 2561 ถึงปี พ.ศ. 2565 เป็นข้อมูล จังหวัดที่ปลูกอ้อยในเขตภาคตะวันออกเฉียงเหนือ, ปีการผลิต, พื้นที่ปลูกอ้อย, ผลผลิตอ้อยเฉลี่ยต่อไร่ สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย (2566) ปริมาณน้ำฝนเฉลี่ย, ข้อมูลอุณหภูมิเฉลี่ย กรมอุตุนิยมวิทยา (2566) ราคาอ้อย,มันสำปะหลัง,ข้าวโพดเลี้ยงสัตว์ เฉลี่ย/ตัน สำนักงานเศรษฐกิจการเกษตร (2566) ที่ถูกจัดเก็บที่ไว้ใน ไฟล์ Excel รวมทั้งสิ้น 100 รายการ 9 แอตทริบิวต์ ดังตารางที่ 1

ตารางที่ 1 แสดงรายละเอียดแอตทริบิวต์ทั้งหมดของข้อมูลในระบบฐานข้อมูล

ลำดับที่	แอตทริบิวต์	คำอธิบาย	การเข้ารหัส
1	Prov.	จังหวัดภาคตะวันออกเฉียงเหนือ	1.Udon Thani
		ที่ปลูกอ้อย	2.Khon Kaen
		ปีการผลิต	3,4,5.....
2	Year	ปริมาณน้ำฝนเฉลี่ย	ข้อมูลที่เกิดขึ้นจริง
3	Rain	อุณหภูมิเฉลี่ย	ข้อมูลที่เกิดขึ้นจริง
4	Temp	พื้นที่ปลูกอ้อย	ข้อมูลที่เกิดขึ้นจริง
5	Area	ผลิตอ้อย/ไร่	ข้อมูลที่เกิดขึ้นจริง
6	A_Cane	ราคาอ้อย	ข้อมูลที่เกิดขึ้นจริง
7	P_Cane	ราคามันสำปะหลัง	ข้อมูลที่เกิดขึ้นจริง
8	P_Cassava	ราคาข้าวโพดเลี้ยงสัตว์	ข้อมูลที่เกิดขึ้นจริง
9	P_Corn		ข้อมูลที่เกิดขึ้นจริง

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation) ตัวแปรที่นำมาวิจัยครั้งนี้ประยุกต์จากการรวบรวมปัจจัยที่มีผลต่อปริมาณอ้อยโดยการรวบรวมจากงานวิจัยที่เคยมีการศึกษาได้แก่ วิภาพร (2552); Alvarez et al. (1982) และ สุรเชษฐ (2552) ผู้วิจัยได้คัดเลือกปัจจัยจำนวน 9 ปัจจัยได้แก่ จังหวัดที่ปลูกอ้อยในเขตภาคตะวันออกเอียงเหนือซึ่งเป็นพื้นที่เป้าหมาย ปีการผลิต พื้นที่ปลูกอ้อย ผลผลิตอ้อยเฉลี่ยต่อไร่ ปริมาณน้ำฝนเฉลี่ยข้อมูล อุณหภูมิเฉลี่ย, ราคาอ้อย, มันสำปะหลัง, ข้าวโพดเลี้ยงสัตว์ เฉลี่ย/ตัน ปัจจัยที่คัดเลือกมาได้รวบรวมชุดข้อมูลการเรียนรู้เป็นข้อมูลรายปี จำนวน 5 ปีย้อนหลัง นับตั้งแต่ ปี 2561-2565

3.1. การนำเข้าข้อมูลที่ใช้สำหรับวิเคราะห์ข้อมูล เพื่อสร้างตัวแบบในการพยากรณ์งานวิจัยนี้ ผู้วิจัยได้ทำการเก็บรวบรวมให้อยู่ในรูปแบบของไฟล์ Excel และนำข้อมูลมาทำการประมวลผลในโปรแกรม RapidMiner Studio แสดงในตารางที่ 1

3.2 การแปลงข้อมูล เนื่องจากข้อมูลที่ได้มายังไม่เหมาะสมสำหรับบางเทคนิคที่ใช้สำหรับการสร้างตัวแบบดังนั้นผู้วิจัยจึงปรับช่วงค่าของข้อมูล เนื่องจากข้อมูลปัจจัยต่าง ๆ ที่นำมาใช้สำหรับเป็นข้อมูล มีความแตกต่างกันในเรื่องของชนิดข้อมูล หน่วยนับ และช่วงของข้อมูล ก่อนมีการนำข้อมูลเข้าสู่การสร้างแบบจำลองจึงมีการปรับช่วงค่าของข้อมูลให้มีรูปแบบและค่าใกล้เคียงกันโดยให้มีค่าอยู่ระหว่าง 0 – 1 เพื่อให้อยู่ในมาตรฐานเดียวกัน ซึ่งสมการที่ใช้สามารถแสดงได้ดังสมการ Begg and Palaniswami (2006) ตัวอย่างข้อมูลจริงกับข้อมูลที่มีการปรับค่าสามารถแสดงได้ในตารางที่ 2

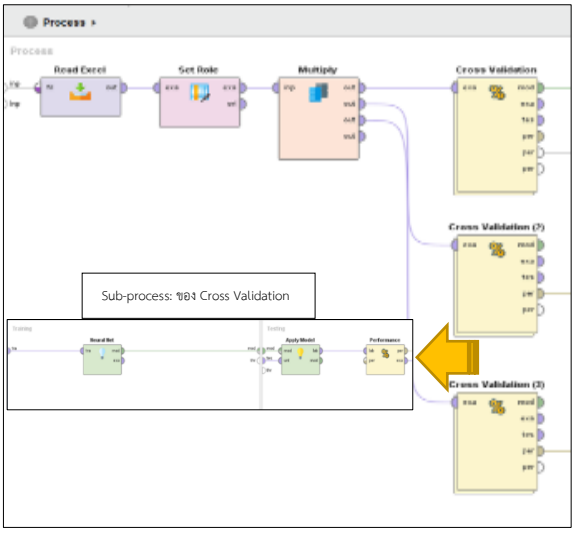
$$\text{output} = \frac{x - \min}{\max - \min}$$

โดยที่ x คือ ข้อมูลที่นำเข้ามาคำนวณ
min คือ ข้อมูลที่น้อยที่สุดของชุดข้อมูลนั้นๆ
max คือ ข้อมูลที่มากที่สุดของชุดข้อมูลนั้นๆ

ตารางที่ 2 แสดงตัวอย่างข้อมูลที่มีการปรับช่วงค่า

Prov.	Year	Rain	Temp	Area	A_Cane	P_Cane	P_Cassava	P_Corn
1	1	0.40	0.33	0.00	0.90	0.24	0.81	0.12
2	1	0.24	0.33	0.92	0.78	0.24	0.81	0.12
3	1	0.21	0.63	0.89	0.86	0.24	0.81	0.12
4	1	0.18	0.56	0.82	0.88	0.24	0.81	0.12
5	1	0.31	0.22	0.61	1.00	0.24	0.81	0.12
6	1	0.32	0.33	0.46	0.89	0.24	0.81	0.12
7	1	0.22	-	0.42	0.93	0.24	0.81	0.12
8	1	0.66	0.26	0.30	0.98	0.24	0.81	0.12

ขั้นตอนที่ 4 การสร้างแบบจำลอง (Modeling Phase) ขั้นตอนนี้เป็นขั้นตอนการสร้างตัวแบบสำหรับการพยากรณ์โดยใช้เทคนิคการจำแนกประเภทข้อมูลได้แก่ เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) ด้วยโปรแกรม RapidMiner Studio Version 10 โดยผู้วิจัยได้ทำการแบ่งข้อมูลสำหรับการสร้างตัวแบบ (Training Set) และข้อมูลที่ใช้สำหรับทดสอบประสิทธิภาพ (Testing Set) ด้วยวิธี Cross Validation ซึ่งมีวิธีการดังภาพที่ 1



ภาพที่ 1 ขั้นตอนการสร้างตัวแบบสำหรับการพยากรณ์โดยใช้โปรแกรม RapidMiner Studio

ขั้นตอนที่ 5 การประเมินผล (Evaluation Phase) ขั้นตอนนี้เป็นการวิเคราะห์ข้อมูลทดสอบแบบจำลองโดยชุดข้อมูลทดสอบ (Testing Data) โดยวัดประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) จากนั้นทำการเปรียบเทียบประสิทธิภาพในแต่ละอัลกอริทึม เพื่อให้ได้แบบจำลองจากอัลกอริทึมที่ดีที่สุด เพราะก่อนนำผลลัพธ์ที่ได้ไปใช้งานต่อต้องมีการวัดประสิทธิ์ภาพของผลลัพธ์ว่าตรงกับวัตถุประสงค์ที่ตั้งไว้ตั้งแต่ข้อแรกหรือไม่ มีความน่าเชื่อถือมากน้อยเพียงใด ในที่นี้จะใช้เทคนิค 10-fold cross validation เพราะเป็นเทคนิคที่นิยมในการทำงานวิจัยเพื่อใช้ในการทดสอบ ดังแสดงในตารางเปรียบเทียบประสิทธิภาพของตัวแบบ โดยแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพของตัวแบบออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 การแบ่งข้อมูลสำหรับการสร้างตัวแบบ (Training Set) และส่วนที่ 2 ข้อมูลที่ใช้สำหรับทดสอบประสิทธิภาพ (Testing Set)

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ด้วยเทคนิคต่าง ๆ

Classification Technique	Regression Performance			
	Mean Square Error	Root Mean Square Error	Root Relative Square Error	R-square
Neural Network *	0.008	0.090	0.302	93.50%
Linear Regression	0.008	0.087	0.281	93.10%
k-Nearest Neighbor	0.067	0.255	0.856	46.50%

*คือเทคนิคที่มีความเหมาะสมสำหรับการนำไปสร้างตัวแบบพยากรณ์ปริมาณอ้อยในครั้งนี้

จากตารางที่ 5 พบว่าเทคนิคที่มีความหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือ คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.008 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 0.090 ค่าความคลาดเคลื่อนสมบูรณ์ เท่ากับ 0.302 ค่าความคลาดเคลื่อนกำลังสองเท่ากับ 0.483 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่นๆ และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่นๆ ซึ่งมีค่าเท่ากับ 0.935 หรือ 93.50%

ขั้นตอนที่ 6 การนำไปใช้งาน (Deployment Phase) หลังจากการวิเคราะห์ข้อมูล ผลลัพธ์ที่ได้จากการสร้างตัวแบบ และการเปรียบเทียบประสิทธิภาพของตัวแบบ เป็นผลลัพธ์มีประสิทธิภาพดีที่สุด ซึ่งรัฐบาลสามารถนำข้อมูลที่ได้จากการวิจัยนี้ไปเป็นกลยุทธในการกำหนดนโยบายส่งเสริมให้การผลิตอ้อยและน้ำตาล มีประสิทธิภาพสามารถเพิ่มผลผลิตอ้อยส่งโรงงานแปรรูปเป็นน้ำตาล ทำให้ประเทศไทยมีน้ำตาลส่งออกเป็นอันดับ 1 ของโลกในอนาคต นอกจากนี้การพยากรณ์ปริมาณอ้อยเป็นการคาดการณ์ผลผลิตอ้อยล่วงหน้าเกษตรกรผู้ปลูกอ้อยและกลุ่มอุตสาหกรรมอ้อยและน้ำตาลสามารถนำข้อมูลไปวางแผนจัดการผลผลิตและทรัพยากรที่มีอยู่ได้อย่างมีประสิทธิภาพให้เกิดประโยชน์สูงสุดในอนาคต

อภิปรายผลการวิจัย

งานวิจัยนี้ได้สร้างตัวแบบในการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ ผลการทดสอบประสิทธิภาพเทคนิคการพยากรณ์ปริมาณผลผลิตอ้อยของภาคตะวันออกเฉียงเหนือพบว่าเทคนิคที่ให้ค่าประสิทธิภาพที่ดีที่สุดคือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) รองลงมาคือเทคนิคการถดถอยเชิงเส้น (Linear Regression), และเทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor) โดยมีเปอร์เซ็นต์ประสิทธิภาพในการพยากรณ์ 93.5 %, 93.10% และ 46.50% ตามลำดับ สอดคล้องกับงานวิจัยของ สริญญา ศาลางาม (2566) พบว่าวิธีโครงข่ายประสาทเทียมแบบหลายชั้นป้อนไปข้างหน้า ด้วยเทคนิคแบบ แพร่กระจายย้อนกลับ (เรียนรู้แบบมีผู้ฝึกสอน) สามารถพยากรณ์ปริมาณก๊าซ CO_2 ได้อย่างเหมาะสมและมีประสิทธิภาพกว่าวิธีการวิเคราะห์ถดถอย เนื่องจากเป็นวิธีที่มีความยืดหยุ่น อาศัยหลักการทำงานคล้าย ๆ กับเซลล์สมองของมนุษย์ พันธ์ทิพา คนฉลาด และคณะ (2557) ศึกษาหาตัวแบบที่เหมาะสมที่สุดที่จะใช้ในการพยากรณ์การใช้พลังงานรวมทั้งหมดโดยใชวิธีโครงข่ายประสาทเทียมเปรียบเทียบกับวิธีบอกซ์และเจนกินส์ พบว่าวิธีโครงข่ายประสาทเทียมเป็นวิธีที่มีความเหมาะสมที่สุดสำหรับใช้ในการพยากรณ์ เช่นเดียวกับ วิรากานต์ กิตติบวรกุล และคณะ (2566) ศึกษาเรื่องการเปรียบเทียบประสิทธิภาพเทคนิคการเรียนรู้ของเครื่องสำหรับการบำรุงรักษาเชิงคาดการณ์สำหรับเครื่องยนต์อากาศยาน นำเทคนิคการเรียนรู้ของเครื่องทั้งหมด 5 เทคนิคได้แก่ ต้นไม้ตัดสินใจ ป่าแบบสุ่ม ขั้นตอนวิธีเพื่อนบ้านใกล้สุด โครงข่ายประสาทเทียม และ Gradient Boosting with XGBoost ถูกนำมาใช้สำหรับสร้างโมเดลเพื่อคาดการณ์ความเสียหาย ในส่วนของการประเมินโมเดลใช้ค่าสัมประสิทธิ์การกำหนด (R2) และค่ารากที่

สองของความผิดพลาดกำลังสองเฉลี่ย (RMSE) จากการฝึกฝนและทดสอบโมเดลทั้งหมด 5 เทคนิคพบว่าเทคนิคโครงข่ายประสาทเทียมให้ผลลัพธ์ที่ดีที่สุด

สรุปผลการวิจัย

เทคนิคโครงข่ายประสาทเทียม (Neural Network) เป็นตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ ที่เหมาะสมที่สุด ในงานวิจัยครั้งนี้ จากข้อมูลที่ใช้ในการเรียนรู้จำนวน 9 ตัวแปร ได้แก่ 1. จังหวัด 2. ปีการผลิต 3. ปริมาณน้ำฝนเฉลี่ย 4. อุณหภูมิเฉลี่ย 5. พื้นที่ปลูกอ้อย 6. ผลผลิต/ไร่ 7. ราคาอ้อย 8. ราคามันสำปะหลัง 9. ราคาข้าวโพดเลี้ยงสัตว์ ตัวแบบได้ดีที่สุด ผลการทดสอบประสิทธิภาพการถดถอย พบว่าค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์เท่ากับ 9.35 ซึ่งสามารถนำมาประยุกต์ใช้เป็นตัวแบบในการพยากรณ์ปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือในปีต่อไปเพื่อช่วยประกอบการตัดสินใจในการวางแผนการจัดการปริมาณผลผลิตอ้อยให้มีประสิทธิภาพเกิดปัญหาน้อยที่สุด

ข้อเสนอแนะ

ตัวแบบที่สร้างขึ้นในงานวิจัยครั้งนี้มีความเหมาะสมกับชุดข้อมูลปริมาณผลผลิตอ้อยภาคตะวันออกเฉียงเหนือ ดังนั้นผู้ที่มีความสนใจที่จะต่อยอดงานวิจัยต้องพิจารณาปัจจัยอื่น ๆ เพิ่ม เช่น สภาพดิน สภาพแวดล้อม ศัตรูพืช โรคพืช เป็นต้น เนื่องจากสภาพแวดล้อมและภูมิประเทศในแต่ละภูมิภาคมีความแตกต่างกัน อีกทั้งค่าที่ทำนายได้จากตัวแบบเป็นเพียงสารสนเทศส่วนหนึ่งที่ช่วยประกอบการตัดสินใจในการวางแผนการจัดการปริมาณผลผลิตอ้อย นอกจากนี้ผู้วิจัยควรศึกษาตัวแบบการจำแนกประเภทข้อมูลอื่น เช่น ต้นไม้ตัดสินใจ (Decision Tree), Support Vector Machines (SVM), ต้นไม้สุ่ม (Random Forest) เป็นต้น เพื่อเพิ่มประสิทธิภาพในการจำแนกประเภทข้อมูล และนำตัวแบบที่เหมาะสมไปพัฒนาเว็บแอปพลิเคชันสำหรับอำนวยความสะดวกให้แก่ อุตสาหกรรมอ้อยและน้ำตาล รวมถึงเกษตรกรผู้ปลูกอ้อยในการวางแผนการจัดการปริมาณผลผลิตอ้อยให้มีประสิทธิภาพต่อไป

บรรณานุกรม

Begg, R. and Palaniswami, M. (2006). Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques, 1-396. doi:10.4018/978-1-59140-836-9.

จตุภัทร เมฆพายัพ[1] กิดาการ สายธนู[2]. (2560). การพยากรณ์ผลผลิตอ้อยในบริเวณภาคตะวันออกของประเทศไทยด้วย เทคนิค ANN. คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา, http://dspace.lib.buu.ac.th/xmlui/handle/1234567890/3489.

ณัฐนนท์ ลีลาตระกูล[1] สุนิสา ริมเจริญ[2] สุภาวดี ศรีคาด[3]. (2559). ขั้นตอนวิธีสำหรับการพยากรณ์ผลผลิตอ้อยเพื่อป้อนเข้าสู่โรงงานอุสาหกรรม. คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา. https://buuir.buu.ac.th/xmlui/handle/1234567890/1974.

ธนะรัตน์ รัตนกูล[1] กันต์ธมน สุขกระจ่า[2]. (2565). การทำนายปริมาณการใช้เชื้อเพลิงขนส่งด้วยโครงข่ายประสาทเทียมกรณีศึกษา รถบรรทุกดินถมในพื้นที่จังหวัดชายแดนใต้ประเทศไทย. วารสารวิชาการบัณฑิตศึกษา, 1(1), 35-44.

พงศ์ หลวงมูล[1] ถาวร อ่อนประไพ[2]. (2563). การประมาณผลผลิตข้าวด้วยค่าดัชนีพืชพรรณโดยใช้ข้อมูลภาพแบบหลายช่วงคลื่นจากอากาศยานไร้คนขับ. วารสารเกษตร, 37(2), 193-205.

พันธ์ทิพา คนฉลาด[1] นิภาพร ชุติมันต์[2] บังอร กุมพล[3]. (2557). การพยากรณ์การใช้พลังงานรวมทั้งหมดของประเทศไทย โดยวิธีของบอกซ์-เจนกินส์และโครงข่ายประสาทเทียม. วารสารวิจัยและพัฒนา มหาวิทยาลัยราชภัฏเลย, 9(28), 31-38.

ภัคจิรา ทวีกาญจน์. (2562). การวิเคราะห์ศักยภาพทางการแข่งขันน้ำตาลทรายของไทยในตลาดอาเซียน. วารสารการศึกษาและการพัฒนาสังคม, 5(2), 72-80.

วรินทรา แซ่โล้[1] พาพิศ วงชัยสุวัฒน์[2] วรรณดี ไทยสยาม[3]. (2566). การพยากรณ์อัตราการไหลโดยใช้แบบจำลองโครงข่ายประสาทเทียมและพารามิเตอร์การตอบสนองของเวลา. วารสาร “Engineering and Technology Horizons”, 40(4),

วิรากานต์ กิตติบวรกุล[1] ศรายุทธ นนท์ศิริ[2] พิชิตชัย คำอินทร์[3]. (2566). การเปรียบเทียบประสิทธิภาพเทคนิคการเรียนรู้ของเครื่องสำหรับการบำรุงรักษาเชิงคาดการณ์สำหรับเครื่องยนต์อากาศยาน. วารสารวิชาการสมาคมสถาบันอุดมศึกษาแห่งประเทศไทยในพระบรมราชูบถัมภ์สมเด็จพระเทพรัตนราชสุดาสยามบรมราชกุมารี, 11(1).

ศรีรักษ์ ศรีทองชัย.(2566). การเปรียบเทียบวิธีการประมาณค่าสูญหายในการพยากรณ์ความเข้มข้นของ PM2.5 ด้วยโครงข่ายประสาทเทียม LSTM. วารสารนวัตกรรมการเรียนรู้และเทคโนโลยี, 3(1), 59-67.

ศรุดา สมพอง[1] จินดา แก่นสมบัติ[2] อภิวัฒน์ บุญอเนก[3]. (2564). น้ำกับความแห้งแล้งในภาคอีสาน. วารสารมนุษยศาสตร์ ฉบับบัณฑิตศึกษา, 10(2), 67-80.

สริญญา ศาลางาม. (2565). ตัวแบบการพยากรณ์ปริมาณก๊าซคาร์บอนไดออกไซด์ด้วยวิธีการ วิเคราะห์การถดถอยและวิธีโครงข่ายประสาทเทียม. วารสารวิชาการเทคโนโลยีอุตสาหกรรม : มหาวิทยาลัยราชภัฏสวนสุนันทา, 10(2). 54-65.

สำนักงานคณะกรรมการอ้อยและน้ำตาลทราย. (2566). รายงานพื้นที่ปลูกอ้อยของประเทศไทย. http://www.ocsb.go.th

สำนักงานเศรษฐกิจการเกษตร. (2566). ราคาอ้อยโรงงาน ราคามันสำปะหลัง ราคาข้าวโพดเลี้ยงสัตว์. http://www.oae.go.th

สำนักงานสถิติแห่งชาติ. (2566). สถิติปริมาณน้ำฝนและอุณหภูมิ ณ สถานีอุตุนิยมวิทยา. http://service.nso.go.th

เอกวิจัย เมยไธสง[1] ฉวีวรรณ สีสม[2] สุเทพ เมยไธสง[3]. (2565). การพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษา สาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง. วารสารพุทธปรัชญาวิวัฒน์, 6(2), 329-340.