การเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ราคาข้าวเปลือกหอมมะลิ
A Comparison of the Forecasting Technique Efficiency for the Price of Jasmine Paddy

กนกวรรณ วามะขันธ์[1], ณัฏฐ์ ศรีเมืองทอง[2] และอนุพงศ์ สุขประเสริฐ[3*]
[1,2,3*]คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม
Kanogwan Wamakhan[1] Nath Srimaungtong[2] and Anupong Sukprasert[3*]
[1,2,3*]Mahasarakham Business School, Mahasarakham University

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ราคาข้าวเปลือกหอมมะลิด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลมาวิเคราะห์ตามกระบวนการมาตรฐานในการทำเหมือง ตั้งแต่เดือนมกราคม พ.ศ.2557 ถึง เดือนธันวาคม พ.ศ.2564 จำนวน 96 เดือน ด้วยเทคนิคการทำเหมืองข้อมูล ทั้งหมด 4 เทคนิค ได้แก่ 1) เทคนิคโครงข่ายประสาทเทียม 2) เทคนิคเพื่อนบ้านใกล้ที่สุด 3) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน 4) เทคนิคสมการถดถอยเชิงเส้น จากนั้นทำการทดสอบประสิทธิภาพของตัวแบบเพื่อหาตัวแบบที่มีความเหมาะสมที่สุดในการสร้างตัวแบบพยากรณ์ฯ ผลการวิจัยพบว่า เทคนิคโครงข่ายประสาทเทียม มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบพยากรณ์ราคาข้าวเปลือกหอมมะลิ ซึ่งทำให้ความคลาดเคลื่อนในการพยากรณ์ต่ำที่สุด โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1,892,401.917 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 1,375.646 ค่าความคลาดเคลื่อนสัมบูรณ์กับ 0.697 ค่าสัมประสิทธิ์ในการตัดสินใจเท่ากับ 67.00%.

คำสำคัญ : 1) การพยากรณ์ 2) เทคนิคโครงข่ายประสาทเทียม 3) เทคนิควิธีเพื่อนบ้านใกล้ที่สุด 4) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน 5) เทคนิคสมการถดถอยเชิงเส้น

Abstract

This research aims to compare the effectiveness of forecasting models for the price of Jasmine Paddy using data mining techniques. The analysis is conducted on data following cross industry standard process for data mining (CRISP-DM) from January 2014 to December 2021, spanning 96 months. Four data mining techniques were employed, namely: 1) Neural Network 2) k-Nearest Neighbor 3) Support Vector Machines and 4) Linear Regression. Afterward, model performance assessments were undertaken to determine the optimal model for the development of forecasting models. The investigation revealed that Neural Network are particularly adept at constructing predictive models for the price of Jasmine Paddy, thereby yielding minimal forecast errors with Mean Squared Error (MSE) of 1,892,401.917, root Mean Squared Error (RMSE) of 1,375.646, Absolute Error of 0.697 and Coefficient of Determination of 67.00%.

Keywords: 1) Forecasting 2) Neural Network 3) Nearest Neighbor 4) Support Vector Machines 5) Linear Regression

---

[1]นิสิตปริญญาโท หลักสูตรบัญชีมหาบัณฑิต (Graduate Student, Master of Accountancy Program)
[2,3*] อาจารย์ประจำคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม E-mail: anupong.s@acc.msu.ac.th

บทนำ

ประเทศไทยเป็นประเทศเกษตรกรรม ที่มีพื้นที่เพาะปลูกจำนวนมากกระจายตัวอยู่ในทุกภูมิภาค ส่งผลให้มีสินค้าภาคเกษตรจำนวนมากจำหน่ายตามท้องตลาดทั้งในประเทศ และส่งออกไปยังต่างประเทศ หนึ่งในสินค้าดังกล่าวคือ ข้าว ซึ่งเป็นสินค้าที่สร้างรายได้มหาศาลให้กับประเทศไทยในแต่ละปีตั้งแต่อดีตจนถึงปัจจุบัน โดยข้าวที่เกษตรกรปลูกมีหลากหลายสายพันธ์ พบว่าพันธ์ข้าวที่ดีที่สุด คือ ข้าวหอมมะลิ ซึ่งเป็นข้าวที่มีคุณภาพดีที่สุด ราคาแพงที่สุด และเป็นที่ต้องการของตลาดต่างประเทศ อีกทั้งรัฐบาลส่งเสริมให้มีการขยายพื้นที่การเพาะปลูก และเพิ่มปริมาณการส่งออกข้าวหอมมะลิ ทำให้ปัจจุบันเกษตรกรผู้ปลูกข้าวหันมานิยมปลูกข้าวหอมมะลิกันเป็นจำนวนมาก เพื่อให้เพียงพอกับความต้องการของผู้บริโภคทั้งในประเทศไทยและการส่งออกไปยังตลาดต่างประเทศ โดยเฉพาะในทวีปเอเชีย ข้าวหอมมะลินอกจากจะเป็นข้าวที่มีคุณภาพดีที่สุดแล้ว ยังมีคุณสมบัติเฉพาะ คือมีความหอมเฉพาะตัวหอมคล้ายใบเตย (ศศิธร โกฏสืบ และกัลยา บุญหล้า, 2559) ที่มาพร้อมกับคุณค่าทางอาหาร และกลูโตสสูง สามารถนำมาประกอบอาหารและแปรรูปเป็นผลิตภัณฑ์อื่นๆอีกมากมาย ทั้งยังมีคุณสมบัติทางเศรษฐศาสตร์ที่ดี เนื่องจากมีราคาสูงเป็นที่ต้องการของตลาด ทำให้เกษตรกรผู้ปลูกข้าวมีรายได้จากการจำหน่ายผลผลิตข้าวหอมมะลิ ประเทศไทยถือเป็นแหล่งผลิตข้าวหอมมะลิที่มีคุณภาพดีที่สุดแห่งหนึ่ง โดยมีแหล่งเพาะปลูกสำคัญในภาคตะวันออกเฉียงเหนือ (เขตทุ่งกุลาร้องไห้) และมีพื้นที่เพาะปลูกครอบคลุมกว่า 19 ล้านไร่ทั่วประเทศ รองลงมาคือภาคเหนือ เนื่องจากสภาพดินฟ้า-อากาศและพื้นที่เพาะปลูกของทั้งสองภาคคล้ายคลึงกันเหมาะแก่การเจริญเติบโตของข้าวหอมมะลิ (กรมฝนหลวงและการบินเกษตร, 2564) ฝนจะเริ่มตกตั้งแต่เดือนพฤษภาคม ชาวนาจะเริ่มหว่านไถในเดือนมิถุนายน และเพาะปลูกอยู่ในช่วงเดือนกรกฎาคม – สิงหาคม เมื่อฝนเริ่มหมดปลายเดือนตุลาคมจนถึงต้นเดือนพฤศจิกายน จึงเริ่มเก็บเกี่ยว ช่วงเดือนพฤศจิกายนความชื้นจะน้อยเพราะเป็นช่วงที่ลมหนาวจากเมืองจีนเริ่มพัดเข้ามาในสองภาคนี้ทำให้อากาศแห้งเหมาะในการเก็บเกี่ยว การตาก การนวด ก็ทำได้ง่ายเพราะน้ำแห้งนาหมดแล้ว ไม่มีฝน จึงทำให้ได้เมล็ดข้าวที่มีคุณภาพ สำหรับการปลูกข้าวหอมมะลิจะทำกันได้ดีเฉพาะพื้นดินที่เป็นดินปนทรายและดินมีความเค็มเล็กน้อยบวกกับสภาพภูมิอากาศที่แห้งแล้งสลับชื้น ลักษณะดังกล่าวทำให้ข้าวหอมมะลิมีคุณภาพดี ยิ่งในฤดูกาลใดหากช่วงเก็บเกี่ยวพื้นนาแห้งสนิทไม่มีน้ำขังผสมกับอากาศหนาวเย็นจะทำให้ข้าวมีความหอมมากยิ่งขึ้น (รักบ้านเกิด, 2555) การส่งออกข้าวหอมมะลิของไทย มีปริมาณการส่งออก ในช่วงมกราคม พ.ศ. 2560 - ธันวาคม 2564 มีปริมาณการส่งออกปรับลดที่ลดลงอย่างต่อเนื่อง อีกทั้งราคาข้าวเปลือกหอมมะลิมีการเปลี่ยนแปลงตลอดเวลาขึ้นอยู่กับสถานการณ์ตลาด, ปริมาณการผลิต, และปัจจัยอื่น ๆ ที่มีผลกระทบต่อการค้าข้าว นอกจากนี้ยังขึ้นอยู่กับประเทศและภูมิภาค พื้นที่เพาะปลูก และอุณหภูมิ ที่เกี่ยวข้องกับการผลิตข้าวหอมมะลิด้วย (ศศิธร โกฏสืบ และกัลยา บุญหล้า, 2559) เนื่องจากสภาพดินฟ้าอากาศและภูมิประเทศที่เหมาะสมต่อการเจริญเติบโตของข้าวหอมมะลิ ราคาข้าวเปลือกหอมมะลิในประเทศไทยขึ้นอยู่กับปัจจัยต่างๆ หลายปัจจัยได้แก่ ปริมาณการส่งออกข้าว ปริมาณการผลิตข้าว อัตราแลกเปลี่ยนเงินบาทต่อดอลลาร์สหรัฐ ราคาพืชอื่นๆที่มีผลต่อราคาข้าว และปริมาณน้ำฝนเฉลี่ย นอกจากนี้ ราคาข้าวเปลือกหอมมะลิยังขึ้นอยู่กับคุณภาพของข้าวด้วย โดยข้าวเปลือกหอมมะลิที่มีคุณภาพดีจะมีราคาสูงกว่าข้าวเปลือกหอมมะลิที่มีคุณภาพต่ำ (กรมการค้าต่างประเทศ กระทรวงพาณิชย์, 2567) ปัญหาของราคาข้าวจะเกิดจากหลายสาเหตุ ทั้งด้านปัญหาต้นทุนการผลิต และผลผลิตต่อไร่ ต้นทุนสูงแต่ ผลผลิตต่อไร่ต่ำ ซึ่งหากเปรียบเทียบราคาส่งออกข้าวหอมมะลิของไทยและคู่แข่งข้าวหอมมะลิก็จะเห็นได้ว่าจีนหันมานำเข้าข้าวหอมมะลิจาก กัมพูชาและเวียดนามเพิ่มมากขึ้น ด้วยราคาที่ถูกกว่าของไทยและคุณภาพดีใกล้เคียงกับของไทย ซึ่งราคาส่งออกข้าวหอมมะลิของไทยมีราคาสูงกว่าคู่แข่ง (เฉลิมวุฒิ คำเมือง และวชิร รารักษ์ โอรสรัมย์, 2564).

ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ราคาข้าวเปลือกหอมมะลิโดยใช้เทคนิคการทำเหมืองข้อมูล ด้วยวิธีการจำแนกประเภทข้อมูลทั้ง 4 เทคนิค ได้แก่ เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors : K-NN) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) และเทคนิคสมการถดถอยเชิงเส้น (Linear Regression) เพื่อหาตัวแบบที่มีความเหมาะสมที่สุดในการนำไปสร้างตัวแบบการพยากรณ์ราคาข้าวเปลือกหอมมะลิต่อไป

การทบทวนวรรณกรรม

การพยากรณ์ คือ การคาดคะเนการทำนายเหตุการณ์หรือสภาพการณ์ในอนาคตโดยศึกษารูปแบบการเกิดเหตุการณ์หรือสภาพการณ์จากข้อมูลที่เก็บรวบรวมไว้อย่างมีระบบหรือจากความรู้ ความสามารถ ประสบการณ์และวิจารณญาณของผู้พยากรณ์(ทรงศิริ แต้สมบัติ, 2549)

เฉลิมวุฒิ คำเมือง และวชิรารักษ์ โอรสรัมย์ (2562) ได้ศึกษาเกี่ยวกับการพยากรณ์ราคาข้าวเปลือกหอมมะลิด้วยเทคนิคการทำเหมืองข้อมูล ผลการศึกษาพบว่า แบบจำลองทางคณิตศาสตร์ ทำนายราคาข้าวเปลือกหอมมะลิและวิธีแรนดอมฟอเรสต์ พบว่าวิธีแรนดอมฟอเรสต์ได้ทำนายราคาข้าวเปลือกหอมมะลิใกล้เคียงกับค่าจริงที่สุด และเมื่อคำนวณค่า RMSE ได้ 1,699.64 และได้ค่า MAPE เป็น 9.80

ศศิธร โกฏิสืบ และกัลยา บุญหล้า (2559) ได้ศึกษาการสร้างตัวแบบเพื่อการพยากรณ์ราคาข้าวหอมมะลิ 105 โดยใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2540 ถึงเดือนธันวาคม พ.ศ.2558 ผลการศึกษาพบว่าข้อมูลของการพยากรณ์ราคาข้าวเปลือกหอมมะลิ 105 วิธีบ๊อกซ์และเจนกินส์ โดยให้ค่าคลาดเคลื่อนกำลังสองน้อยที่สุดเท่ากับ0.293 พันล้านบาท/ตัน เมื่อเปรียบเทียบกับวิธีการปรับให้เรียบแบบเอกซ์โปเนนเชียลซ้ำสองครั้ง วิธีการพยากรณ์ที่เหมาะสมที่สุด คือ วิธีบ็อกซ์และเจนกิ้นส์ โดยมีตัวแบบที่เหมาะสมคือ ARIMA (1,1,2) ทั้งนี้ควรมีการนำปัจจัยต่างๆ ที่เกี่ยวข้องกับค่าพยากรณ์ เช่น พื้นที่เพาะปลูก อุณหภูมิ มาใช้ในการพยากรณ์ร่วมด้วย เพื่อเพิ่มความสมบูรณ์ให้กับค่าพยากรณ์

ธนกร สุทธสนธ์ (2565) ได้ศึกษาวิจัยเรื่องการประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเชิงลึกสำหรับการพยากรณ์ราคาพืชไร่ของประเทศไทย เพื่อเลือกตัวแบบการพยากรณ์ที่มีความเหมาะสมกับข้อมูลอนุกรมเวลาราคาพืชไร่ ผลการวิจัย ผู้วิจัยนำเสนอวิธีการสร้างและคัดเลือกตัวแบบพยากรณ์ที่เหมาะสมกับข้อมูลอนุกรมเวลาราคาพืชไร่ของประเทศไทย ด้วยตัวแบบ ANN และ ตัวแบบ LSTM และตัวแบบ Hybrid EMD-LSTM ผลการวิจัยพบว่า ตัวแบบ Hybrid EMD-LSTM มีความแม่นยำมากกว่าตัวแบบ ANN และ ตัวแบบ LSTM ดังนั้น สรุปได้ว่าตัวแบบผสมที่ทีมผู้วิจัยนำเสนอมีความเหมาะสมที่จะนำไปพยากรณ์ข้อมูลอนุกรมเวลาราคาพืชไร่ทั้ง 3 ชุด

พนิดา สมบัติมาก, ภัสสร จันทร์หอม, ศุภกร รัศมี, โอฬาร รุ่งมณีธรรมคุณ และสายชล สินสมบูรณ์ทอง. (2561) ได้ศึกษาวิจัยเรื่องการเปรียบเทียบประสิทธิภาพในการจำแนกเมื่อข้อมูลมีค่านอกเกณฑ์ในการทำเหมืองข้อมูล พบว่า จากการเปรียบเทียบข้อมูลโรคมะเร็งเต้านมของรัฐวิสคอนซิน วิธีที่มีประสิทธิภาพสูงสุด คือ วิธีโครงข่ายประสาทเทียม โดยการสุ่มของโปรแกรม SPSS โรคเบาหวานของชาวพม่า ในประเทศอินเดีย วิธีที่มีประสิทธิภาพสูงสุด คือ วิธีเพื่อนบ้านใกล้ที่สุด k ตัว โดยการสุ่ม SPSS และ WEKA และการชำระเงินด้วยบัตรเครดิตของลูกค้า วิธีที่มีประสิทธิภาพสูงสุด คือ วิธีเพื่อนบ้านใกล้ที่สุด k ตัว โดยการสุ่ม SPSS และ WEKA ชุดข้อมูลที่มีค่านอกเกณฑ์อยู่ในระดับปานกลางและสูงให้ผลการจำแนกที่เหมือนกัน ซึ่งแตกต่างจากชุดข้อมูลที่มีค่านอกเกณฑ์ในระดับที่ต่ำ

ขนิษฐา กุลนาวิน วันเพ็ญ โพธิ์เกษม และ ศันสนีย์ เลี้ยงพานิชย์ (2565) การออกแบบแบบจำลองการพยากรณ์ผลผลิตมันสาปะหลังในเขตพื้นที่ จังหวัดนครราชสีมาด้วยเทคนิคเหมืองข้อมูล ซึ่งใช้วิธีการวิเคราะห์องค์ประกอบ จากนั้นนาไปพัฒนาแบบจาลองเพื่อการพยากรณ์ โดยใช้เทคนิคโครงสร้างต้นไม้ เทคนิค โครงข่ายประสาทเทียม เทคนิคเครือข่ายเบย์ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน พบว่าประสิทธิภาพ ของแบบจำลองโดยใช้เทคนิคเครือข่ายเบย์ให้ค่าความถูกต้อง เมื่อทดสอบด้วยวิธีการแบ่งข้อมูลแบบสุ่ม ด้วยการแบ่งร้อยละ ได้แสดงค่าความแม่นยา ค่าความระลึกได้ และค่าการวัดประสิทธิภาพในภาพรวม ดีที่สุด มีค่าดังนี้ 0.673, 0.606, 0.685 และ 0.629 ตามลาดับ เมื่อเทียบกับค่าความผิดพลาดของรากที่สอง ของค่าความคลาดเคลื่อนเฉลี่ยกาลังสอง และยังวัดค่าความแม่นยา โดยใช้เกณฑ์ในการเปรียบเทียบจาก รากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกาลังสอง และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย มีค่าดังนี้ 0.303 และ 0.201 ตามลาดับ แสดงให้เห็นว่าค่าความคลาดเคลื่อนน้อยแบบจาลองที่ได้ยิ่ง มีความแม่นยำของการพยากรณ์ดี

อนุพงศ์ สุขประเสริฐ, (2564) **การทำเหมืองข้อมูล (Data mining)** เป็นเทคนิคกระบวนการการวิเคราะห์ข้อมูลที่มีจำนวนมากเพื่อหาแนวโน้มหรือความสัมพันธ์ โดยใช้เทคนิคทางสถิติ เพื่อเชื่อมโยงฐานข้อมูลเพื่อหาความน่าจะเป็นที่จะเกิดขึ้นเพื่อนำไปพัฒนาหรือเพื่อประกอบการตัดสินใจที่สำคัญในด้านที่เกี่ยวข้อง หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมี จำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และ ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัย หลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

วัตถุประสงค์

เพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ราคาข้าวเปลือกหอมมะลิด้วยเทคนิคการทำเหมืองข้อมูล

วิธีการดำเนินงานวิจัย

ขั้นตอนวิธีการดำเนินงานวิจัยผู้วิจัยได้อ้างอิงจากกระบวนการมาตรฐานในการทำเหมืองข้อมูล (Cross Standard Process for Data Mining : CRISP-DM) แบ่งออกเป็น 6 ขั้นตอนดังนี้ (อนุพงศ์ สุขประเสริฐ, 2564)

ขั้นตอนที่ 1 การทำความเข้าใจปัญหา (Business Understanding)

ผู้วิจัยได้ทบทวนวรรณกรรมทางวิชาการที่เกี่ยวข้องและศึกษารวบรวมข้อมูลปัจจัยที่ส่งผลกระทบต่อราคาข้าวเปลือกหอมมะลิ ซึ่งปัญหาของราคาข้าวเกิดจากหลายสาเหตุ เช่น ปัญหาด้านการผลิตต่อไร่ ระบบชลประทานในพื้นที่เพาะปลูกมีจำกัด และราคาส่งออกข้าวไทยสูงกว่าประเทศเพื่อนบ้าน ทำให้การส่งออกข้าวไปยังต่างประเทศลดลง ดังนั้นการพยากรณ์ราคาจึงสามารถช่วยทำให้ประเมินราคาสินค้าในอนาคตได้ ดังนั้นการศึกษาแนวโน้มและพยากรณ์ราคาข้าวเปลือกหอมมะลิในประเทศไทยจึงมีความจำเป็นอย่างมาก เพื่อที่จะนำผลการศึกษาที่ได้จากการพยากรณ์มาเป็นแนวทางในการตัดสินใจสำหรับผู้ผลิตและผู้ส่งออก ในการวางแผนการผลิตและส่งออกให้มีประสิทธิภาพดียิ่งขึ้นและมียอดการส่งออกเพิ่มขึ้นและสอดคล้องกับความต้องการของตลาด

ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลปัจจัยที่ส่งผลต่อราคาข้าวเปลือกหอมมะลิ ทั้ง 6 ตัวแปร ประกอบไปด้วย ปริมาณการผลิตข้าว (Product), ปริมาณการส่งออกข้าว (Export), ปริมาณน้ำฝนเฉลี่ย (Rain), ราคาข้าวโพด (P-corn), ราคามันสำปะ (P-cassava), อัตราแลกเปลี่ยนเงินบาทต่อดอลลาร์สหรัฐ (THB/USD) ดังแสดงในภาพที่ 1 โดยได้รวบรวมมาจากสำนักงานเศรษฐกิจการเกษตร ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา และธนาคารแห่งประเทศไทย ซึ่งถูกจัดเก็บในรูปแบบรายเดือน โดยผู้วิจัยได้ทำการรวบรวมข้อมูลตั้งแต่เดือน มกราคม พ.ศ.2557 จนถึงเดือน ธันวาคม พ.ศ.2564 รวมทั้งสิ้น 96 เดือน

ภาพที่ 1 กรอบแนวคิดการวิจัย

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูล คือขั้นตอนในการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมก่อนการวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการเตรียมข้อมูลโดยมีขั้นตอนดังนี้

3.1 การนำข้อมูลสำหรับการวิเคราะห์โดยรวบรวมข้อมูลให้อยู่ในรูปแบบไฟล์ Excel และนำข้อมูลเข้ามาประมวลผลโดยใช้โปรแกรม RapidMiner Studio Version 10.2

3.2 การทำความสะอาดข้อมูล เนื่องจากข้อมูลที่ได้มายังมีข้อมูลที่สูญหาย (Missing Value) เกิดขึ้น ผู้วิจัยจึงได้ทำการแทนที่ค่าสูญหายเหล่านั้นด้วยค่าเฉลี่ย (Average)

3.3 การกำหนดหน้าที่ให้กับแอตทริบิวต์ โดยกำหนดให้แอตทริบิวต์ ปีการผลิต (Year) ให้ทำหน้าที่เป็นไอดี (ID) และกำหนดแอตทริบิวต์ ราคาข้าวเปลือกหอมมะลิ ให้ทำหน้าที่เป็นตัวแปรตาม (Label) เพื่อใช้สำหรับการพยากรณ์ และแอตทริบิวต์ที่เหลือจะทำหน้าที่เป็นตัวแปรอิสระ ดังแสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลที่ใช้ในงานวิจัย

ลำดับ	ตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	Year (ID)	ปีการผลิต	Integer
2	Product	ปริมาณการผลิตข้าว	Real
3	Export	ปริมาณการส่งออกข้าว	Real
4	Rain	ปริมาณน้ำฝนเฉลี่ย	Real
5	P_corn	ราคาข้าวโพด	Real
6	P_cassava	ราคามันสำปะหลัง	Real
7	THB/USD	อัตราแลกเปลี่ยนเงินบาทต่อดอลล่าห์สหรัฐ	Real
8	Price (Label)	ราคาข้าวเปลือกหอมมะลิ	Real

ขั้นตอนที่ 4 การสร้างแบบจำลอง (Modeling)

ขั้นตอนนี้ผู้วิจัยได้นำข้อมูลที่ผ่านขั้นตอนของการเตรียมข้อมูลเรียบร้อยแล้ว มาสร้างตัวแบบสำหรับการพยากรณ์ราคาข้าวเปลือกหอมมะลิ โดยใช้โปรแกรม RapidMiner Studio Version 10.2 ซึ่งมีทั้งหมด 4 เทคนิค ประกอบด้วย

4.1 เทคนิคโครงข่ายประสาทเทียม (Neural Network) เป็นเครื่องมือที่มีแนวคิดมาจากการจำลองโครงสร้างระบบประสาทภายในสมองของมนุษย์ โครงข่ายประสาทมีความสามารถในการจำแนกข้อมูลและทำนายผลลัพธ์จากข้อมูลที่ให้มา เซลล์ประสาท (Neuron) ในโครงข่ายประสาทมีค่าน้ำหนักสำหรับการเชื่อมต่อระหว่างชั้นและการคำนวณผลรวมของข้อมูลและค่าหน้ำหนักเพื่อให้คำตอบของระบบไปตามที่ถูกต้อง

4.2 เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: K-NN) เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใด (K) มากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุดนั้นลักษณะการทำงานแบบไม่ได้ใช้ข้อมูลชุดเรียนรู้ (Training data) ในการสร้างแบบจำลองแต่จะใช้ข้อมูลนี้มาเป็นตัวแบบจำลองเลย

4.3 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) เป็นเทคนิคการเรียนรู้จากกลุ่มข้อมูลที่ใช้ในงานที่ต้องการจำแนกข้อมูลอย่างชัดเจนเป็นกลุ่ม เป็นประเภทต่างๆ เป็นวิธีเทคนิคที่มีคุณภาพสูงสำหรับการจำแนกข้อมูลที่มีมากมายและซับซ้อน เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีหลักการคือการหาขอบเขตที่เรียกว่า แบบจำแนก หรือ แบบพื้นระหว่างกลุ่ม ที่แยกข้อมูลในแต่ละกลุ่มออกจากกันอย่างชัดเจน

4.4 เทคนิคสมการถดถอยเชิงเส้น (Linear Regression) เป็นการศึกษารูปแบบความสัมพันธ์ของตัวแปรเพื่อใช้ทำนายค่าตัวแปรที่ต้องการศึกษา โดยอาศัยความรู้เกี่ยวกับค่าของตัวแปรอื่นที่เกี่ยวข้อง (ตัวแปรที่กำหนด) ซึ่งอาจจะมีหนึ่งตัวหรือมากกว่า ตัวแปรที่ต้องการศึกษา เรียกว่า ตัวแปรตาม (Dependent variable) แทนด้วยตัวแปร y ตัวแปรที่กำหนดให้ เรียกว่า ตัวแปรอิสระ (Independent variable) แทนด้วยตัวแปร x ซึ่งทั้งสองตัวแปรนี้จะต้องมีรูปแบบความสัมพันธ์ในเชิงเส้นตรง (Linear) และจะมีรูปแบบสมการถดถอยเชิงเส้นดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

เมื่อ Y คือ ตัวแปรตาม
X คือ ตัวแปรอิสระ
$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ คือ พารามิเตอร์
ε คือ ความคลาดเคลื่อน

ขั้นตอนที่ 5 การประเมินผล (Evaluation)

เมื่อได้ตัวแบบทั้งหมดจากขั้นตอนที่ 4 ผู้วิจัยได้ทำการวัดประสิทธิภาพของตัวแบบ เพื่อหาตัวแบบที่มีความเหมาะสมที่สุดที่จะนำไปใช้สำหรับการพยากรณ์ต่อไป โดยทำการแบ่งข้อมูลออกเป็น 2 ส่วน ด้วยวิธีการ 10-fold cross validation เพื่อสร้างตัวแบบ (Training Set) และทดสอบประสิทธิภาพของตัวแบบ (Testing Set) ซึ่งเกณฑ์ที่ใช้สำหรับการทดสอบประสิทธิภาพของตัวแบบนี้ ประกอบด้วย

5.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เป็นการวัดค่าความคลาดเคลื่อนโดยการนำค่าความคลาดเคลื่อนมายกกำลังแล้วนำไปหาค่าเฉลี่ย ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก โดยมีสูตรดังนี้

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้
y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.2. รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) เป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการที่ใช้ในการวัดค่าความแม่นยำดังนี้

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.3 ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root Relative Squared Error: RRSE) เป็นหนึ่งในวิธีการวัดความแม่นยำของการทำนายหรือโมเดลทางสถิติ โดยให้ความสัมพันธ์กับค่าเฉลี่ยผลรวมกำลังสองของความคลาดเคลื่อน โดยมีหน่วยวัดเป็นหน่วยวัดค่าสังเกต (ground truth) ที่ถูกยกกำลังสอง ของหน่วยวัดค่าสังเกต โดยมีสูตรดังนี้

$$RRSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t-\hat{y})^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2}}$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

$\bar{y}$ คือ ค่าเฉลี่ยผลลัพธ์ที่สังเกตได้

5.4 สัมประสิทธิ์การตัดสินใจพหุคูณ (Coefficient of Determination: R^2) เป็นค่าที่แสดงอิทธิพลของตัวแปรอิสระทั้งหมดในสมการที่มีต่อตัวแปรตาม ซึ่งจะแสดงถึงขอบเขตความสามารถในการคาดเดาการแปรผันของตัวแปรตามของโมเดล โดยมีสูตรดังนี้

$$R^2 = 1 - \left(\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}\right)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

$\bar{y}$ คือ ค่าเฉลี่ยผลลัพธ์ที่สังเกตได้

ขั้นตอนที่ 6 การนำไปใช้ (Deployment)

ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลในครั้งนี้ จะได้ตัวแบบที่มีความเหมาะสมที่สุดสำหรับการนำไปพยากรณ์ราคาข้าวเปลือกหอมมะลิ ซึ่งสามารถนำผลวิจัยในครั้งนี้ไปพัฒนาเป็นระบบสารสนเทศสำหรับการพยากรณ์ราคา

ข้าวเปลือกหอมมะลิ หรือนำไปประยุกต์ใช้สำหรับการพยากรณ์ราคาสินค้าเกษตรที่สำคัญของประเทศไทยได้ โดยอาจจะมีการเพิ่มปัจจัยด้านอื่นๆ เพื่อให้มีความแม่นยำสำหรับการพยากรณ์มากยิ่งขึ้น

ผลการวิจัย

การเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ราคาข้าวเปลือกหอมมะลิด้วยเทคนิคการทำเหมืองข้อมูล โดยนำข้อมูลปัจจัยที่ส่งผลต่อการเปลี่ยนแปลงของราคาข้าวเปลือกหอมมะลิทั้ง 6 ปัจจัย มาวิเคราะห์ตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล ด้วยเทคนิคประมาณค่าข้อมูล ทั้งหมด 4 เทคนิค ได้แก่ 1) เทคนิคโครงข่ายประสาท 2) เทคนิคเพื่อนบ้านใกล้ที่สุด 3) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และ 4) เทคนิคสมการถดถอยเชิงเส้น จากนั้นทำการทดสอบประสิทธิภาพของการประมาณค่าด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error)ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ค่ารากที่สองของความผิดพลาดกำลังสองเฉลี่ย (Root Relative Squared Error) และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ซึ่งผลการวิเคราะห์ข้อมูลได้แสดงดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของตัวแบบ

Regression technique	Regression Performance			
	Mean Squared Error	Root Mean Squared Error	Root Relative Squared Error	Coefficient of Determination
Neural Network*	189,240.917	1,375.646	0.697	67.00%
Liner Regression	2,178,519.912	1,475.981	0.732	61.00%
K-Nearest Neighbor	4,885,289.052	2,201.269	1.096	8.00%
Support Vector Machine	4,481,070.857	2,116.854	1.048	47.00%

* คือเทคนิคที่มีความเหมาะสมสำหรับการนำไปสร้างตัวแบบพยากรณ์ราคาข้าวเปลือกหอมมะลิ

จากตารางที่ 2 พบว่า เทคนิคโครงข่ายประสาทเทียม (Neural Network) เป็นเทคนิคเหมาะสมที่สุดสำหรับการสร้างตัวแบบพยากรณ์ราคาข้าวเปลือกหอมมะลิ ซึ่งทำให้ความคลาดเคลื่อนในการพยากรณ์ต่ำที่สุด โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1,892,401.917 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองเท่ากับ 1,375.646 ค่าความคลาดเคลื่อนสัมบูรณ์เท่ากับ 0.697 ค่าสัมประสิทธิ์ในการตัดสินใจเท่ากับ 67.00 %

อภิปรายผล

1. ตัวแบบที่ได้จากการศึกษาในครั้งนี้สามารถนำไปใช้สำหรับพยากรณ์ราคาข้าวเปลือกหอมมะลิ ซึ่งจะช่วยให้เกษตรกรสามารถวางแผนการผลิตข้าวได้อย่างเหมาะสม โดยสามารถวางแผนการปลูกข้าวให้สอดคล้องกับราคาข้าวเปลือกในอนาคต เช่น หากราคาข้าวเปลือกหอมมะลิมีแนวโน้มที่จะสูงขึ้น เกษตรกรอาจพิจารณาเพิ่มพื้นที่ปลูกข้าว หรือหากราคาข้าวเปลือกมีแนวโน้มที่จะลดลง เกษตรกรอาจพิจารณาลดพื้นที่ปลูกข้าวลง หรือหาพืชชนิดอื่นที่ให้ราคาดีกว่ามาปลูกทดแทน ซึ่งจะช่วยลดความเสี่ยงในการขาดทุนของเกษตรกรผู้ผลิตข้าวได้

2. การพยากรณ์ราคาข้าวเปลือกหอมมะลิช่วยให้ผู้ประกอบการสามารถวางแผนการตลาดข้าวได้อย่างเหมาะสม โดยสามารถวางแผนการจัดเก็บข้าวเปลือกหรือวางแผนการขายข้าวเปลือกให้สอดคล้องกับราคาข้าวเปลือกในอนาคต เช่น หากราคาข้าวเปลือกมีแนวโน้มที่จะสูงขึ้น ผู้ประกอบการอาจพิจารณาจัดเก็บข้าวเปลือกไว้เพื่อขายในราคาที่สูงขึ้น หรือหากราคาข้าวเปลือกมีแนวโน้มที่จะลดลง ผู้ประกอบการอาจพิจารณาเร่งขายข้าวเปลือกในราคาที่ต่ำกว่า

3. การวางแผนนโยบาย การพยากรณ์ราคาข้าวเปลือกช่วยให้หน่วยงานที่เกี่ยวข้องสามารถวางแผนนโยบายต่างๆ ที่เกี่ยวข้องกับข้าวได้อย่างเหมาะสม เช่น หากราคาข้าวเปลือกมีแนวโน้มที่จะสูงขึ้น หน่วยงานที่เกี่ยวข้องอาจ

พิจารณาออกมาตรการเพื่อควบคุมราคาข้าวเปลือก หรือหากราคาข้าวเปลือกมีแนวโน้มที่จะลดลง หน่วยงานที่เกี่ยวข้องอาจพิจารณาออกมาตรการเพื่อช่วยเหลือเกษตรกรโดยกำหนดมาตรการต่าง ๆ เพื่อรองรับปัญหาที่อาจจะเกิดขึ้นกับเกษตรกรผู้ปลูกข้าวได้ล่วงหน้า

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม ที่ให้การสนับสนุนการทำวิจัย รวมถึงคณาจารย์ในคณะฯ ที่ได้ประสิทธิ์ประสาทวิชาความรู้อันมีค่ายิ่งในการทำวิจัยในครั้งนี้ และขอขอบคุณสำนักงานเศรษฐกิจการเกษตร ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา และธนาคารแห่งประเทศไทย ที่เอื้อเฟื้อข้อมูลที่ใช้งานวิจัยนี้

บรรณานุกรม

กรมการค้าต่างประเทศ กระทรวงพาณิชย์. (2567). สถิตินำเข้า-ส่งออก. สืบค้นเมื่อวันที่ 18 กุมภาพันธ์ 2567, เข้าถึงได้จาก https://www.dft.go.th/th-th/dft-service-data-statistic/cid/604

กรมฝนหลวงและการบินเกษตร. (2564). ข้าวตาพระยา. สืบค้นเมื่อวันที่ 18 กุมภาพันธ์ 2567, เข้าถึงได้จาก https://www.royalrain.go.th/royalrain/ShowDetail.aspx?DetailId=11763

ขนิษฐา กุลนาวิน วันเพ็ญ โพธิ์เกษม และ ศันสนีย์ เลี้ยงพานิชย์ (2022) การออกแบบแบบจาลองการพยากรณ์ผลผลิตมันสาปะหลังในเขตพื้นที่ จังหวัดนครราชสีมาด้วยเทคนิคเหมืองข้อมูล, วารสารวิทยาศาสตร์และเทคโนโลยี

เฉลิมวุฒิ คำเมือง และวชิรรารักษ์ โอรสรัมย์. (2564). การพยากรณ์ราคาข้าวเปลือกหอมมะลิ ด้วยเทคนิคการทำเหมืองข้อมูล. วารสารวิทยาศาสตร์ประยุกต์, 21(1), 244210-244210.

ทรงศิริ แต้สมบัติ. (2549). การพยากรณ์เชิงปริมาณ สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร.

ธนกร สุทธิสนธ์. (2565). การประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเชิงลึกสำหรับการพยากรณ์ราคาพืชไร่ของประเทศไทย. วารสารวิชาการเทคโนโลยีอุตสาหกรรม, 18(3), 208-227.

ธนาคารแห่งประเทศไทย. (2567) อัตราแลกเปลี่ยนเงินตราต่างประเทศ. สืบค้นเมื่อวันที่ 20 กุมภาพันธ์ 2565, เข้าถึงได้จาก https://www.bot.or.th

พนิดา สมบัติมาก, ภัสสร จันทร์หอม, ศุภกร รัศมี, โอฬาร รุ่งมณีธรรมคุณ และสายชล สินสมบูรณ์ทอง. (2561)การเปรียบเทียบประสิทธิภาพในการจำแนกเมื่อข้อมูลมีค่านอกเกณฑ์ในการทำเหมืองข้อมูล.

รักบ้านเกิด. (2555). ข้าวขาวดอกมะลิ105 หรือ ข้าวหอมมะลิ105. สืบค้นเมื่อวันที่ 18 กุมภาพันธ์ 2567, เข้าถึงได้จาก https://www.rakbankerd.com/agriculture/print.php?id=4695&s=tblrice#:~:text=พื้นดินเป็นดินป,ความหอมมากยิ่งขึ้น

ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา. (2567). ข้อมูลปริมาณฝนรายเดือนรายภาคเปรียบเทียบกับค่าปกติของประเทศไทย. สืบค้นเมื่อวันที่ 20 กุมภาพันธ์ 2567, เข้าถึงได้จาก http://climate.tmd.go.th/gge/

สำนักงานเศรษฐกิจการเกษตร. (2567). ราคาสินค้าเกษตรรายเดือน. สืบค้นเมื่อวันที่ 20 กุมภาพันธ์ 2567, เข้าถึงได้จาก https://www.oae.go.th/view/1/ราคาสินค้าเกษตรรายเดือน/TH-TH

อนุพงศ์ สุขประเสริฐ. (2566) การทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio (พิมพ์ครั้งที่ 5). มหาสารคาม: คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.