



A Machine Learning–Based Market Value Estimation System for High-Value Local Chickens

Kaidtiphong Chaemchuen¹ & Pak Satanasawapak^{2*}

^{1,2} Department of Computer Engineering, Faculty of Engineering,
Rajabhat Maha Sarakham University

*Corresponding author: pak.satanasaowapak@gmail.com

Received
12/03/2026

Reviewed
25/04/2026

Revised
29/04/2026

Accepted
30/04/2026

Abstract

This study aims to develop and evaluate a market value estimation system for high-value indigenous chickens using machine learning techniques, in order to reduce valuation errors arising from subjective expert judgment. The dataset comprises 1,008 records covering key features including physical characteristics, competitive performance history, and market price ranges, all of which were preprocessed and subjected to feature selection prior to model construction. Three algorithms were compared: Random Forest, Gradient Boosting, and Extra Trees. Model performance was evaluated using 5-Fold Cross-Validation with Accuracy, Precision, Recall, and F1-score as classification metrics, alongside MAE, RMSE, and R^2 for regression analysis. Results show that Random Forest achieved the highest performance, with Accuracy of 0.958, Precision of 0.869, Recall of 0.948, and F1-score of 0.905 under a $\pm 20\%$ acceptance threshold, together with an R^2 of 0.784. These findings demonstrate the potential of artificial intelligence in valuing specialized economic animals and developing decision-support systems for the agricultural sector.

Keywords: Machine Learning; Price Estimation; Indigenous Chickens; Model Accuracy; K-Fold Cross-Validation



ระบบประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูง โดยใช้เทคนิคการเรียนรู้ของเครื่อง

เกียรติพงศ์ แซ่มชื่น¹ และภาคย์ สอนเสาวภาคย์^{2*}

^{1,2}สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยราชภัฏมหาสารคาม

ผู้นิพนธ์หลัก: pak.satanasaowapak@gmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและประเมินระบบประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูงโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อลดความคลาดเคลื่อนจากการประเมินโดยอาศัยประสบการณ์ส่วนบุคคล ข้อมูลที่ใช้ประกอบด้วยคุณลักษณะสำคัญจำนวน 1,008 รายการ ครอบคลุมลักษณะทางกายภาพ ประวัติสมรรถนะการแข่งขัน และช่วงราคาตลาด ซึ่งผ่านกระบวนการเตรียมข้อมูลและคัดเลือกคุณลักษณะก่อนสร้างแบบจำลอง งานวิจัยนี้เปรียบเทียบอัลกอริทึมสามประเภท ได้แก่ Random Forest, Gradient Boosting และ Extra Trees การประเมินประสิทธิภาพดำเนินการด้วยเทคนิค 5-Fold Cross-Validation โดยใช้ตัวชี้วัด Accuracy, Precision, Recall และ F1-score ร่วมกับการวิเคราะห์ค่าความคลาดเคลื่อน MAE, RMSE และ R^2 ผลการทดลองพบว่า Random Forest ให้ประสิทธิภาพสูงสุด โดยมีค่า Accuracy เท่ากับ 0.958 ค่า Precision เท่ากับ 0.869 ค่า Recall เท่ากับ 0.948 และ F1-score เท่ากับ 0.905 ภายใต้เกณฑ์การประเมิน $\pm 20\%$ ของราคาจริง รวมถึงค่า R^2 เท่ากับ 0.784 ผลการวิจัยแสดงให้เห็นถึงศักยภาพของการประยุกต์ใช้ปัญญาประดิษฐ์ในการประเมินมูลค่าสัตว์เศรษฐกิจเฉพาะทาง และการพัฒนาระบบสนับสนุนการตัดสินใจในภาคเกษตรกรรม

คำสำคัญ : การเรียนรู้ของเครื่อง; การประมาณราคา; ไก่พื้นเมือง; ความแม่นยำของแบบจำลอง; การตรวจสอบความถูกต้องแบบเค-โฟลด์

บทนำ

โกพื้นเมืองมูลค่าสูง โดยเฉพาะโกที่ถูกเลี้ยงดูเพื่อวัตถุประสงค์ทางวัฒนธรรมและเศรษฐกิจในชุมชนเกษตรกรรมของไทย มีมูลค่าทางการตลาดที่แตกต่างกันอย่างมีนัยสำคัญจากปศุสัตว์เชิงอุตสาหกรรมทั่วไป ราคาของสัตว์กลุ่มนี้ถูกกำหนดจากปัจจัยหลายมิติ ได้แก่ ลักษณะทางกายภาพ ประวัติ และสมรรถนะในการแข่งขัน รูปแบบพฤติกรรม และการรับรู้ของตลาด อย่างไรก็ตาม กระบวนการกำหนดราคาในปัจจุบัน ยังคงอาศัยประสบการณ์ของผู้เชี่ยวชาญและการต่อรองในตลาดเป็นหลัก ส่งผลให้การตั้งราคาขาดความเป็นมาตรฐาน ยากต่อการตรวจสอบ และก่อให้เกิดความไม่สมดุลของข้อมูลระหว่างผู้ซื้อและผู้ขาย ซึ่งเห็นได้จากความแปรปรวนของราคาในชุดข้อมูลที่รวบรวมในงานวิจัยนี้พบว่า มีช่วงราคาของโกพื้นเมืองตั้งแต่ 2,000 บาทถึง 20,000,000 บาท คิดเป็นอัตราส่วนสูงสุดต่อต่ำสุดถึง 10,000 เท่า และมีส่วนเบี่ยงเบนมาตรฐานสูงถึง 2,580,263 บาท บ่งบอกถึงความไม่แน่นอนของการกำหนดราคาที่ยังขาดมาตรฐานเชิงปริมาณ

แม้ว่าจะมีความพยายามนำข้อมูลเชิงปริมาณ เช่น อายุ น้ำหนัก และประวัติการเลี้ยงดู มาประกอบการประเมินราคา แต่ข้อมูลดังกล่าวมักถูกพิจารณาแบบแยกส่วน และขาดแบบจำลองที่สามารถอธิบายความสัมพันธ์ของตัวแปรหลายมิติได้อย่างเป็นระบบ งานทบทวนวรรณกรรมด้านการประยุกต์ใช้การเรียนรู้ของเครื่องในภาคการเกษตร ชี้ให้เห็นว่าปัญหาที่มีความซับซ้อนในลักษณะนี้ มีความเหมาะสมอย่างยิ่งต่อการประยุกต์ใช้เทคนิคดังกล่าว (Liakos et al., 2018)

ด้านการพยากรณ์ราคาสินค้าเกษตร งานวิจัยจำนวนมากยืนยันว่า เทคนิคการเรียนรู้ของเครื่องให้ความแม่นยำสูงกว่าวิธีการทางสถิติแบบดั้งเดิม โดยเฉพาะเมื่อข้อมูลมีความผันผวนและความสัมพันธ์แบบไม่เชิงเส้น (Krishna et al., 2025) แสดงให้เห็นว่าแบบจำลองในกลุ่ม boosting เช่น XGBoost สามารถพยากรณ์ราคาสินค้าเกษตรได้อย่างมีประสิทธิภาพ ขณะที่ (Zhang et al., 2025) และ (Hossain et al., 2025a) ยืนยันว่าแบบจำลองในกลุ่ม ensemble learning เช่น LightGBM สามารถลดความคลาดเคลื่อนและเพิ่มความเสถียรของการพยากรณ์ได้อย่างมีนัยสำคัญ นอกจากนี้ (Mallory et al., 2025) และ (Chuluunsai Khan et al., 2024) ชี้ให้เห็นว่าราคาสินค้าเกษตรไม่ได้ขึ้นอยู่กับปัจจัยเพียงอย่างเดียว แต่เป็นผลรวมของตัวแปรหลายด้านตลอดห่วงโซ่อุปทาน ซึ่งการเรียนรู้ของเครื่องสามารถเรียนรู้ความสัมพันธ์ของความซับซ้อนนี้ได้ดีกว่าการใช้สูตรหรือสถิติแบบเดิม

ด้านปศุสัตว์ (Rahman et al., 2021) แสดงให้เห็นว่า การใช้ข้อมูลภาพถ่ายและคุณลักษณะทางกายภาพสามารถจำแนกระดับราคาของวัวได้อย่างมีประสิทธิภาพ สะท้อนถึงศักยภาพของข้อมูลเชิงลักษณะในการประเมินมูลค่าของสัตว์เศรษฐกิจ ขณะที่งานวิจัยด้านการพยากรณ์การเจริญเติบโตของสัตว์ก็แสดงให้เห็นว่า การรวมข้อมูลหลายมิติ เช่น ข้อมูลทางกายภาพ พฤติกรรม และสิ่งแวดล้อม ช่วยเพิ่มประสิทธิภาพของแบบจำลองได้ (Hossain et al., 2025a; Oliveira et al., 2025; Lyu et al., 2023) ในส่วนของการพยากรณ์พฤติกรรมสัตว์ (Sanjuan et al., 2025) นำเสนอแบบจำลอง Random Forest และ Neural Network ที่สามารถจัดการกับข้อมูลพฤติกรรมที่ซับซ้อนและไม่เป็นเชิงเส้นได้อย่างมีประสิทธิภาพ ซึ่งมีความเกี่ยวข้องโดยตรงกับการประเมินสมรรถนะของโกพื้นเมืองมูลค่าสูง ในขณะเดียวกัน (Papadopoulos et al., 2025) มีแนวคิดในการใช้ข้อมูลจากระบบสารสนเทศในฟาร์มปศุสัตว์ เพื่อเพิ่มประสิทธิภาพในการจัดการและการตัดสินใจ โดยการบูรณาการข้อมูลจากหลายแหล่งสามารถช่วยสนับสนุนการพัฒนาระบบเกษตรอัจฉริยะได้อย่างมีประสิทธิภาพ

อย่างไรก็ตาม เมื่อพิจารณางานวิจัยที่ผ่านมา พบช่องว่างสำคัญสามประเด็น ประเด็นแรก งานส่วนใหญ่มุ่งเน้นตัวแปรเชิงปริมาณพื้นฐาน เช่น น้ำหนักและอายุ และยังขาดการบูรณาการข้อมูลเชิงพฤติกรรมและการรับรู้ของตลาดอย่างเป็นระบบ ประเด็นที่สอง แบบจำลองส่วนใหญ่ให้ผลลัพธ์เป็นค่าพยากรณ์เดียว ซึ่งไม่สามารถสะท้อนความไม่แน่นอนของตลาดได้อย่างเพียงพอ ในขณะที่การพยากรณ์เชิงช่วงราคาและเชิงความน่าจะเป็น ได้รับความสนใจมากขึ้นในงานวิจัยสมัยใหม่ (Ma et al., 2025; Rahmani et al., 2024) ประเด็นที่สาม ยังไม่มีงานวิจัยที่ประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องกับการประเมินมูลค่าของสัตว์พื้นเมืองที่มีปัจจัยกำหนดราคาเชิงวัฒนธรรม ซึ่งต่างจากปศุสัตว์เชิงอุตสาหกรรม (Hossain et al., 2025b; Sun et al., 2023) โดยเฉพาะอย่างยิ่ง ยังไม่มีงานวิจัยใดที่นำเทคนิคการเรียนรู้ของเครื่องมาประเมินมูลค่าสัตว์ ที่ราคาถูกกำหนดโดยสมรรถนะการแข่งขันเป็นหลัก ซึ่งแตกต่างจากปศุสัตว์ทั่วไปที่ราคาขึ้นอยู่กับน้ำหนักตัวหรือผลผลิตเนื้อ

ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาระบบประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูงโดยใช้เทคนิคการเรียนรู้ของเครื่อง ด้วยการบูรณาการข้อมูลสถิติ ได้แก่ ข้อมูลทางกายภาพ ข้อมูลพฤติกรรม ประวัติสมรรถนะในการแข่งขัน และข้อมูลตลาด เพื่อสร้างแบบจำลองที่สามารถอธิบายความสัมพันธ์ของปัจจัยหลายมิติได้อย่างเป็นระบบ รวมทั้งสะท้อนความไม่แน่นอนของราคาในรูปแบบช่วงพยากรณ์ที่เหมาะสม ระบบที่พัฒนาขึ้นเป็นเครื่องมือสนับสนุนการตัดสินใจสำหรับเกษตรกรและผู้มีส่วนเกี่ยวข้องในตลาด ซึ่งจะช่วยเพิ่มความโปร่งใสในการกำหนดราคา ลดความไม่สมดุลของข้อมูล และส่งเสริมการพัฒนาาระบบเกษตรอัจฉริยะในบริบทของสัตว์พื้นเมืองมูลค่าสูง

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาแบบจำลองการเรียนรู้ของเครื่องสำหรับการประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูง โดยอาศัยข้อมูลคุณลักษณะเชิงกายภาพและข้อมูลตลาดที่เกี่ยวข้อง
2. เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องในกลุ่ม Tree-based Ensemble ได้แก่ Random Forest, Gradient Boosting และ Extra Trees ภายใต้กระบวนการ 5-Fold Cross-Validation
3. เพื่อประเมินความแม่นยำและความเสถียรของแบบจำลองทั้งในมิติการพยากรณ์เชิงปริมาณ และการจำแนกช่วงราคา โดยใช้ตัวชี้วัดทางสถิติที่เหมาะสม
4. เพื่อวิเคราะห์ศักยภาพของแบบจำลองในการประยุกต์ใช้เป็นระบบสนับสนุนการตัดสินใจด้านการประเมินมูลค่า

สมมติฐานการวิจัย

งานวิจัยนี้ตั้งสมมติฐานว่า แบบจำลองการเรียนรู้ของเครื่องในกลุ่ม Tree-based Ensemble มีแนวโน้มที่จะสามารถประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูงได้อย่างมีประสิทธิภาพ ภายใต้ข้อมูลเชิงตารางที่มีความสัมพันธ์ไม่เชิงเส้นและมีความผันผวนของราคา โดยคาดว่าแบบจำลองมีแนวโน้มที่จะให้ค่าความคลาดเคลื่อนเชิงปริมาณ MAE และ RMSE อยู่ในระดับต่ำ และมีความสามารถในการอธิบายความแปรปรวนของข้อมูล (R^2) ในระดับที่เหมาะสมต่อการประยุกต์ใช้งานจริง นอกจากนี้ งานวิจัยนี้ตั้งสมมติฐานว่า อัลกอริทึม Random Forest มีแนวโน้มที่จะให้ประสิทธิภาพสูงกว่าอัลกอริทึมอื่นภายใต้เงื่อนไขของชุดข้อมูลที่ศึกษา ทั้งในด้านความแม่นยำและความเสถียรภายใต้กระบวนการ 5-Fold Cross-Validation นอกจากนี้ งานวิจัยนี้ตั้งสมมติฐานว่า การกำหนดเกณฑ์การประเมินภายใต้ช่วงความคลาดเคลื่อนไม่เกิน $\pm 20\%$ มีแนวโน้มที่จะสะท้อนความเหมาะสมในการนำไปใช้จริงของแบบจำลอง สำหรับการพัฒนาเป็นระบบสนับสนุนการตัดสินใจด้านการประเมินมูลค่าในภาคเกษตรกรรม

ขอบเขตการวิจัย

งานวิจัยนี้มุ่งพัฒนาและประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องสำหรับการประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูง โดยใช้ข้อมูลเชิงตารางที่ประกอบด้วยคุณลักษณะทางกายภาพ และข้อมูลตลาดที่สามารถวัดเชิงปริมาณได้ การศึกษาเปรียบเทียบอัลกอริทึมในกลุ่ม Tree-based Ensemble ได้แก่ Random Forest, Gradient Boosting และ Extra Trees ภายใต้กระบวนการ 5-Fold Cross-Validation และประเมินผลทั้ง Regression และ Classification ภายใต้เกณฑ์ความคลาดเคลื่อนไม่เกิน $\pm 20\%$

ระเบียบวิธีวิจัย

งานวิจัยนี้ออกแบบกระบวนการวิจัยเชิงทดลองที่มีโครงสร้างชัดเจน ครอบคลุมตั้งแต่การจัดเตรียมข้อมูล การสร้างแบบจำลอง การตรวจสอบความถูกต้องเชิงสถิติ และการวิเคราะห์เปรียบเทียบผลลัพธ์เชิงลึก โดยมีกระบวนการวิจัยประกอบด้วย 6 ขั้นตอนหลัก ได้แก่ การรวบรวมข้อมูล การเตรียมและปรับปรุงข้อมูล การพัฒนาแบบจำลองการเรียนรู้ของเครื่อง การประเมินประสิทธิภาพของแบบจำลอง การพัฒนาเว็บแอปพลิเคชัน การวิเคราะห์ผลลัพธ์และการเปรียบเทียบอัลกอริทึม สำหรับรายละเอียดของแต่ละขั้นตอนดังนี้

ผู้วิจัยได้ออกแบบสถาปัตยกรรมระบบ โดยการนำชุดข้อมูลจากตลาดซื้อขายที่เกิดขึ้นจริงมาเข้าสู่กระบวนการจัดเตรียมข้อมูล และเรียนรู้ข้อมูลด้วยการเรียนรู้ของเครื่อง ผู้วิจัยได้เขียน Python เพื่อทำงานในส่วนของ Backend สำหรับเชื่อมต่อกับโมเดลการเรียนรู้ และสร้าง Web UI เมื่อผู้ใช้กรอกข้อมูลคุณลักษณะของไก่พื้นเมือง ระบบจะประมาณราคาที่เหมาะสมสำหรับไก่พื้นเมืองที่มีคุณลักษณะดังกล่าว ภาพรวมการทำงานของระบบแสดงดัง Figure 1

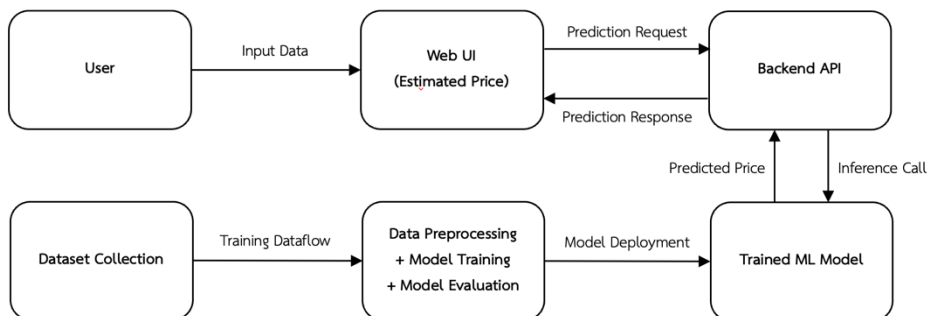


Figure 1. Architecture of the proposed web-based machine learning price estimation system



การรวบรวมและเตรียมข้อมูล

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลเชิงตาราง ประกอบด้วยคุณลักษณะของสัตว์พื้นเมืองมูลค่าสูงที่เกี่ยวข้องกับลักษณะทางกายภาพ ประวัติการแข่งขัน และมูลค่าทางเศรษฐกิจ เช่น อายุของไก่ จำนวนครั้งที่ชนะ ยกที่ชนะล่าสุด นาที่ที่ชนะล่าสุด มูลค่ารางวัลล่าสุดที่ได้รับ สไตล์การต่อสู้ เทคนิคการต่อสู้ สีของขน สีของขา ลักษณะของหาง น้ำหนัก ลักษณะความสูง ค่าตัวที่มีการซื้อขายจริง เป็นต้น รายละเอียดของแต่ละคุณลักษณะ แสดงดัง Table 1

Table 1. Description of input features and target variable used for machine learning based price estimation

ข้อมูล	ตัวอย่างรายละเอียดของข้อมูล
อายุของไก่ (วัน)	อายุของไก่ในหน่วยวัน แปลงจากเดือนและวัน เพื่อใช้เป็นตัวแปรเชิงปริมาณต่อเนื่อง
ครั้งที่ชนะ	จำนวนครั้งที่ไก่ชนะการแข่งขันทั้งหมด ใช้สะท้อนประสบการณ์และสมรรถนะ
ยกที่ชนะล่าสุด	ยกที่ไก่ชนะในการแข่งขันครั้งล่าสุด แสดงความได้เปรียบเชิงกลยุทธ์
นาที่ที่ชนะล่าสุด	ระยะเวลาที่ใช้ในการชนะการแข่งขันครั้งล่าสุด
มูลค่ารางวัลล่าสุดที่ได้รับ	จำนวนเงินรางวัลจากการแข่งขันครั้งล่าสุด
สไตล์การต่อสู้	ลักษณะการต่อสู้ เช่น เจริญรุก เจริญรับ หรือผสม
เทคนิคการต่อสู้	เทคนิคเด่นในการต่อสู้ เช่น ตีหัว ตีลำตัว หรือเข้าทำเร็ว
สีของขน	ของขนไก่ เช่น แดง เหลือง ดำ
สีของขา	สีของขา เช่น เหลือง เขียว ดำ
ลักษณะของหาง	รูปแบบของหาง เช่น หางยาว หางสั้น หางชี้
น้ำหนัก	น้ำหนักตัวไก่ (กิโลกรัม) ใช้เป็นตัวแปรเชิงปริมาณ
ลักษณะความสูง	ระดับความสูงของลำตัว เช่น เตี้ย ปานกลาง สูง
ค่าตัวที่มีการซื้อขายจริง	ราคาซื้อขายจริงในตลาด ใช้เป็นตัวแปรเป้าหมาย

ชุดข้อมูลเหล่านี้ ผู้วิจัยได้รวบรวมมาจากแหล่งข้อมูลต่าง ๆ ได้แก่ การซื้อขายจริงตามตลาด การซื้อขายจริงบนสื่อสังคมออนไลน์ กลุ่มไลน์ซื้อขาย เป็นต้น ซึ่งได้รวบรวมข้อมูล จำนวน 1,008 ข้อมูล โดยเป็นข้อมูลที่มีการซื้อขายสำเร็จ ข้อมูลทั้งหมดถูกจัดเก็บในรูปแบบตารางและผ่านกระบวนการตรวจสอบความถูกต้องก่อนเข้าสู่กระบวนการสร้างแบบจำลอง โดยการดำเนินการดังนี้ 1) ตรวจสอบค่าที่หายไป 2) ตรวจสอบความผิดปกติของข้อมูล 3) แปลงข้อมูลเชิงหมวดหมู่ด้วยวิธี One-Hot Encoding 4) แปลงอายุให้อยู่ในรูปแบบวันเพื่อให้เป็นตัวแปรเชิงปริมาณที่ต่อเนื่อง 5) ตรวจสอบการกระจายของตัวแปรเป้าหมาย และ 6) กำหนดตัวแปรตาม คือ ราคาซื้อขายจริงในหน่วยบาท

เพื่อให้เห็นโครงสร้างของข้อมูลที่ใช้ในการวิจัย ตัวอย่างชุดข้อมูลบางส่วนแสดงดัง Table 2 ซึ่งประกอบด้วยคุณลักษณะเชิงปริมาณและเชิงคุณภาพที่เกี่ยวข้องกับมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูง

**Table 2.** Sample dataset used in this study for estimating the market value of high-value indigenous chickens, consisting of quantitative and qualitative features related to performance and physical characteristics

อายุ (วัน)	ครั้งที่ ชนะ	ยกที่ ชนะ ล่าสุด	นาที่ที่ ชนะ ล่าสุด	มูลค่ารางวัล ล่าสุด ที่ได้รับ	สไตล์การ ต่อสู้	เทคนิคการ ต่อสู้	สีของ ขน	สีของ ขา	ลักษณะ ของหาง	น้ำหนัก (กก.)	ลักษณะ ความสูง	ค่าตัวที่มี การซื้อขายจริง
435	3	2	19	1.1M	เดินถอย	แข่งหน้าจัด บ้าดี	ส้มดำ	เทา	สั้น	2.7	สูง	2M
408	3	1	15	220k	ลอยยาว	ตีหู บ้าดี	เหลือง ดำ	เหลือง	ยาว	2.5	เตี้ย	300k
425	4	2	17	55k	เดินถอย	หงอนได้ กางได้	เทา	เทา	ยาว	2.7	สูง	60k
309	2	1	15	4.4k	เดิน	หน้าคาง	เทา	ขาว	สั้น	2.5	สูง	7k

การแบ่งชุดข้อมูล

เพื่อประเมินความสามารถของแบบจำลอง งานวิจัยนี้ใช้วิธี 5-Fold Cross-Validation ซึ่งเป็นเทคนิคมาตรฐานในการลดความแปรปรวนของผลลัพธ์จากการแบ่งข้อมูล โดยมีขั้นตอนดังนี้ 1) แบ่งข้อมูลออกเป็นห้าส่วนเท่า ๆ กัน 2) ใช้สี่ส่วนสำหรับฝึกโมเดล 3) ใช้หนึ่งส่วนสำหรับทดสอบ 4) ทำซ้ำจนครบห้ารอบ และ 5) คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของผลลัพธ์ ซึ่งการใช้ Cross-Validation ทำให้ช่วยลดการเกิดอคติจากการแบ่งข้อมูลเพียงครั้งเดียว และเพิ่มความน่าเชื่อถือของผลการทดลอง

กำหนดให้ D แทนชุดข้อมูลทั้งหมด โดยแบ่งออกเป็น 5 ส่วนเท่า ๆ กัน คือ $D_1, D_2, \dots, D_5$ ในการวนรอบที่ i ให้ D_i เป็น Testing Set และใช้ส่วนที่เหลือเป็น Training Set ทำซ้ำจนครบ 5 รอบ และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวชี้วัดทั้งหมด

อย่างไรก็ตาม งานวิจัยนี้ไม่ได้แยกชุดข้อมูลทดสอบอิสระ (Hold-out Test Set) ออกจากกระบวนการ 5-Fold Cross-Validation เนื่องจากชุดข้อมูลที่ใช้มีจำนวนจำกัดที่ 1,008 ข้อมูล ซึ่งเป็นการยากหาของข้อมูลการซื้อขายไก่พื้นเมืองมูลค่าสูง ซึ่งถือเป็นข้อจำกัดของงานวิจัยนี้ ทั้งนี้ กระบวนการ 5-Fold Cross-Validation ที่ใช้ มีคุณสมบัติเทียบเท่ากับการทำ Random Split Test Set ซ้ำถึง 5 รอบ โดยทุกรายการในชุดข้อมูลถูกใช้เป็นชุดทดสอบ 1 ครั้ง ซึ่งให้การประเมินที่ครอบคลุมและน่าเชื่อถือกว่าการแบ่ง Hold-out Test Set เพียงครั้งเดียว

อัลกอริทึมที่ใช้ในการวิจัย

งานวิจัยนี้เลือกใช้อัลกอริทึมประเภท Ensemble Learning จำนวน 3 วิธี ได้แก่ 1) Random Forest 2) Gradient Boosting และ 3) Extra Trees โดยมีเหตุผลในการเลือกอัลกอริทึม ดังนี้

อัลกอริทึมกลุ่ม Tree-based Ensemble มีคุณสมบัติที่เหมาะสมกับลักษณะข้อมูลในงานวิจัยนี้หลายประการ ได้แก่ 1) รองรับข้อมูลเชิงตาราง ที่ผสมระหว่างตัวแปรเชิงปริมาณและตัวแปรเชิงหมวดหมู่ได้โดยตรง โดยไม่จำเป็นต้องทำ Feature Scaling ซึ่งลดความเสี่ยงจากการปรับแต่งข้อมูลที่ไม่เหมาะสม 2) สามารถจัดการความสัมพันธ์แบบไม่เชิงเส้น ระหว่างคุณลักษณะของไก่กับมูลค่าตลาดได้ดี 3) มีความทนทานต่อ Noise และ Outlier ซึ่งเป็นลักษณะที่พบได้บ่อยในข้อมูลราคาตลาดที่มีความ

ผันผวนเชิงพฤติกรรม และ 4) สามารถคำนวณค่า Feature Importance ได้โดยตรง ทำให้สามารถวิเคราะห์ปัจจัยที่มีอิทธิพลต่อการกำหนดราคาได้อย่างโปร่งใส

ชุดข้อมูลในงานวิจัยนี้มีลักษณะเฉพาะที่สนับสนุนการเลือกอัลกอริทึมกลุ่มนี้ กล่าวคือ ข้อมูลประกอบด้วยตัวแปรที่มีธรรมชาติแตกต่างกัน ทั้งตัวแปรเชิงปริมาณ เช่น อายุ น้ำหนัก จำนวนครั้งที่ชนะ และตัวแปรเชิงหมวดหมู่ เช่น สโมสรการต่อสู้ สีของขน ซึ่งหลังจากแปลงด้วย One-Hot Encoding แล้ว อัลกอริทึมในกลุ่มนี้สามารถเรียนรู้รูปแบบจากตัวแปรดังกล่าวได้อย่างมีประสิทธิภาพ นอกจากนี้ ยังมีงานวิจัยที่ยืนยันประสิทธิภาพของอัลกอริทึมเหล่านี้ในบริบทที่ใกล้เคียง เช่น Sanjuan et al. (2025) พบว่า Random Forest ให้ความสมมูลระหว่างความแม่นยำ RMSE เท่ากับ 14.97 และความสามารถในการอธิบายผลได้ดีกว่า Neural Networks ซึ่งแม้จะแม่นยำกว่าเล็กน้อยแต่ตีความได้ยากกว่า จึงเหมาะสมกว่าสำหรับการนำไปใช้จริงในระบบสนับสนุนการตัดสินใจในฟาร์มปศุสัตว์

การเปรียบเทียบอัลกอริทึมทั้งสามในงานวิจัยนี้มีเหตุผลเพิ่มเติม กล่าวคือ Random Forest ใช้หลักการ Bagging ซึ่งเน้นลด Variance และมีความทนทานต่อ Outlier ส่วน Gradient Boosting ใช้หลักการ Boosting ซึ่งเน้นลด Bias ผ่านการเรียนรู้แบบลำดับขั้น และ Extra Trees เป็นรูปแบบที่สุ่มจุดแบ่งของต้นไม้อย่างสมบูรณ์เพื่อลด Variance ให้มากกว่า Random Forest การเปรียบเทียบอัลกอริทึมทั้งสามจึงครอบคลุมแนวทาง Ensemble ที่แตกต่างกัน และช่วยให้สามารถระบุได้ว่าแนวทางใดเหมาะสมที่สุดกับบริบทของข้อมูลไก่พื้นเมืองมูลค่าสูง สำหรับรายละเอียดของแต่ละอัลกอริทึมมีดังต่อไปนี้

Random Forest

เป็นอัลกอริทึมประเภท Bagging ที่สร้างต้นไม้ตัดสินใจหลายต้นจากการสุ่มตัวอย่างแบบ Bootstrap และสุ่มเลือกคุณลักษณะในแต่ละโหนดเพื่อลดความสัมพันธ์ระหว่างต้นไม้ สำหรับงาน Regression ผลลัพธ์สุดท้ายคำนวณจากค่าเฉลี่ยของต้นไม้ทั้งหมด จากสูตร

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x)$$

โดยที่

T คือ จำนวนต้นไม้

$f_t(x)$ คือ ผลลัพธ์จากต้นไม้ลำดับที่ t



Gradient Boosting

เป็นอัลกอริทึมประเภท Boosting ที่พัฒนาแบบจำลองเชิงลำดับ โดยแต่ละขั้นตอนจะเรียนรู้ค่าความคลาดเคลื่อน ของแบบจำลองก่อนหน้าเพื่อลดข้อผิดพลาดรวมของระบบ ส่งผลให้ค่าฟังก์ชันสูญเสียลดลงอย่างต่อเนื่อง อัลกอริทึมนี้เหมาะสำหรับข้อมูลที่มีรูปแบบซับซ้อน แต่ต้องควบคุมพารามิเตอร์เพื่อป้องกัน Overfitting สามารถนิยามแบบจำลองในรูปสมการได้ ดังนี้

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x)$$

โดยที่

η คือ Learning Rate

$h_m(x)$ คือ โมเดลที่เรียนรู้ Residual

Extra Trees

เป็นอัลกอริทึมที่เพิ่มระดับความสุ่มในกระบวนการแบ่งโหนด โดยทำการสุ่มค่า threshold แทนการคัดเลือกค่าที่เหมาะสมที่สุดตามหลัก Greedy Optimization แนวทางนี้ช่วยลดความแปรปรวนของแบบจำลอง และเพิ่มประสิทธิภาพเชิงคำนวณในกระบวนการฝึกสอน การสุ่ม threshold ยังมีบทบาทสำคัญในการลดความสัมพันธ์เชิงสหสัมพันธ์ระหว่างต้นไม้ภายใน Ensemble ส่งผลให้แบบจำลองมีเสถียรภาพสูงขึ้น และลดความเสี่ยงของการเกิด Overfitting เมื่อประยุกต์ใช้กับข้อมูลที่มีความซับซ้อนและความไม่เชิงเส้นสูง

เกณฑ์การประเมินประสิทธิภาพ

งานวิจัยนี้แบ่งการประเมินแบ่งออกเป็น 2 แบบ ได้แก่ Regression และ Classification โดยมีรายละเอียดดังต่อไปนี้

การประเมินเชิง Regression

ในการประเมินเชิง Regression ใช้ตัวชี้วัด 3 ค่า ได้แก่ 1) Mean Absolute Error (MAE) 2) Root Mean Squared Error (RMSE) และ 3) Coefficient of Determination (R^2) โดยแต่ละตัวชี้วัดมีสมการ ดังนี้

1) Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$



2) Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

3) Coefficient of Determination (R^2)

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

โดยที่

y_i คือ ราคาจริง

$\hat{y}_i$ คือ ราคาที่โมเดลพยากรณ์

$\bar{y}$ คือ ค่าเฉลี่ยของราคาจริง

การประเมินเชิง Classification ภายได้เกณฑ์ $\pm 20\%$

เกณฑ์ $\pm 20\%$ ในงานวิจัยนี้เป็นการกำหนดโดยอิงจากบริบทเชิงปฏิบัติของตลาดไก่พื้นเมืองมูลค่าสูง โดยพิจารณาจาก 3 ปัจจัย ได้แก่

1) ด้านลักษณะข้อมูล ชุดข้อมูลมีค่า Coefficient of Variation สูงถึง 218.6% (SD เท่ากับ 2,580,263 บาท และ Mean เท่ากับ 1,180,245 บาท) แสดงให้เห็นว่า ราคามีความผันผวนสูงมาก หากกำหนดเกณฑ์ที่เข้มงวดเกินไป เช่น $\pm 5\%$ หรือ $\pm 10\%$ โมเดลจะผ่านได้น้อยมาก ตัวอย่างเช่น ไก่ที่ราคาจริง 100,000 บาท โมเดลต้องประเมินได้ในช่วง 95,000–105,000 บาทเท่านั้น จึงจะถือว่าถูกต้อง ซึ่งเป็นเกณฑ์ที่ยากเกินไปสำหรับข้อมูลที่มีความแปรปรวนสูงในระดับนี้ ในทางตรงกันข้าม หากกำหนดเกณฑ์ที่หลวมเกินไป เช่น $\pm 30\%$ หรือสูงกว่า โมเดลจะผ่านได้แทบทุกกรณี ทำให้ผลการประเมินขาดความหมายในเชิงปฏิบัติ ดังนั้น $\pm 20\%$ จึงเป็นจุดสมดุลที่เหมาะสมสำหรับบริบทนี้

2) ด้านความเป็นจริงของตลาด ราคาไก่พื้นเมืองขึ้นอยู่กับปัจจัยที่วัดไม่ได้เชิงปริมาณ เช่น ความนิยมส่วนตัว อารมณ์การต่อรอง และบริบทของวันแข่งขัน ความคลาดเคลื่อน $\pm 20\%$ จึงสะท้อนขีดจำกัดที่สมเหตุสมผลของโมเดลการเรียนรู้ของเครื่อง

3) ด้านการใช้งานจริง หากโมเดลประเมินราคาไก่ที่มีมูลค่าจริง 100,000 บาท อยู่ในช่วง 80,000 ถึง 120,000 บาท ถือว่าเพียงพอสำหรับการใช้เป็นแนวทางตัดสินใจซื้อขาย ซึ่งดีกว่าการประเมินโดยผู้เชี่ยวชาญที่ไม่มีข้อมูลเชิงปริมาณรองรับเลย

ในการการประเมินเชิง Classification งานวิจัยนี้ใช้ตัวชี้วัด Accuracy, Precision, Recall และ F1-score รวมไปถึงวิเคราะห์ Confusion Matrix ทั้งแบบ 2-class คือ Correct กับ Incorrect และแบบ 3-class ได้แก่ Low, Medium, High

การวิเคราะห์ความสำคัญของตัวแปร

นอกเหนือจากการประเมินประสิทธิภาพของแบบจำลองด้วยตัวชี้วัดเชิงตัวเลข งานวิจัยนี้ดำเนินการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance) เพื่อระบุว่าตัวแปรใดมีอิทธิพลสูงสุดต่อการพยากรณ์มูลค่าตลาดของโกพื้นเมืองมูลค่าสูง การวิเคราะห์นี้ใช้ค่า Feature Importance ที่คำนวณจากอัลกอริทึม Tree-based ทั้ง 3 ตัว โดยอาศัยหลักการ Mean Decrease in Impurity (MDI) ซึ่งวัดว่าตัวแปรแต่ละตัวมีส่วนลดความไม่บริสุทธิ์ในโหนดของต้นไม้ตัดสินใจ โดยเฉลี่ยยังมีค่าสูง แสดงว่าตัวแปรนั้นมีบทบาทสำคัญในการแบ่งแยกกลุ่มราคามากกว่า

การจำแนกระดับราคา

เพื่อวิเคราะห์พฤติกรรมของการจำแนกของโมเดลเชิงลึกเพิ่มเติม งานวิจัยได้แบ่งช่วงราคาจากการเรียงลำดับข้อมูลจริงทั้ง 1,008 ข้อมูล ออกเป็น 3 ระดับ โดยใช้ค่า Quartile ของข้อมูลจริง ได้แก่ $Low \leq Q1$, $Medium Q1 < ราคา \leq Q3$ และ $High > Q3$ โดย $Q1$ เท่ากับ 25,000 บาท หมายความว่า ใกล้เคียง 25% ของชุดข้อมูล มีราคาต่ำกว่า 25,000 บาท และ $Q3$ เท่ากับ 850,000 บาท หมายความว่า ใกล้เคียง 75% ของชุดข้อมูล มีราคาต่ำกว่า 850,000 บาท การจำแนกนี้ช่วยให้สามารถวิเคราะห์ความสามารถของโมเดลในการแยกระดับราคาต่ำ ราคากลาง และราคาสูงได้อย่างชัดเจน

Figure 2 แสดง distribution ของมูลค่าตลาดโกพื้นเมืองในรูปแบบ Histogram (ซ้าย) และ Boxplot จำแนกตามกลุ่มราคา (ขวา) โดยแกน X และแกน Y แสดงในรูปแบบ Log Scale เพื่อให้เห็นการกระจายตัวของข้อมูลในทุกช่วงราคาได้อย่างชัดเจน จาก Histogram พบว่า ข้อมูลมีการกระจายแบบ Right-skewed กล่าวคือ ข้อมูลส่วนใหญ่กระจุกตัวอยู่ในช่วงราคาต่ำกว่า $Q1$ เท่ากับ 25,000 บาท และ $Q3$ เท่ากับ 850,000 บาท โดยมีข้อมูลจำนวนน้อยที่มีราคาสูงถึง 20,000,000 บาท ลักษณะ Long Tail ทางขวานี้ แสดงให้เห็นว่าค่าเฉลี่ยเท่ากับ 1,180,245 บาท สูงกว่าค่ากลาง 77,000 บาท ซึ่งเป็นผลจากโกพื้นเมืองมูลค่าสูงที่มีราคาเป็น Outlier ในตลาด ดังนั้นผู้วิจัยจึงเลือกใช้ Quartile เป็นเกณฑ์แบ่งกลุ่มราคา แทนการแบ่งแบบ Equal Interval เนื่องจาก Quartile มีความทนทานต่อ Outlier ได้ดีกว่า

จาก Boxplot พบว่า ทั้งสามกลุ่มมีลักษณะการกระจายที่แตกต่างกัน กลุ่ม Low มีจำนวน 278 ข้อมูล มีช่วงราคา 2,000 ถึง 25,000 บาท โดยมีการกระจายตัวน้อย และค่ากลางอยู่ในระดับต่ำ กลุ่ม Medium มีจำนวน 481 ข้อมูล มีช่วงราคา 25,001 ถึง 850,000 บาท และมีการกระจายตัวกว้างที่สุดในสามกลุ่ม จากความหลากหลายของโกในระดัปล่าง ขณะที่กลุ่ม High จำนวน 249 ข้อมูล มีช่วงราคาตั้งแต่ 850,001 บาทขึ้นไปจนถึง 20,000,000 บาท โดยมี Outlier จำนวนมากในช่วงราคาสูง ซึ่งสะท้อนถึงความหายากและคุณค่าพิเศษของโกพื้นเมืองมูลค่าสูงในตลาด

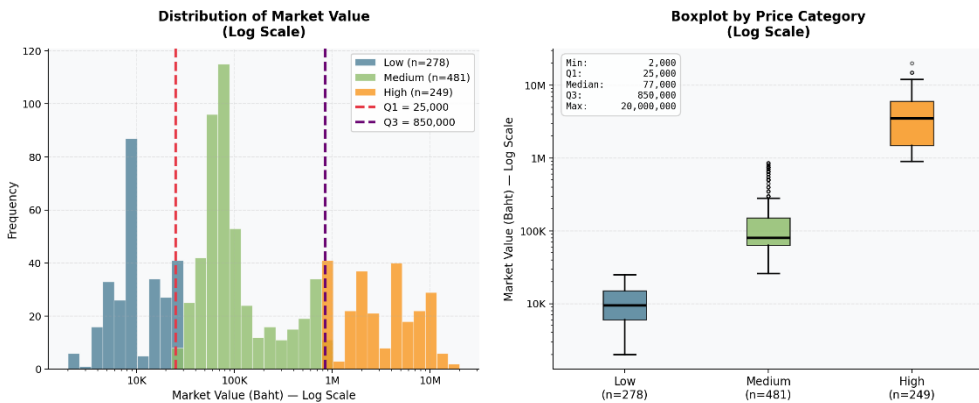


Figure 2. Distribution of market values for high-value indigenous chickens, illustrated by a Histogram (left) and a Boxplot classified by price category: Low, Medium, and High (right). Both charts are displayed on a Log Scale. The red and purple dashed lines indicate Q1 = 25,000 THB and Q3 = 850,000 THB, respectively, which serve as the price category boundaries

การพัฒนาเว็บแอปพลิเคชัน

ระบบที่พัฒนาขึ้นถูกออกแบบและนำไปใช้งานในรูปแบบ Web-based Application ดัง Figure 3 เพื่อให้สามารถประยุกต์ใช้แบบจำลองการเรียนรู้ของเครื่องในสภาพแวดล้อมจริงได้อย่างเป็นรูปธรรม ส่วนติดต่อผู้ใช้ ถูกออกแบบให้เป็นแบบฟอร์มกรอกข้อมูลเชิงโครงสร้าง โดยสอดคล้องกับคุณลักษณะ ที่ใช้ในการฝึกแบบจำลอง เช่น อายุ จำนวนครั้งที่ชนะ ผลการแข่งขันล่าสุด ลักษณะทางกายภาพ และปัจจัยด้านตลาด ทั้งนี้เพื่อให้กระบวนการทำนายสอดคล้องกับขั้นตอนการเตรียมข้อมูลในระยยะฝึกแบบจำลอง เมื่อผู้ใช้กรอกข้อมูลและส่งคำขอประเมิน ระบบจะส่งข้อมูลไปยัง Backend API ผ่านกลไก Prediction Request โดยฝั่งเซิร์ฟเวอร์จะดำเนินการประมวลผลข้อมูลล่วงหน้าในรูปแบบเดียวกับกระบวนการฝึกแบบจำลอง และเรียกใช้โมเดลที่ผ่านการฝึกแล้ว เพื่อทำการทำนายมูลค่าทางการตลาด จากนั้นผลลัพธ์การประมาณราคาและค่าความเชื่อมั่นของโมเดลจะถูกส่งกลับไปยังส่วนติดต่อผู้ใช้แบบเรียลไทม์ สถาปัตยกรรมดังกล่าวช่วยให้เกิดความสอดคล้องระหว่างกระบวนการฝึกและการใช้งานจริง การพัฒนาในรูปแบบ Web-based ทำให้แบบจำลองเชิงทฤษฎีสามารถแปลงสู่ระบบสนับสนุนการตัดสินใจได้จริง เหมาะสำหรับเกษตรกร ผู้เพาะเลี้ยง และผู้มีส่วนเกี่ยวข้องในตลาดไก่พื้นเมืองมูลค่าสูง

 **ระบบประมาณการมูลค่าทางการตลาดของไก่พื้นเมืองมูลค่าสูง**

กรอกข้อมูลเพื่อประเมินราคาซื้อ-ขายอย่างเป็นธรรม

อายุ (เดือน)	อายุ (วัน)	จำนวนครั้งหั้นะ
13	6	5
ยก	นาที่	เงินรางวัลล่าสุด
2	6	200000.0
สไตล์การต่อสู้	เทคนิคการต่อสู้	ตัวสี
หมา	বাদ	ดำ
ตำแหน่ง	ลักษณะนาง	น้ำหนัก (กก.)
ดำ	ขาว	2.6
ช่วงตัว		
สูง		

คำนวณราคากลาง

 ราคากลางประมาณการ: 405,247 บาท
 ความมั่นใจของ AI: 84%

Figure 3. Web-based interface of the proposed machine learning system for high-value local chicken price estimation

ผลการวิจัย

การประเมินประสิทธิภาพของแบบจำลอง ดำเนินการภายใต้การตรวจสอบความถูกต้องแบบ 5-Fold Cross-Validation เพื่อประเมินความสามารถของแบบจำลองอย่างเป็นระบบ ลดอคติและความแปรปรวนที่อาจเกิดจากการแบ่งชุดข้อมูลเพียงครั้งเดียว การวิเคราะห์ผลการทดลองแบ่งออกเป็นสองแบบ ได้แก่ 1) การประเมินความแม่นยำเชิงปริมาณของการพยากรณ์ โดยใช้ตัวชี้วัดด้านความคลาดเคลื่อน และ 2) การประเมินความแม่นยำเชิงการตัดสินใจภายใต้เงื่อนไขค่าคลาดเคลื่อนยอมรับได้ $\pm 20\%$ ซึ่งสะท้อนประสิทธิภาพของระบบในบริบทการใช้งานจริง ผลการวิจัยมีรายละเอียดดังต่อไปนี้

การวิเคราะห์ผลราย Fold

Table 3 แสดงค่า Accuracy ของแต่ละอัลกอริทึมในแต่ละ fold ผลการทดลองพบว่าแบบจำลอง Random Forest ให้ค่า Accuracy เฉลี่ยสูงสุดที่ 0.958 รองลงมาคือ Gradient Boosting ที่ 0.953 และ Extra Trees ที่ 0.949 แสดงให้เห็นว่า ประสิทธิภาพโดยรวมที่อยู่ในระดับสูงของทั้งสามอัลกอริทึม เมื่อพิจารณาในราย Fold พบว่า Random Forest มีค่าความแม่นยำสูงและค่อนข้างสม่ำเสมอในทุก Fold โดยเฉพาะใน Fold 5 ที่ให้ค่า Accuracy สูงสุดที่ 0.985 แสดงถึงศักยภาพในการเรียนรู้โครงสร้างข้อมูลได้อย่างมีประสิทธิภาพ ขณะที่ Gradient Boosting และ Extra Trees แสดงผลลัพธ์ที่ใกล้เคียงกันในหลาย Fold โดย Gradient Boosting มีค่าอยู่ในช่วง 0.941–0.970 และ Extra Trees อยู่ในช่วง 0.931–0.975 ซึ่งทั้งสองบ่งชี้ถึงเสถียรภาพของแบบจำลองภายใต้การเปลี่ยนแปลงของชุดข้อมูลฝึกและทดสอบ

ช่วงความแตกต่างของค่า Accuracy ระหว่าง Fold สูงสุดและต่ำสุดของแต่ละโมเดลอยู่ในระดับที่ไม่สูงมาก (RF: 0.049, GB: 0.029, ET: 0.044) และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ในช่วงแคบ (RF: ± 0.018 , GB: ± 0.012 , ET: ± 0.018) แสดงให้เห็นว่า ประสิทธิภาพของโมเดลมีความสม่ำเสมอในทุก Fold โดยไม่มี Fold ใดที่ให้ผลดีหรือแย่กว่า Fold อื่นอย่างมีนัยสำคัญ หากโมเดลเกิด Overfitting ค่า Accuracy จะแสดงความผันผวนสูงระหว่าง Fold เช่น สูงในบาง Fold และต่ำในอีก Fold อย่างชัดเจน แต่ผลการทดลองที่ได้ ไม่แสดงรูปแบบดังกล่าว

ผลการทดลองแสดงให้เห็นว่า Random Forest มีความสามารถในการเรียนรู้รูปแบบข้อมูลที่ซับซ้อนได้ดีกว่าโมเดลอื่นอย่างสม่ำเสมอ ซึ่งสอดคล้องกับกลไกของ Random Forest ที่สร้าง Decision Tree จำนวนมาก โดยใช้การสุ่มทั้งข้อมูลและตัวแปรทำให้แต่ละ Tree เรียนรู้รูปแบบที่หลากหลายและลด Variance ของการพยากรณ์โดยรวม ในขณะที่ Gradient Boosting แม้จะมีค่า S.D. ต่ำที่สุด (± 0.012) แต่ให้ค่า Accuracy เฉลี่ยต่ำกว่า Random Forest เนื่องจากกลไก Sequential Learning อาจมีความไวต่อโครงสร้างข้อมูลในแต่ละ Fold มากกว่า ส่วน Extra Trees แม้จะให้ความหลากหลายสูงจากการสุ่ม Threshold แบบสมบูรณ์ แต่ความสุ่มที่มากเกินไปอาจส่งผลให้ค่าเฉลี่ยต่ำกว่า Random Forest เล็กน้อย ผลลัพธ์แสดงให้เห็นว่า กลไก Bagging ของ Random Forest มีความเหมาะสมกับโครงสร้างข้อมูลราคาไก่พื้นเมืองที่มีความผันผวนสูงและมี Outlier จำนวนมากได้ดีที่สุดเมื่อเทียบกับอัลกอริทึมอื่น ๆ ที่นำมาทดสอบ

Table 3. Classification accuracy (within $\pm 20\%$ error threshold) of each model across five folds

Fold	Random Forest	Gradient Boosting	Extra Trees
Fold 1	0.960	0.950	0.931
Fold 2	0.960	0.960	0.960
Fold 3	0.936	0.941	0.936
Fold 4	0.951	0.945	0.945
Fold 5	0.985	0.970	0.975
Average	0.958	0.953	0.949

ผลการประเมินเชิง Regression

จาก Table 4 แสดงถึงผลการทดลอง ซึ่งพบว่า Random Forest ให้ค่าความคลาดเคลื่อนต่ำที่สุดทั้งในแง่ MAE และ RMSE สะท้อนถึงความสามารถในการลดความคลาดเคลื่อนสัมบูรณ์และความคลาดเคลื่อนกำลังสองได้ดีกว่าโมเดลอื่น ค่า R^2 ของ Random Forest เท่ากับ 0.784 แสดงให้เห็นว่าแบบจำลองสามารถอธิบายความแปรปรวนของราคาจริงได้ประมาณ 78.4% ซึ่งถือว่าอยู่ในระดับสูงเมื่อพิจารณาถึงลักษณะข้อมูลตลาดที่มีความผันผวนเชิงพฤติกรรมและปัจจัยเชิงคุณภาพ ที่ไม่สามารถวัดได้ทั้งหมด ผลลัพธ์ดังกล่าวสะท้อนว่าอัลกอริทึม Random Forest มีความทนทานต่อ Noise และ Outlier มากกว่าแนวทาง Boosting ภายใต้โครงสร้างข้อมูลชุดนี้ เมื่อเปรียบเทียบระหว่าง Gradient Boosting และ Extra Trees พบว่า Gradient Boosting มีค่า MAE สูงกว่าแต่ค่า RMSE ต่ำกว่า Extra Trees



ซึ่งแสดงให้เห็นว่า Gradient Boosting มีแนวโน้มเกิดความคลาดเคลื่อนขนาดเล็กจำนวนมากกว่า ขณะที่ Extra Trees มีความคลาดเคลื่อนขนาดใหญ่เฉพาะกรณีมากกว่า สะท้อนความแตกต่างของกลไก Boosting และ Randomized Ensemble ในการจัดการกับ Outlier

Table 4. Comparison of regression performance (MAE, RMSE, and R^2) of the evaluated machine learning algorithms

Algorithm	MAE (บาท)	RMSE (บาท)	R^2
Random Forest	391,225 $\pm$ 69,889	1,099,341 $\pm$ 259,654	0.784
Gradient Boosting	469,392 $\pm$ 61,204	1,147,861 $\pm$ 237,554	0.748
Extra Trees	432,360 $\pm$ 58,561	1,198,580 $\pm$ 224,412	0.721

เมื่อพิจารณา MAE และ RMSE เทียบกับช่วงราคาของข้อมูลจริง พบว่า Random Forest ให้ค่า MAE เท่ากับ 391,225 บาท หรือคิดเป็นประมาณ 33% ของค่าเฉลี่ยราคา 1,180,245 บาท สำหรับ RMSE นั้น มีค่าสูงกว่า MAE อย่างเห็นได้ชัด เนื่องจาก RMSE ให้น้ำหนักกับความคลาดเคลื่อนขนาดใหญ่มากกว่า MAE กล่าวคือ โกที่มีราคาสูงถึง 20,000,000 บาท มีจำนวนน้อยมากในชุดข้อมูล ทำให้โมเดลไม่ได้เรียนรู้รูปแบบของกลุ่มนี้เพียงพอ และเมื่อพยากรณ์ผิดพลาดในกรณีเหล่านี้ จึงส่งผลให้ RMSE สูงขึ้น

ผลการประเมินเชิง Classification ภายใต้เกณฑ์ $\pm 20\%$

ผู้วิจัยได้กำหนดเกณฑ์การประเมินเชิงปฏิบัติการ โดยพิจารณาว่าการพยากรณ์ที่มีค่าคลาดเคลื่อนไม่เกิน $\pm 20\%$ ของราคาจริง ถือว่าโมเดลพยากรณ์ถูกต้องและยอมรับได้ ผลการทดลองแสดงใน Table 5 ทั้งนี้ ค่า Precision, Recall และ F1-score ใน Table 5 คำนวณจากค่าเฉลี่ยของทั้ง 5 Fold ซึ่งอาจแตกต่างเล็กน้อยจากค่าที่คำนวณตรงจาก Aggregate Confusion Matrix ใน Table 6 เนื่องจากความแตกต่างของการกระจายข้อมูลในแต่ละ Fold

Table 5. Average classification performance (mean $\pm$ standard deviation) of the evaluated models based on 5-Fold Cross-Validation

Algorithm	Accuracy	Precision	Recall	F1-score
Random Forest	0.958 $\pm$ 0.018	0.869 $\pm$ 0.070	0.948 $\pm$ 0.052	0.905 $\pm$ 0.039
Gradient Boosting	0.953 $\pm$ 0.012	0.847 $\pm$ 0.034	0.948 $\pm$ 0.048	0.894 $\pm$ 0.024
Extra Trees	0.949 $\pm$ 0.018	0.843 $\pm$ 0.061	0.933 $\pm$ 0.067	0.884 $\pm$ 0.042

ผลการทดลองแสดงให้เห็นว่า Random Forest มีค่า F1-score สูงที่สุด ซึ่งสะท้อนถึงสมดุลระหว่าง Precision และ Recall ที่เหมาะสมที่สุด ภายใต้เงื่อนไขข้อมูลที่มีความไม่สมดุล ค่า Accuracy ของทุกโมเดล มีค่ามากกว่า 0.94 แสดงถึงความสามารถในการนำไปใช้งานได้โดยไม่เกิดการทำนายที่ผิดพลาดเกินไป

การวิเคราะห์ Confusion Matrix

จากการวิเคราะห์ Confusion Matrix ใน Table 6 พบว่า True Negative (TN) มีค่าสูงที่สุดในทุกโมเดล บ่งบอกถึงการกระจายของชุดข้อมูลจริง กล่าวคือ กลุ่ม Negative (ค่าตลาดเคลื่อนไหว $\pm 20\%$) มีจำนวนประมาณ 800 รายการ คิดเป็นประมาณ 79.4% ของข้อมูลทั้งหมด 1,008 รายการ ขณะที่กลุ่ม Positive มีเพียงประมาณ 208 รายการ คิดเป็น 20.6% แสดงให้เห็นว่า ชุดข้อมูลมีลักษณะ Class Imbalance ซึ่งเกิดจากธรรมชาติของตลาดหุ้นเมืองที่มีความผันผวนของราคาสูง ทำให้การพยากรณ์ที่อยู่ภายในเกณฑ์ $\pm 20\%$ เป็นเหตุการณ์ที่เกิดขึ้นในสัดส่วนน้อยกว่า

เมื่อพิจารณา Recall แยกรายกลุ่มจาก Table 8 ถึง Table 10 พบว่า Random Forest ให้ค่า Recall ที่สม่ำเสมอในทุกกลุ่มราคา ได้แก่ กลุ่ม Low เท่ากับ 0.957, กลุ่ม Medium เท่ากับ 0.955 และกลุ่ม High เท่ากับ 0.964 Gradient Boosting ให้ค่ากลุ่ม Low เท่ากับ 0.950, กลุ่ม Medium เท่ากับ 0.950 และกลุ่ม High เท่ากับ 0.948 และ Extra Trees ให้ค่ากลุ่ม Low เท่ากับ 0.940, กลุ่ม Medium เท่ากับ 0.938 และกลุ่ม High เท่ากับ 0.935 ความสม่ำเสมอของ Recall ในทุกกลุ่มราคาแสดงให้เห็นว่า โมเดลไม่แสดงอคติต่อกลุ่มใดกลุ่มหนึ่งเกินไป ซึ่งโมเดลมีความเสถียรพอสำหรับการใช้งานในระบบประมาณการราคา ที่ต้องการลดความเสี่ยงจากการประเมินต่ำหรือสูงเกินจริง

Table 6. Aggregated confusion matrix results (TP, FP, FN, TN) of the three algorithms based on 5-Fold Cross-Validation

Algorithm	TP	FP	FN	TN
Random Forest	197	31	11	769
Gradient Boosting	197	36	11	764
Extra Trees	194	37	14	763

หมายเหตุ: ค่าในตารางนี้เป็นผลรวมสะสมจากทั้ง 5 Fold และอาจแตกต่างเล็กน้อยจากค่าเฉลี่ยราย Fold ที่รายงานใน Table 5 เนื่องจากความแตกต่างของการกระจายข้อมูลในแต่ละ Fold

จากการพิจารณา Confusion Matrix พบว่า ความผิดพลาดส่วนใหญ่มักเกิดในกรณีที่ราคาจริงอยู่ใกล้ขอบเขต $\pm 20\%$ ซึ่งสะท้อนลักษณะข้อมูลที่มีความต่อเนื่อง ไม่ใช่ความผิดพลาดรุนแรง ซึ่งเป็นลักษณะปกติของข้อมูลประเภท threshold-based classification

ผลการวิเคราะห์ความสำคัญของตัวแปร

จากผลการวิเคราะห์ Feature Importance ดัง Table 7 พบว่า ตัวแปรที่มีอิทธิพลสูงสุดต่อการพยากรณ์มูลค่าตลาดของหุ้นเมืองสามอันดับแรก ได้แก่ มูลค่ารางวัลล่าสุดที่ได้รับ (RF: 0.5498) สีของชน (RF: 0.0803) และน้ำหนัก (RF: 0.0583) ตามลำดับ ซึ่งทั้งสามตัวแปรมีความ Importance รวมกันคิดเป็นสัดส่วน 68.84% ของน้ำหนักทั้งหมด บ่งบอกว่าการกำหนดมูลค่าของหุ้นเมืองมูลค่าสูงถูกขับเคลื่อนโดยตัวแปรกลุ่มหลักเพียงไม่กี่ตัว มากกว่าการกระจายน้ำหนักเท่า ๆ กันในทุกตัวแปร



ผลการเปรียบเทียบ Feature Importance ระหว่างโมเดลทั้งสามพบความสอดคล้องในตัวแปรที่มีความสำคัญสูง กล่าวคือ ตัวแปรมูลค่ารางวัลล่าสุดที่ได้รับและสีของขน ปรากฏในลำดับต้นของทั้ง RF (0.5498, 0.0803), GB (0.8045, 0.0345) และ ET (0.3608, 0.0872) ความสอดคล้องนี้เพิ่มความน่าเชื่อถือว่า ตัวแปรดังกล่าวมีความสัมพันธ์ที่แท้จริงกับมูลค่าตลาด ไม่ใช่ผลจาก Overfitting ของโมเดลใดโมเดลหนึ่ง นอกจากนี้ ตัวแปรลักษณะความสูงมีค่า Importance เป็นศูนย์ในทุกโมเดล แสดงว่าตัวแปรดังกล่าวไม่มีนัยสำคัญต่อการพยากรณ์มูลค่าในชุดข้อมูลนี้ และอาจพิจารณาตัดออกหรือนิยามใหม่ในการพัฒนาโมเดลรุ่นถัดไป เพื่อให้เห็นภาพชัดเจนขึ้น ผลที่ได้นำเสนอในรูปแบบ Bar Chart ดัง Figure 4 โดยเรียงลำดับจากตัวแปรที่มีความสำคัญสูงสุดไปต่ำสุด และเปรียบเทียบความสอดคล้องของ Feature Importance ระหว่างโมเดลทั้งสาม

Table 7. Feature importance scores of input variables across Random Forest, Gradient Boosting, and Extra Trees models

ลำดับ	ตัวแปร	RF	GB	ET	เฉลี่ย
1	มูลค่ารางวัลล่าสุดที่ได้รับ	0.5498	0.8045	0.3608	0.5717
2	สีของขน	0.0803	0.0345	0.0872	0.0673
3	น้ำหนัก	0.0583	0.0322	0.0915	0.0607
4	เทคนิคการต่อสู้	0.0654	0.0268	0.0746	0.0556
5	อายุของไก่ (วัน)	0.0682	0.0281	0.0692	0.0552
6	นาฬิกาที่ชนะล่าสุด	0.0597	0.0157	0.0797	0.0517
7	ครั้งที่ชนะ	0.0343	0.0101	0.0556	0.0333
8	ยกที่ชนะล่าสุด	0.0267	0.0190	0.0481	0.0313
9	สไตล์การต่อสู้	0.0239	0.0156	0.0506	0.0300
10	สีของขา	0.0196	0.0060	0.0510	0.0255
11	ลักษณะของหาง	0.0138	0.0075	0.0317	0.0177
12	ลักษณะความสูง	0.0000	0.0000	0.0000	0.0000

หมายเหตุ: เรียงตามค่าเฉลี่ยของทั้งสามโมเดล

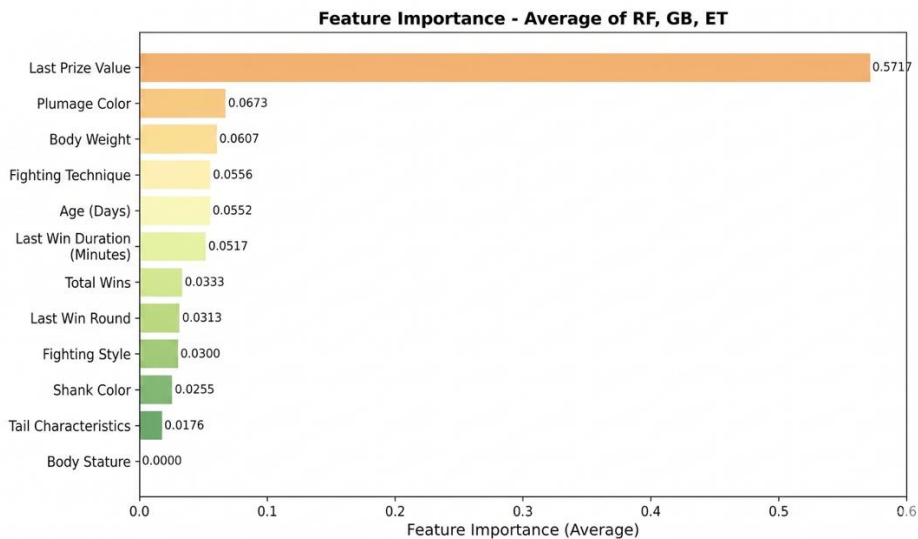


Figure 4. Average feature importance (RF, GB, ET)

การวิเคราะห์แบบ 3 ระดับราคา

งานวิจัยนี้ได้พิจารณา Confusion Matrix ของอัลกอริทึมที่ให้ผลลัพธ์ที่ดีที่สุด โดยแบ่งช่วงราคาออกเป็นดังนี้ 1) ช่วง Low มีราคาน้อยกว่าหรือเท่ากับ 25,000 บาท 2) ช่วง Medium ราคาอยู่ระหว่างมากกว่า 25,000 บาท และน้อยกว่าหรือเท่ากับ 850,000 บาท 3) ช่วง High มีราคามากกว่า 850,000 บาท ซึ่ง Table 8, 9, 10 แสดงผลของ Confusion Matrix ของอัลกอริทึม Random Forest, Gradient Boosting และ Extra Trees ตามลำดับ

Table 8. Confusion matrix of the Random Forest model for three price categories (Low, Medium, High)

Actual \ Predicted	Low	Medium	High
Low	287	8	5
Medium	7	382	11
High	4	7	297

Table 9. Confusion matrix of the Gradient Boosting model for three price categories (Low, Medium, High)

Actual \ Predicted	Low	Medium	High
Low	285	10	5
Medium	9	380	11
High	6	10	292

Table 10. Confusion matrix of the Extra Trees model for three price categories (Low, Medium, High)

Actual \ Predicted	Low	Medium	High
Low	282	12	6
Medium	12	375	13
High	8	12	288

ผลการทดลองพบว่า แบบจำลองสามารถจำแนกกลุ่มราคาสูงได้อย่างมีความแม่นยำในระดับสูง โดยค่าการทำนายถูกต้องกระจุกตัวอยู่ในแนวทแยงของเมทริกซ์ ความคลาดเคลื่อนส่วนใหญ่เกิดขึ้นระหว่างกลุ่มราคาที่ติดกัน ได้แก่ Low และ Medium ขณะที่การจำแนกผิดพลาดจากกลุ่มราคาสูงไปยังกลุ่มราคาต่ำโดยตรงพบในสัดส่วนที่ต่ำมาก ซึ่งสะท้อนถึงความสามารถของแบบจำลองในการรักษาความถูกต้องของการประเมินในกลุ่มราคาสูงได้อย่างมีประสิทธิภาพ ในบริบทของภาคเกษตรและตลาดสินค้าเกษตร ความสามารถในการจำแนกสินค้าที่มีมูลค่าสูงได้อย่างถูกต้องมีความสำคัญต่อการสนับสนุนการตัดสินใจเชิงเศรษฐกิจของผู้มีส่วนเกี่ยวข้อง

การวิเคราะห์เสถียรภาพของโมเดล

การพิจารณาค่าเฉลี่ยร่วมกับส่วนเบี่ยงเบนมาตรฐานจาก 5-Fold Cross-Validation พบว่า ความแปรปรวนของค่าประเมินอยู่ในระดับต่ำ สะท้อนถึงความเสถียรของโมเดลภายใต้การสุ่มแบ่งข้อมูล โดย Random Forest แสดงทั้งค่าเฉลี่ย Accuracy สูงสุดเท่ากับ 0.958 และมีส่วนเบี่ยงเบนมาตรฐานต่ำสุดเท่ากับ ± 0.018 เมื่อเทียบกับ Gradient Boosting ที่มีค่าเฉลี่ย Accuracy เท่ากับ 0.953 มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.012 และ Extra Trees ที่มีค่าเฉลี่ย Accuracy เท่ากับ 0.949 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.018 นอกจากนี้ ช่วงความแตกต่างของ Accuracy ระหว่าง Fold สูงสุดและต่ำสุดของ Random Forest อยู่ที่ 0.049 ซึ่งมีค่าไม่สูงมาก แสดงให้เห็นว่าโมเดลไม่แสดงพฤติกรรม Overfitting ต่อชุดข้อมูลย่อยใดชุดหนึ่ง จึงถือเป็นโมเดลที่เหมาะสมที่สุดสำหรับการพัฒนาเป็นระบบ Web-based Price Estimation

อภิปรายผลการวิจัย

ผลการวิจัยแสดงให้เห็นว่า แบบจำลองในกลุ่ม Tree-based Ensemble สามารถประมาณการมูลค่าทางการตลาดได้อย่างมีประสิทธิภาพ โดยเฉพาะอัลกอริทึม Random Forest ให้ประสิทธิภาพสูงสุดในทุกตัวชี้วัด ได้แก่ MAE เท่ากับ 391,225 บาท, RMSE เท่ากับ 1,099,341 บาท และ R^2 เท่ากับ 0.784 ซึ่งต่ำกว่า Gradient Boosting ที่มี MAE เท่ากับ 469,392, RMSE เท่ากับ 1,147,861 และ Extra Trees ที่มี MAE เท่ากับ 432,360 และ RMSE เท่ากับ 1,198,580 ผลการทดลองนี้สอดคล้องกับ Sanjuan et al. (2025) ที่ยืนยันว่า Random Forest มีสมดุลระหว่างความแม่นยำและความสามารถในการตีความผลที่เหมาะสมสำหรับการนำไปใช้งานจริง อย่างไรก็ตาม งานวิจัยนี้มีข้อจำกัดสำคัญ ได้แก่ ชุดข้อมูลเก็บจากบริบทตลาดเฉพาะพื้นที่ ซึ่งอาจขยายผลไปยังตลาดอื่นได้ไม่ครอบคลุมทันที และค่า R^2 ที่ 0.784 แสดงให้เห็นว่า ยังมีความแปรปรวนประมาณ 21.6% ที่ไม่สามารถ



อธิบายได้ด้วยตัวแปรในชุดข้อมูลปัจจุบัน ซึ่งอาจมาจากปัจจัยเชิงพฤติกรรมตลาดที่ไม่ได้บันทึกเป็นตัวแปร

การใช้ 5-Fold Cross-Validation ช่วยลดอิทธิพลของการสุ่มแบ่งข้อมูลเพียงครั้งเดียว และทำให้สามารถประเมินความเสถียรของโมเดลในเชิงสถิติได้อย่างเป็นระบบ ผลการทดลองพบว่า ค่าเฉลี่ยของตัวชี้วัดมีความสอดคล้องในทุก Fold และมีส่วนเบี่ยงเบนมาตรฐานอยู่ในระดับต่ำ ไม่มี Fold ใดที่เกิดการลดลงของประสิทธิภาพอย่างรุนแรง ลักษณะดังกล่าวสะท้อนว่าโมเดลมีความสามารถในการประยุกต์ใช้กับข้อมูลใหม่ได้ดี และไม่เกิดการ Overfitting ต่อชุดข้อมูลย่อยใดชุดหนึ่ง ความเสถียรนี้เป็นปัจจัยสำคัญต่อการนำโมเดลไปใช้งานจริง เนื่องจากระบบประเมินราคาจะต้องรองรับข้อมูลใหม่ที่ไม่เคยปรากฏมาก่อน

เมื่อเปรียบเทียบผลลัพธ์ที่ได้กับงานวิจัยในบริบทที่ใกล้เคียง พบว่าแบบจำลองที่พัฒนาในงานวิจัยนี้ให้ผลลัพธ์ที่อยู่ในระดับที่สามารถแข่งขันได้ (Rahman et al., 2021) ใช้ CNN ในการจำแนกระดับราคาจากภาพถ่ายและคุณลักษณะทางกายภาพ และรายงานค่า Accuracy ในระดับ 0.705 ขณะที่งานวิจัยนี้ได้ค่า Accuracy เฉลี่ยสูงสุดที่ 0.958 ด้วย Random Forest อย่างไรก็ตาม การเปรียบเทียบโดยตรงมีข้อจำกัด เนื่องจากความแตกต่างของลักษณะข้อมูล กล่าวคือ งานของ (Rahman et al., 2021) ใช้ภาพถ่ายเป็นข้อมูลนำเข้าและประเมินมูลค่าจากลักษณะกายภาพเพียงมิติเดียว ในขณะที่งานวิจัยนี้ใช้ข้อมูลเชิงตารางที่ครอบคลุมทั้งลักษณะทางกายภาพ ประวัติการแข่งขัน และข้อมูลตลาด ซึ่งสะท้อนความซับซ้อนของปัจจัยกำหนดราคาในตลาดไก่พื้นเมืองได้ครอบคลุมกว่า นอกจากนี้ Krishna et al. (2025) รายงานค่า R^2 สูงถึง 0.988 ในการพยากรณ์ราคาพืชผลโดยใช้แบบจำลอง XGBoost แสดงถึงความสามารถในการอธิบายความแปรปรวนของข้อมูลได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม งานวิจัยดังกล่าวใช้ข้อมูลเชิงโครงสร้างที่ประกอบด้วยปัจจัยด้านสิ่งแวดล้อมและเศรษฐกิจ เช่น อุณหภูมิ ปริมาณน้ำฝน อุปสงค์ อุปทาน และต้นทุนการขนส่ง ซึ่งเป็นตัวแปรเชิงปริมาณที่วัดได้อย่างชัดเจนและมีรูปแบบค่อนข้างสม่ำเสมอ แตกต่างจากราคาไก่พื้นเมืองมูลค่าสูงซึ่งมีความผันผวนจากปัจจัยเชิงพฤติกรรมและการรับรู้ของตลาดที่ไม่สามารถวัดได้เชิงปริมาณ ส่งผลให้การพยากรณ์ราคามีความซับซ้อนมากยิ่งขึ้น ดังนั้น ค่า R^2 ที่ 0.784 ของงานวิจัยนี้จึงถือว่ามีความเหมาะสมสำหรับบริบทข้อมูลดังกล่าว

Feature Importance ต่อการประยุกต์ใช้จริง

ผลการวิเคราะห์ Feature Importance มีนัยสำคัญทั้งในเชิงวิชาการและเชิงปฏิบัติการ ในเชิงวิชาการ การที่ตัวแปรมูลค่ารางวัลล่าสุดที่ได้รับมีค่า Importance สูงสุดมีค่าเฉลี่ย 0.5717 หรือคิดเป็น 57.17% ของน้ำหนักทั้งหมด สอดคล้องกับสมมติฐานที่ว่าตลาดไก่พื้นเมืองมูลค่าสูงให้น้ำหนักกับฟอร์มปัจจุบันของไก่พื้นเมืองมากกว่าคุณลักษณะสะสมในระยะยาว กล่าวคือ ผู้ซื้อในตลาดนี้ใช้ผลการแข่งขันครั้งล่าสุดเป็นสัญญาณหลักในการประเมินศักยภาพและความคุ้มค่าของการลงทุน ซึ่งสะท้อนพฤติกรรมตัดสินใจที่อิงกับข้อมูลเชิงประจักษ์ล่าสุดมากกว่าการประเมินจากลักษณะภายนอกเพียงอย่างเดียว รองลงมา คือ สีของขน ค่าเฉลี่ย 0.0673 และน้ำหนักค่าเฉลี่ย 0.0607 ซึ่งเป็นตัวแปรเชิงกายภาพที่ตลาดใช้ประกอบการพิจารณาในลำดับรอง บ่งชี้ว่าแม้ลักษณะภายนอกจะมีผลต่อมูลค่า แต่น้ำหนักน้อยกว่าประวัติสมรรถนะอย่างมีนัยสำคัญ

ผลลัพธ์ Feature Importance สามารถนำไปใช้เป็นแนวทางสำหรับเกษตรกร ในการให้ความสำคัญกับการพัฒนาคุณลักษณะที่มีอิทธิพลสูงต่อมูลค่าตลาด กล่าวคือ เกษตรกรควรให้ความสำคัญกับการบันทึกและรักษาประวัติผลการแข่งขัน อย่างเป็นระบบและต่อเนื่อง เนื่องจากข้อมูลนี้มีผลต่อการพยากรณ์มูลค่ามากที่สุด การมีบันทึกผลการแข่งขันที่ครบถ้วนและน่าเชื่อถือจึงไม่เพียงช่วยให้โมเดลทำงานได้แม่นยำขึ้น แต่ยังเพิ่มความโปร่งใสและความน่าเชื่อถือในการตั้งราคาต่อผู้ซื้อด้วย นอกจากนี้ ตัวแปรลักษณะความสูงที่มีค่า Importance เป็นศูนย์ในทุกโมเดล อาจพิจารณาตัดออกในรุ่นถัดไปของโมเดล เพื่อลดความซับซ้อนและเพิ่มความเร็วในการประมวลผลโดยไม่กระทบประสิทธิภาพโดยรวม

การวิเคราะห์เชิงโครงสร้างของแบบจำลอง

Random Forest แสดงให้เห็นถึงความเหมาะสมกับข้อมูลชุดนี้ อย่างชัดเจน โดยมีค่า R^2 เท่ากับ 0.784 ซึ่งสูงกว่าทุกโมเดลที่เปรียบเทียบ ในเชิงทฤษฎี Ensemble Learning แบบ Bagging มีบทบาทสำคัญในการลด Variance ของโมเดล ลดความไวต่อ Noise และลดผลกระทบของ Outlier เมื่อพิจารณาจากลักษณะข้อมูลที่มีความผันผวนเชิงพฤติกรรมตลาดและปัจจัยแฝงจำนวนมาก การใช้แนวทาง Bagging จึงช่วยรักษาเสถียรภาพของผลลัพธ์ได้ดีกว่าแนวทาง Boosting ที่เน้นลด Bias ผ่านการเรียนรู้แบบลำดับขั้น

การตีความค่า R^2 และความสามารถในการอธิบายความแปรปรวน

ค่า R^2 ของ Random Forest เท่ากับ 0.784 แสดงว่าแบบจำลองสามารถอธิบายความแปรปรวนของราคาจริงได้ประมาณ 78.4% แม้ว่าจะยังมีความแปรปรวนบางส่วนที่ไม่สามารถอธิบายได้ เช่น ปัจจัยเชิงพฤติกรรมของผู้ซื้อ ปัจจัยทางสังคมและความนิยมเฉพาะกลุ่ม ความผันผวนตามฤดูกาล ข้อมูลเชิงคุณภาพที่ไม่ได้ถูกบันทึกเป็นตัวเลข เป็นต้น

การวิเคราะห์ภายใต้เกณฑ์ $\pm 20\%$ และมิติการใช้งานจริง

การกำหนดเกณฑ์ $\pm 20\%$ มีความสำคัญในเชิงการประยุกต์ใช้งานจริง เนื่องจากสะท้อนถึงช่วงความคลาดเคลื่อนที่ยอมรับได้ในเชิงตลาด ค่า Accuracy มากกว่า 0.94 ในทุกโมเดล และค่า F1-score ของ Random Forest ที่สูงที่สุดที่ 0.905 บ่งบอกว่าโมเดลสามารถสร้างสมดุลระหว่าง Precision และ Recall ได้ดี ในเชิงระบบสนับสนุนการตัดสินใจ ผลลัพธ์ดังกล่าวแสดงว่าโมเดลสามารถลดความเสี่ยงจากการประเมินต่ำเกินไปหรือสูงเกินไปจริง ได้ในระดับที่เหมาะสม

Class Imbalance และผลกระทบต่อโมเดล

การวิเคราะห์ Confusion Matrix พบว่า ข้อมูลมีลักษณะไม่สมดุล โดยกลุ่มที่พยากรณ์ไม่อยู่ในช่วง $\pm 20\%$ มีจำนวนมากกว่ากลุ่มที่อยู่ในช่วงยอมรับได้ อย่างไรก็ตาม ค่า Recall ยังคงอยู่ในระดับสูงประมาณ 0.95 แสดงว่าโมเดลมีแนวโน้มตรวจจับกลุ่มตัวอย่างขนาดเล็กได้ในระดับที่น่าพอใจ เมื่อพิจารณาพร้อมกับค่า Precision และ F1-score รวมใน Table 5 พบว่า Random Forest ให้ค่า Precision ที่ 0.869 และ F1-score ที่ 0.905 แสดงให้เห็นว่า โมเดลยังรักษาสมดุลระหว่างความแม่นยำ

และความครอบคลุมได้ในภาพรวม อย่างไรก็ตาม ค่าดังกล่าวเป็นค่าเฉลี่ยรวมทั้งชุดข้อมูล การประเมินผลกระทบของ Class Imbalance อย่างครบถ้วนยังคงต้องอาศัย Precision และ F1-score แยกรายกลุ่มราคา ซึ่งเป็นข้อจำกัดของการวิเคราะห์ในงานวิจัยนี้ ทั้งนี้งานวิจัยนี้ไม่ได้ดำเนินการเปรียบเทียบกับโมเดลประเภทอื่น จึงไม่สามารถสรุปได้ว่า Ensemble Learning จัดการกับ Class Imbalance ได้ดีกว่าแนวทางอื่นในบริบทนี้

การจำแนกระดับราคาและผลกระทบเชิงเศรษฐกิจ

การวิเคราะห์แบบ 3 ระดับราคา Low, Medium และ High พบว่าโมเดลสามารถจำแนกกลุ่ม High ได้อย่างแม่นยำสูง และไม่พบกรณีที่ราคาสูงถูกจำแนกเป็นราคาต่ำอย่างรุนแรง ซึ่งมีนัยสำคัญเชิงเศรษฐกิจอย่างยิ่ง เนื่องจากความผิดพลาดประเภทนี้จะก่อให้เกิดความเสียหายทางการเงินสูงสุดต่อผู้ขาย และอาจทำลายความน่าเชื่อถือของระบบ การที่โมเดลสามารถป้องกันความผิดพลาดประเภทนี้ได้แสดงให้เห็นว่า แบบจำลองมีความปลอดภัยเชิงการตัดสินใจในระดับที่ยอมรับได้สำหรับการนำไปใช้จริงในตลาดสัตว์พื้นเมืองมูลค่าสูง ความผิดพลาดที่เกิดขึ้นจริงส่วนใหญ่อยู่ระหว่างกลุ่ม Low และ Medium ซึ่งมีช่วงราคาที่ทับซ้อนกันมากกว่า และผลกระทบทางเศรษฐกิจต่ำกว่าการสับสนระหว่างกลุ่ม High กับกลุ่มอื่น

เหตุผลที่ Random Forest มีประสิทธิภาพสูงกว่าโมเดลอื่น

Random Forest ให้ประสิทธิภาพสูงในงานวิจัยนี้ สามารถอธิบายได้จากลักษณะเฉพาะของชุดข้อมูลและกลไกของอัลกอริทึม กล่าวคือ ชุดข้อมูลไก่พื้นเมืองมูลค่าสูงมีความผันผวนของราคาสูงและมี Outlier จากการซื้อขายที่ราคาผิดปกติอยู่จำนวนหนึ่ง Random Forest ซึ่งใช้หลักการ Bagging และสร้างต้นไม้ตัดสินใจหลายต้นอย่างอิสระจากกัน มีความทนทานต่อ Outlier สูงกว่า Gradient Boosting ซึ่งเรียนรู้แบบลำดับขั้นและมีแนวโน้มให้น้ำหนักกับข้อมูลที่พยากรณ์ผิดในรอบก่อนหน้ามากขึ้นเรื่อย ๆ ทำให้ Gradient Boosting ไวต่อ Outlier กว่าในบริบทข้อมูลนี้ เมื่อเปรียบเทียบกับ Extra Trees ซึ่งสุ่มจุดแบ่งของต้นไม้อย่างสมบูรณ์ พบว่า Random Forest ให้ผลดีกว่า เนื่องจากชุดข้อมูลในงานวิจัยนี้มีตัวแปรที่มีอิทธิพลสูงอย่างชัดเจน โดยเฉพาะมูลค่ารางวัลล่าสุดที่ได้รับที่มีค่า Importance สูงถึง 57.17% การที่ Random Forest ค้นหาจุดแบ่งที่ดีที่สุดจากกลุ่มตัวแปรสุ่ม แทนที่จะสุ่มจุดแบ่งอย่างสมบูรณ์แบบ Extra Trees ทำให้โมเดลสามารถใช้ประโยชน์จากตัวแปรที่มีอำนาจในการจำแนกสูงได้อย่างมีประสิทธิภาพมากกว่า ผลนี้สอดคล้องกับข้อค้นพบของ Sanjuan et al. (2025) ที่อธิบายว่า Random Forest เป็นแบบจำลองที่ให้สมดุลระหว่างความแม่นยำและความสามารถในการตีความ โดยให้ผลการพยากรณ์ที่มีประสิทธิภาพใกล้เคียงกับ Neural Network แต่สามารถอธิบายผลลัพธ์ได้ดีกว่า จึงเหมาะสมต่อการนำไปใช้งานจริงในบริบทของระบบฟาร์มปศุสัตว์

ความหมายของผลลัพธ์ต่อการประยุกต์ใช้จริงในระบบประเมินราคา

ผลลัพธ์ที่ได้มีนัยสำคัญต่อการออกแบบระบบประเมินราคาในเชิงปฏิบัติกรอย่างน้อย 3 ด้าน ด้านแรก คือ ความเหมาะสมสำหรับการใช้งานจริง ค่า Accuracy เท่ากับ 0.958 และ F1-score ที่สูงของ Random Forest บ่งชี้ว่าระบบสามารถพยากรณ์มูลค่าได้ถูกต้องภายในเกณฑ์ $\pm 20\%$ ในกรณีส่วนใหญ่ ซึ่งในเชิงปฏิบัติหมายความว่าเกษตรกรและผู้ประกอบการสามารถใช้ระบบนี้เป็นจุดอ้างอิงในการ

เจรจาราคาได้จริง โดยไม่ต้องพึ่งพาการประเมินจากผู้เชี่ยวชาญเพียงอย่างเดียว ซึ่งมักมีความลำเอียงเชิงส่วนตัวและขาดความโปร่งใส ด้านที่สอง คือ การลดความไม่สมมาตรของข้อมูล ในตลาดปัจจุบัน ผู้ซื้อที่มีประสบการณ์มีความได้เปรียบเหนือเกษตรกรรายย่อยในการประเมินมูลค่าสัตว์ ระบบที่พัฒนาในงานวิจัยนี้สามารถช่วยลดช่องว่างดังกล่าว โดยให้ราคาอ้างอิงที่อิงจากข้อมูลเชิงประจักษ์ ทำให้การซื้อขายมีความเป็นธรรมมากขึ้นสำหรับทุกฝ่าย และด้านที่สาม คือ ข้อจำกัดที่ควรตระหนักก่อนนำไปใช้จริง แม้ว่าโมเดลจะมีประสิทธิภาพสูง แต่ความแม่นยำขึ้นอยู่กับคุณภาพและความสมบูรณ์ของข้อมูลที่ป้อนเข้าระบบ โดยเฉพาะข้อมูลมูลค่ารางวัลล่าสุดที่ได้รับที่มีอิทธิพลสูงสุด หากข้อมูลนี้ไม่ครบถ้วนหรือไม่น่าเชื่อถือ ความแม่นยำของระบบจะลดลง นอกจากนี้ โมเดลที่ฝึกจากข้อมูลในช่วงเวลาหนึ่งอาจต้องได้รับการปรับปรุงอย่างสม่ำเสมอเมื่อสภาพตลาดเปลี่ยนแปลง เพื่อรักษาระดับความแม่นยำในระยะยาว

การวิเคราะห์กรณีที่โมเดลพยากรณ์ผิดพลาด

เพื่อให้เข้าใจข้อจำกัดของแบบจำลองอย่างชัดเจน งานวิจัยนี้วิเคราะห์ลักษณะของกรณีที่โมเดลพยากรณ์ผิดพลาด กล่าวคือ กรณีที่ค่าตลาดเคลื่อนเกินกว่าเกณฑ์ $\pm 20\%$ จากการวิเคราะห์ Confusion Matrix ใน Table 6 พบว่า False Positive (FP) ซึ่งหมายถึงกรณีที่โมเดลพยากรณ์ว่าอยู่ในช่วงยอมรับได้แต่ความจริงไม่ใช่ มีจำนวน 31 ถึง 37 กรณีขึ้นอยู่กับอัลกอริทึม และ False Negative (FN) ซึ่งหมายถึงกรณีที่โมเดลพยากรณ์ว่าไม่อยู่ในช่วงยอมรับได้แต่ความจริงอยู่ในช่วงดังกล่าว มีจำนวน 11 ถึง 14 กรณี เมื่อพิจารณาพร้อมกับผลการวิเคราะห์แบบ 3 ระดับราคาใน Table 8 ถึง 10 พบว่าความผิดพลาดส่วนใหญ่เกิดขึ้นระหว่างกลุ่มราคาที่อยู่ติดกัน คือ Low และ Medium ขณะที่กรณีที่ไถ่ระดับ High ถูกจำแนกเป็น Low หรือในทางกลับกัน พบในสัดส่วนที่ต่ำมาก กล่าวคือ 4 ถึง 8 กรณีจาก 308 รายการในกลุ่ม High บ่งบอกว่าความผิดพลาดมักเกิดที่บริเวณขอบเขตระหว่างกลุ่มราคาที่ใกล้เคียงกัน มากกว่าความผิดพลาดเชิงโครงสร้างของแบบจำลอง

ลักษณะของกรณีที่พยากรณ์ผิดพลาดมีแนวโน้มสัมพันธ์กับปัจจัยสำคัญ 3 ประการ ได้แก่ 1) ภาวะที่เมืองที่มีคุณลักษณะอยู่ในช่วงรอยต่อระหว่างกลุ่มราคา เช่น มีประวัติการแข่งขันที่ติดแต่ลักษณะทางกายภาพไม่โดดเด่น ทำให้โมเดลประเมินมูลค่าได้ยากกว่ากรณีที่คุณลักษณะทุกด้านชี้ไปในทิศทางเดียวกัน 2) ข้อมูลที่มีความผันผวนของราคาสูงผิดปกติ ซึ่งอาจเกิดจากปัจจัยเชิงพฤติกรรมของผู้ซื้อหรือบริบทตลาด ณ เวลานั้น ที่ไม่ได้ถูกบันทึกเป็นตัวแปรในชุดข้อมูล และ 3) กรณีที่ราคาซื้อขายจริงอยู่ใกล้ขอบเขต $\pm 20\%$ ซึ่งเป็นลักษณะทั่วไปของการจำแนกแบบกำหนดเกณฑ์ ความคลาดเคลื่อนที่เกิดขึ้นบริเวณขอบเขตดังกล่าวจึงไม่ถือเป็นข้อบกพร่องที่รุนแรงของแบบจำลอง แต่สะท้อนถึงความต่อเนื่องของข้อมูลมากกว่า

ข้อค้นพบเหล่านี้ชี้ให้เห็นว่า การเพิ่มตัวแปรที่จับปัจจัยเชิงบริบทตลาด เช่น ช่วงเวลาของปี ภูมิภาค หรือระดับความต้องการของตลาด ณ เวลาที่ทำธุรกรรม อาจช่วยลดจำนวนกรณีที่พยากรณ์ผิดพลาดในกลุ่มขอบเขตได้



ข้อเสนอแนะ

ปัจจัยสำคัญที่สุดที่มีผลต่อประสิทธิภาพของแบบจำลอง คือ คุณภาพและความครอบคลุมของข้อมูล โดยเฉพาะตัวแปรที่มีค่า Feature Importance สูงสุดตามผลใน Table 7 ได้แก่ มูลค่ารางวัลล่าสุดที่ได้รับมีค่าเฉลี่ยที่ 0.5717 สีของขนมีค่าเฉลี่ยที่ 0.0673 และน้ำหนักมีค่าเฉลี่ยที่ 0.0607 ซึ่งคิดเป็น 68.84% ของน้ำหนักความสำคัญทั้งหมด ดังนั้น การพัฒนางานวิจัยต่อยอดจึงควรให้ความสำคัญตามลำดับ ดังนี้ ลำดับที่ 1 มูลค่ารางวัลล่าสุดที่ได้รับ ควรพัฒนาระบบบันทึกประวัติผลการแข่งขันแบบมาตรฐานและต่อเนื่อง เช่น การออกแบบฟอร์มบันทึกผลการแข่งขันที่ใช้ร่วมกันในชุมชนผู้เลี้ยง หรือการเชื่อมต่อกับฐานข้อมูลสมาคมการแข่งขัน เพื่อให้ข้อมูลที่ป้อนเข้าระบบมีความน่าเชื่อถือและตรวจสอบได้ ลำดับที่ 2 ตัวแปรเชิงกายภาพ เช่น สีของขน หรือน้ำหนักตัว ควรพัฒนาโปรโตคอลการวัดและบันทึกที่เป็นมาตรฐานเดียวกัน เช่น การใช้ระบบการจำแนกสีจากภาพถ่ายหรือการชั่งน้ำหนักในสภาพแวดล้อมควบคุม เพื่อลดความคลาดเคลื่อนเชิงโครงสร้างจากการบันทึกที่แตกต่างกันระหว่างผู้เก็บข้อมูล และลำดับที่ 3 ตัวแปรเกี่ยวกับตลาดซื้อขาย ซึ่งจากการวิเคราะห์ข้อผิดพลาด พบว่าความผิดพลาดของโมเดลส่วนหนึ่ง เกิดจากปัจจัยเกี่ยวกับตลาดซื้อขายที่ไม่ได้บันทึกเป็นตัวแปร ดังนั้น การเพิ่มตัวแปรเช่น ช่วงเวลาของปี ภูมิภาค หรือระดับความต้องการตลาดในเวลาทำธุรกรรม อาจช่วยลดจำนวนกรณีที่ยากกรณีผิดพลาดในกลุ่มขอบเขตราคาได้ นอกจากนี้การเชื่อมต่อบนระบบ Web Application เข้ากับฐานข้อมูลตลาดแบบเรียลไทม์ จะช่วยให้โมเดลปรับตัวต่อสภาวะตลาดที่เปลี่ยนแปลงได้อย่างต่อเนื่อง โดยไม่ต้องรอการฝึกโมเดลใหม่ทั้งหมด

References

- Chuluunsai Khan, T., Kim, J.-H., Park, S.-H., & Nasridinov, A. (2024). Analyzing internal and external factors in livestock supply forecasting using machine learning: Sustainable insights from South Korea. *Sustainability*, 16, 6907.
- Hossain, M. R. H., Islam, M. R., McGrath, S. R., Islam, M. Z., & Lamb, D. (2025a). Learning-based estimation of cattle weight gain and its influencing factors. *Computers and Electronics in Agriculture*, 231, Article 109002.
- Hossain, M. R. H., Islam, M. R., McGrath, S. R., Islam, M. Z., & Lamb, D. (2025b). Mob-based cattle weight gain forecasting using machine learning models. *Smart Agricultural Technology*, 12, Article 100412.
- Krishna, P. A., Narayana, G. V. S., Kotha, S. K., & Pattnayak, D. (2025). Machine learning based agricultural price forecasting for major food crops in India using environmental and economic factors. *Biology and Life Sciences Forum*, 54, 7.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Lyu, P., Min, J., & Song, J. (2023). Application of machine learning algorithms for on-farm monitoring and prediction of broilers' live weight: A quantitative study based on body weight data. *Agriculture*, 13, 2193.



- Ma, W., Peng, L., Chen, H., & Yan, H. (2025). Interval price prediction of livestock product based on fuzzy mathematics and improved LSTM. *PLoS ONE*, 20(2), e0318823.
- Mallory, M., Peng, R., Ma, M., & Wang, H. H. (2025). High-dimensional spatial-plus-vertical price relationships and price transmission: A machine learning approach. *Journal of Agricultural Economics*. Advance online publication
- Oliveira, F. M., Ferraz, P. F. P., Ferraz, G. A. e. S., Pereira, M. N., Barbari, M., & Rossi, G. (2025). Prediction of body mass of dairy cattle using machine learning algorithms applied to morphological characteristics. *Animals*, 15, 1054.
- Papadopoulos, G., Papantonatou, M.-Z., Uyar, H., Kriezi, O., Mavrommatis, A., Psiroukis, V., Kasimati, A., Tsiplakou, E., & Fountas, S. (2025). Economic and environmental benefits of digital agricultural technological solutions in livestock farming: A review. *Smart Agricultural Technology*. Advance online publication.
- Rahman, M. A., Kabir, M. A., Haque, M. E., & Hossain, B. M. M. (2021). A machine learning-based price prediction for cows. *American Journal of Agricultural Science, Engineering and Technology*, 5(1), 63–69.
- Rahmani, E., Khatami, M., & Stephens, E. (2024). Using probabilistic machine learning methods to improve beef cattle price modeling and promote beef production efficiency and sustainability in Canada. *Sustainability*, 16, 1789.
- Sanjuan, S., Méndez, D. A., Arnau, R., Calabuig, J. M., Díaz de Otálora Aguirre, X., & Estellés, F. (2025). Soft computing approaches for predicting shade-seeking behavior in dairy cattle under heat stress: A comparative study of random forests and neural networks. *Mathematics*, 13, 2662.
- Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H., & Liu, P. (2023). Agricultural product price forecasting methods: A review. *Agriculture*, 13, 1671.
- Zhang, N., An, Q., Zhang, S., & Ma, H. (2025). Price prediction for fresh agricultural products based on a boosting ensemble algorithm. *Mathematics*, 13, 71.

