

บทที่ 5

สรุปและข้อเสนอแนะ

การดำเนินโครงการวิจัย การพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ในประเทศไทย มีวัตถุประสงค์เพื่อศึกษาและพัฒนาตัวแบบพยากรณ์ราคาที่มีความแม่นยำสูงสุด โดยเปรียบเทียบระหว่างแบบจำลองทางสถิติ (SARIMA) และแบบจำลองการเรียนรู้ของเครื่อง (Linear Regression, Random Forest และ Support Vector Regression) รวมถึงเพื่อพัฒนาระบบสารสนเทศสำหรับการแสดงผลการพยากรณ์ โดยสรุปผลการดำเนินงาน อภิปรายผล และข้อเสนอแนะได้ดังนี้

5.1 สรุปผลการดำเนินงาน

จากการดำเนินการวิเคราะห์ข้อมูลราคาข้าวโพดเลี้ยงสัตว์รายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2552 ถึงเดือนธันวาคม พ.ศ. 2566 รวมระยะเวลา 15 ปี (180 เดือน) โดยใช้กระบวนการมาตรฐาน CRISP-DM และแบ่งชุดข้อมูลเพื่อทดสอบประสิทธิภาพใน 3 สัดส่วน ได้แก่ 70:30, 80:20 และ 90:10 สรุปผลได้ดังนี้

1. ผลการพัฒนาตัวแบบพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ ในการพัฒนาตัวแบบเพื่อพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ ผลการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองทางสถิติและแบบจำลองการเรียนรู้ของเครื่องพบว่า แบบจำลอง Linear Regression ที่สัดส่วนข้อมูล 90:10 มีความเหมาะสมที่สุดสำหรับชุดข้อมูลนี้ โดยสามารถทำนายค่าในชุดทดสอบได้แม่นยำสูงด้วยค่า RMSE 0.652 และค่า MAE 0.539 บาท อีกทั้งยังมีค่าเปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) ต่ำเพียง 5.19% (หรือคิดเป็นความแม่นยำสูงถึงร้อยละ 94.81) ในขณะที่แบบจำลองที่มีความซับซ้อนอย่าง Random Forest และ SVR เกิดปัญหาการเรียนรู้เกิน (Overfitting) อย่างชัดเจน

2. ผลการนำเสนอข้อมูลการพยากรณ์ผ่านระบบสารสนเทศ โครงการนี้ยังได้บรรลุเป้าหมายในการนำเสนอข้อมูลพยากรณ์สู่ผู้ใช้งานจริง โดยการพัฒนาระบบสารสนเทศในรูปแบบแดชบอร์ด (Dashboard) บนเว็บไซต์ ซึ่งนำผลลัพธ์จากแบบจำลองที่มีประสิทธิภาพสูงสุดมาจัดทำเป็นภาพกราฟและข้อมูลแนวโน้มล่วงหน้าที่น่าเข้าใจง่าย ช่วยให้เกษตรกรและผู้ประกอบการสามารถเข้าถึงข้อมูลเพื่อใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการวางแผนการผลิตและการตลาดได้อย่างมีประสิทธิภาพ

5.2 อภิปรายผล

จากผลการดำเนินงาน สามารถนำมาอภิปรายผลโดยเชื่อมโยงกับแนวคิดและทฤษฎีที่เกี่ยวข้องได้ ดังนี้

ความเหมาะสมของแบบจำลองเชิงเส้น (Linearity of Data) การที่แบบจำลอง Linear Regression มีประสิทธิภาพสูงกว่าแบบจำลองอื่นๆ อภิปรายได้ว่า โครงสร้างของราคาข้าวโพดเลี้ยงสัตว์และปัจจัยแวดล้อมทางเศรษฐศาสตร์ที่นำมาศึกษา (เช่น ปริมาณผลผลิต, พื้นที่เพาะปลูก, ราคาน้ำมันดีเซล) มีความสัมพันธ์กันในลักษณะเชิงเส้นตรง (Linear Relationship) อย่างชัดเจน การใช้สมการเชิงเส้นจึงสามารถจับทิศทางและอธิบายน้ำหนักของตัวแปรต่างๆ ได้อย่างตรงไปตรงมาและมีความเสถียร (Generalization) สูงกว่า

ปัญหาการเรียนรู้เกินในแบบจำลองขั้นสูง (Overfitting) จากการทดลองพบว่า แบบจำลอง Random Forest และ SVR มีค่าความคลาดเคลื่อนในชุดฝึกสอน (Train Error) ต่ำมาก แต่ค่าความคลาดเคลื่อนในชุดทดสอบ (Test Error) กลับพุ่งสูงขึ้น ปรากฏการณ์นี้สอดคล้องกับปัญหา Overfitting ซึ่งอธิบายได้ว่า แบบจำลองมีความซับซ้อนมากเกินไปเมื่อเทียบกับขนาดของข้อมูล (180 รายการ) ทำให้แบบจำลองพยายามจดจำสัญญาณรบกวน (Noise) หรือความผันผวนย่อย มากกว่าการเรียนรู้แนวโน้มหลัก ทำให้สูญเสียความสามารถในการพยากรณ์ข้อมูลใหม่

ความสำคัญของปริมาณข้อมูลสอน (Data Splitting) การที่ทุกแบบจำลองทำงานได้ดีที่สุดที่สัดส่วน 90:10 สะท้อนให้เห็นว่า ข้อมูลอนุกรมเวลารายเดือนที่มีจำนวนจำกัดจำเป็นต้องใช้สัดส่วนข้อมูลย้อนหลังสำหรับฝึกสอนให้มากที่สุด เพื่อให้แบบจำลองสามารถมองเห็นรอบวัฏจักร (Seasonality) และแนวโน้ม (Trend) ซ้ำๆ ได้เพียงพอต่อการสร้างกฎการพยากรณ์ที่แม่นยำ

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์ การวางแผนการจำหน่ายของเกษตรกร เกษตรกรสามารถใช้ข้อมูลราคาพยากรณ์จากระบบแดชบอร์ด เพื่อวางแผนชะลอการขายในช่วงที่ราคาตกต่ำ และเตรียมระบายสินค้าในช่วงที่ราคามีแนวโน้มปรับตัวสูงขึ้น การบริหารจัดการต้นทุน ผู้ประกอบการโรงงานอาหารสัตว์สามารถใช้ข้อมูลพยากรณ์ในการวางแผนจัดซื้อวัตถุดิบล่วงหน้า เพื่อบริหารจัดการต้นทุนการผลิตและรับมือกับความเสถียรด้านความผันผวนของราคาได้อย่างมีประสิทธิภาพ

5.3.2 ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

1. การเพิ่มตัวแปรนำเข้า ควรพิจารณนำตัวแปรด้านสภาพภูมิอากาศ (เช่น ปริมาณน้ำฝน หรืออุณหภูมิ) เข้ามาร่วมวิเคราะห์ เนื่องจากเป็นปัจจัยภายนอกที่อาจส่งผลกระทบต่อโดยตรงต่อปริมาณผลผลิตและราคา ซึ่งจะช่วยให้ประสิทธิภาพและความแม่นยำให้กับแบบจำลองได้มากยิ่งขึ้น

2. การศึกษาเปรียบเทียบด้วยเทคนิคขั้นสูง หากในอนาคตสามารถเก็บรวบรวมข้อมูลราคาในเวลาที่สูงขึ้น (เช่น รายสัปดาห์ หรือรายวัน) ควรศึกษาเปรียบเทียบด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เช่น Long Short-Term Memory (LSTM) หรือ Gated Recurrent Unit (GRU) ซึ่งถูกออกแบบมาเพื่อจัดการข้อมูลอนุกรมเวลาขนาดใหญ่ โดยเฉพาะ