ตัวแบบผสมสำหรับการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย

Hybrid Model for Forecasting Monthly Price of Maize in Thailand

ธนกร สุทธิสนธ์*

*Thanakon Sutthison**

สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี อุบลราชธานี

Program of Applied Statistics, Faculty of Science, Ubon Ratchathani Rajabhat University

Received : 2 February 2022
Revised : 30 March 2022
Accepted : 10 July 2022

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อสร้างและคัดเลือกตัวแบบการพยากรณ์ที่เหมาะสมกับข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย โดยใช้ข้อมูลจากสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์ ตั้งแต่เดือนมกราคม 2542 ถึงเดือน พฤศจิกายน 2564 จำนวน 275 ค่า แบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 คือ ข้อมูลการเรียนรู้ (Training dataset) ตั้งแต่เดือนมกราคม 2542 ถึงเดือน ธันวาคม 2562 จำนวน 252 ค่า ใช้สำหรับการสร้างตัวแบบการพยากรณ์ด้วยตัวแบบ SARIMA และตัวแบบ Hybrid SARIMAGASVR ชุดที่ 2 คือ ข้อมูลการทดสอบ (Testing dataset) ตั้งแต่เดือน มกราคม 2563 ถึงเดือนพฤศจิกายน 2564 จำนวน 23 ค่า สำหรับเปรียบเทียบความแม่นยำของตัวแบบการพยากรณ์ โดยใช้เกณฑ์ร้อยละของความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) ผลการวิจัยพบว่า ตัวแบบ Hybrid SARIMAGASVR ซึ่งเป็นตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression :SVR) มีค่าความแม่นยำมากกว่าตัวแบบ SARIMA โดยมีค่า MAPE เท่ากับ 0.02765259 ดังนั้น สรุปได้ว่า ตัวแบบผสมที่นำเสนอมีความเหมาะสมที่จะนำไปพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย

คำสำคัญ : ราคาข้าวโพดเลี้ยงสัตว์ ; ตัวแบบผสม ; ตัวแบบ SARIMA ; ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน

Abstract

This research aimed to create and select a forecast model suitable for the time series data on the monthly price of maize in Thailand. The data were obtained from the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives from January 1999 to November 2021, a total of 275 values. The data were divided into two sets. The first set consisted of 252 values of the training dataset from January 1999 to December 2019. This data set was used for creating the forecast model using the SARIMA model and the Hybrid SARIMAGASVR model. The second set consisted of 23 values of the testing dataset from January 2020 to November 2021. It was used for comparing the accuracy of the forecast model using the Mean Absolute Percentage Error (MAPE). The results showed that the Hybrid SARIMAGASVR model, which was a hybrid of the SARIMA model and the Support Vector Regression model (SVR), was more accurate than the SARIMA model, with a MAPE value of 0.02765259. Therefore, it can be concluded that the hybrid model proposed in this research is suitable for forecasting the time series data on the monthly price of maize in Thailand.

Keywords : price of maize ; hybrid model ; SARIMA model ; Support Vector Regression model

***Corresponding author. E-mail** : thanakon.s@ubru.ac.th

บทนำ

ปัจจุบันภาคการเกษตรของประเทศไทยมีบทบาทสำคัญต่อเศรษฐกิจของประเทศอย่างมาก ข้อมูลจากการทำสำมะโนการเกษตรของสำนักงานสถิติแห่งชาติ พบว่า ประเทศไทยมีจำนวนประชากรในภาคการเกษตรมากถึง 10.52 ล้านคนของประชากรทั้งหมด นับว่าเป็นแหล่งรองรับแรงงานที่มีขนาดใหญ่ที่สุด และภาคการเกษตรสามารถสร้างรายได้ต่อระบบเศรษฐกิจของประเทศประมาณร้อยละ 8.631 ของผลิตภัณฑ์มวลรวมของประเทศ โดยในปี 2563 ประเทศไทยได้รับการจัดลำดับเป็นผู้ผลิตและส่งออกอาหารในอันดับที่ 11 ของโลก ผลิตภัณฑ์จากไก่เป็นหนึ่งในสินค้าทางการเกษตรที่ประเทศไทยสามารถผลิตและส่งออกได้ในปริมาณมาก โดยปี 2563 มีมูลค่าการส่งออกเนื้อไก่และผลิตภัณฑ์รวม 103,891 ล้านบาท มีปริมาณการผลิตรวมทั้งประเทศอยู่ที่ 1,609,497,418 ตัว เมื่อความต้องการผลิตภัณฑ์ไก่จำนวนมาก การผลิตเพื่อให้เพียงพอต่อความต้องการจึงเพิ่มขึ้นตามไปด้วย ส่งผลทำให้เกิดความต้องการวัตถุดิบสำหรับผลิตอาหารสัตว์เพื่อป้อนเข้าสู่อุตสาหกรรมการเลี้ยงไก่จำนวนมาก เช่นกัน (Sujjaviriyasup, 2018) ข้าวโพดเลี้ยงสัตว์เป็นหนึ่งในวัตถุดิบที่สำคัญของกระบวนการผลิตไก่ เนื่องจากเป็นแหล่งพลังงานที่สำคัญและยังมีแคโรทีนซึ่งเป็นสารที่ช่วยให้สีเนื้อไก่และไข่แดงเข้มขึ้นตามความนิยมของผู้บริโภค การผลิตข้าวโพดเลี้ยงสัตว์ให้สอดคล้องกับอุปสงค์และอุปทานเป็นความท้าทายอย่างยิ่ง เนื่องจากผลผลิตในแต่ละปีมีความผันผวนไม่แน่นอนซึ่งมีความสอดคล้องสัมพันธ์กับพื้นที่เพาะปลูก ประกอบกับการปลูกข้าวโพดของเกษตรกรต้องอาศัยน้ำฝนเป็นหลัก ด้วยเหตุนี้จึงทำให้เกษตรกรไม่สามารถควบคุมปริมาณการผลิตให้สอดคล้องกับความต้องการได้ และยังพบว่า ราคาผลผลิตเป็นสาเหตุหลักที่ทำให้เกษตรกรตัดสินใจปลูกข้าวโพดเลี้ยงสัตว์ (Riansut & Thongrit, 2017) แต่ราคาข้าวโพดเลี้ยงสัตว์ในแต่ละเดือนมีลักษณะไม่แน่นอน (OFFICE OF AGRICULTURAL ECONOMICS MINISTRY, 2021) การพยากรณ์ด้วยวิธีการอย่างง่ายอาจทำให้เกิดความคลาดเคลื่อนได้ เมื่อไม่สามารถพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ที่แม่นยำได้ ทำให้เกษตรกรผู้ผลิตข้าวโพดเลี้ยงสัตว์ประสบปัญหาการวางแผนการผลิตเพื่อให้สามารถลดต้นทุนสำหรับการบริหารจัดการและการวางแผนการผลิตให้เกิดประสิทธิภาพ ด้วยเหตุนี้ การพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ที่มีการเปลี่ยนแปลงในอนาคตจึงเป็นสารสนเทศที่สำคัญสำหรับการวางแผนการผลิตที่มีประสิทธิภาพ (Sujjaviriyasup, 2018)

ปัจจุบันตัวแบบการพยากรณ์ข้อมูลอนุกรมเวลาเป็นเครื่องมือหนึ่งที่ได้รับความนิยมและยอมรับว่ามีประสิทธิภาพที่นำมาใช้ในแก้ปัญหาและการพยากรณ์ข้อมูลในอนาคตเพื่อช่วยในการตัดสินใจ (Li *et al.*, 2021;Yang *et al.*, 2020; Liu *et al.*, 2020; Huang *et al.*, 2021; Cheng et al., 2021; Zhu *et al.*, 2019) ถึงแม้จะมีข้อถกเถียงเกี่ยวกับประสิทธิภาพของ ตัวแบบการพยากรณ์สำหรับการวางแผนการตัดสินใจต่าง ๆ (Sujjaviriyasup, 2018) อย่างไรก็ตามมีนักวิชาการจำนวนมากสนับสนุนให้ใช้งานตัวแบบการพยากรณ์ เพื่อประกอบการวางแผนการตัดสินใจในเหตุการณ์ต่าง ๆ โดยอาศัยข้อมูลและประสบการณ์ในอดีตรวมถึงข้อมูลในปัจจุบันเพื่อพยากรณ์ข้อมูลที่จะเกิดขึ้นในอนาคต (de Oliveira *et al.*, 2020) อย่างน้อยทำให้ได้สารสนเทศที่เป็นประโยชน์และช่วยในการตัดสินใจได้ จากข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนที่มีการเปลี่ยนแปลงตามเวลาและมีความไม่แน่นอนแทรกอยู่ การหาตัวแบบที่เหมาะสมและมีประสิทธิภาพมีความแม่นยำในการพยากรณ์สูงจึงมีความท้าทายอย่างมากหากได้ตัวแบบที่มีเหมาะสม สามารถนำตัวแบบดังกล่าวพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ที่แม่นยำเพื่อนำมาประกอบการวางแผนลดต้นทุนการผลิต รวมทั้งบริหารจัดการการผลิตที่มีประสิทธิภาพได้ ด้วยเหตุนี้ นักวิชาการจำนวนมากพยายามเสนอตัวแบบการพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์อย่าง

หลากหลาย โดยตัวแบบการพยากรณ์ที่มีการประยุกต์ใช้สำหรับการพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์สามารถแบ่งได้เป็น 2 กลุ่มใหญ่ คือ (1) ตัวแบบการพยากรณ์เชิงเส้นตรง (Linear approach) ได้แก่ ตัวแบบ ARIMA และตัวแบบ SARIMA เป็นตัวแบบที่มีประสิทธิภาพสูงใช้ได้ดีในกรณีที่ข้อมูลอนุกรมเวลามีลักษณะเชิงเส้นตรง เนื่องจากการสร้างตัวแบบเกิดจากจากสมมติฐานของความสัมพันธ์แบบเชิงเส้นตรงของข้อมูลอนุกรมเวลา ด้วยเหตุนี้ก่อนนำตัวแบบไปใช้งานต้องมีการตรวจสอบข้อตกลงเบื้องต้นทางสถิติทุกครั้งเพื่อให้สอดคล้องสัมพันธ์กับแนวคิดการสร้างตัวแบบ จึงกล่าวได้ว่าวิธีการนี้มีข้อจำกัดทางสถิติค่อนข้างมาก และมีข้อเสียคือไม่สามารถพยากรณ์ข้อมูลอนุกรมเวลาที่มีลักษณะไม่เป็นเส้นตรงได้ ปัจจุบันมีการประยุกต์ใช้ตัวแบบ ARIMA และ SARIMA สำหรับพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย (Riansut & Thongrit, 2017) (2) ตัวแบบการพยากรณ์ไม่เป็นเชิงเส้นตรง (Non - Linear approach) วิธีพยากรณ์นี้มีความยืดหยุ่นและมีความแม่นยำสูงรวมทั้งไม่มีข้อจำกัดเกี่ยวกับข้อตกลงเบื้องต้นทางสถิติใช้ได้ดีกับข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง แต่ตัวแบบนี้มีข้อเสีย คือ ไม่สามารถพยากรณ์ข้อมูลอนุกรมเวลาที่มีลักษณะเป็นเชิงเส้นตรงได้ วิธีการพยากรณ์ตามแนวทางนี้ เช่น ตัวแบบโครงข่ายประสาทเทียม (Artificial Neural Network : ANN) และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) เป็นต้น ปัจจุบันมีการนำตัวแบบ ANN ไปประยุกต์ใช้สำหรับพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยซึ่งพบว่ามีความแม่นยำมากกว่าตัวแบบ SARIMA (Sujjaviriyasup, 2018)

จากข้อค้นพบดังกล่าวข้างต้น พบว่าตัวแบบการพยากรณ์ทั้งแบบเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงต่างมีข้อดีและข้อด้อยแตกต่างกันหากใช้วิธีการพยากรณ์วิธีใดวิธีหนึ่งพยากรณ์ข้อมูลอนุกรมเวลาซึ่งมักมีส่วนประกอบทั้งเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงอาจไม่เหมาะสมและมีความคลาดเคลื่อนสูง ทั้งนี้เพราะตัวแบบอนุกรมเวลาไม่สามารถอธิบายองค์ประกอบของข้อมูลอนุกรมเวลาได้ครบถ้วน เพื่อแก้ปัญหานี้ Zhang (2003) จึงนำเสนอตัวแบบพยากรณ์ข้อมูลอนุกรมเวลาแบบใหม่โดยใช้ชื่อเรียกว่า ตัวแบบผสม (Hybrid model) ตัวแบบนี้มีจุดเด่น คือ รวมจุดเด่นของตัวแบบการพยากรณ์ที่เป็นเชิงเส้นตรงและตัวแบบการพยากรณ์ไม่เชิงเส้นตรงเข้าด้วยกัน มีหลักการคือจะแยกการวิเคราะห์ข้อมูลอนุกรมเวลาออกเป็น 2 ขั้นตอน คือ ขั้นตอนแรกทำการพยากรณ์ข้อมูลอนุกรมเวลาที่เป็นเชิงเส้นตรงด้วยตัวแบบ ARIMA จากนั้นนำค่าพยากรณ์จากตัวแบบอนุกรมเวลาดังกล่าวลบด้วยค่าจริงทำให้ได้ข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง ขั้นตอนที่สองนำข้อมูลไม่เป็นเชิงเส้นตรงพยากรณ์ด้วยตัวแบบ ANN แล้วนำค่าพยากรณ์ทั้งสองมารวมกันในขั้นตอนสุดท้าย ผลการวิจัยบ่งชี้ว่า ตัวแบบผสมมีแม่นยำมากกว่าตัวแบบ ARIMA และตัวแบบ ANN เพียงตัวเดียว ด้วยเหตุนี้มีการนำแนวคิดของ Zhang ไปประยุกต์ใช้สำหรับพยากรณ์ข้อมูลอนุกรมเวลาอย่างกว้างขวาง (Liu *et al.*, 2021) และผลการวิจัยให้ผลสอดคล้องกับแนวคิดของ Zhang กล่าวคือ ตัวแบบผสมมีความแม่นยำมากกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยว ปัจจุบันมีการพัฒนาตัวแบบผสมขึ้นเพื่อพยากรณ์ข้อมูลอนุกรมเวลาอย่างต่อเนื่องมีการนำเสนอตัวแบบผสมแบบใหม่เพื่อให้ได้ค่าพยากรณ์ที่มีความแม่นยำสูง เช่น ตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) (Xu *et al.*, 2019) เป็นต้น

จากเหตุผลและความสำคัญดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะเสนอตัวแบบแบบผสมเพื่อพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย โดยจะขยายแนวคิดของ Zhang โดยการปรับเปลี่ยนตัวแบบการพยากรณ์ไม่เป็นเชิงเส้นตรง

จากตัวแบบ ANN เป็นตัวแบบ SVR เพื่อสร้างทางเลือกที่หลากหลายสำหรับการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ซึ่งตัวแบบ SVR เป็นหนึ่งในตัวแบบการเรียนรู้ของเครื่องแบบมีผู้สอน ปัจจุบันมีการประยุกต์ใช้อย่างกว้างขวาง (Fan *et al*., 2021) การใช้ตัวแบบ SVR จะต้องมีการกำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับตัวแบบเพื่อให้ได้ค่าพยากรณ์มีความแม่นยำ ปัจจุบันมีการประยุกต์ใช้อัลกอริทึมเพื่อใช้ในการกำหนดค่าพารามิเตอร์ดังกล่าว เช่น วิธีการเชิงพันธุกรรม (Genetic Algorithm : GA) เป็นอัลกอริทึมหนึ่งที่มีประสิทธิภาพและมีการนำประยุกต์ใช้สำหรับกำหนดค่าพารามิเตอร์ให้กับตัวแบบ SVR (Chen *et al*., 2015) และจะนำมาประยุกต์ใช้ในครั้งนี้ด้วย ผลการวิจัยที่ได้สามารถใช้เป็นแนวทางสำหรับการเลือกใช้ตัวแบบการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยต่อไป

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ดำเนินการสร้างตัวแบบการพยากรณ์โดยใช้ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย (บาทต่อกิโลกรัม) ตั้งแต่เดือนมกราคม 2542 ถึงเดือนพฤศจิกายน 2564 จำนวน 275 ค่า จากสำนักงานเศรษฐกิจการเกษตร (OFFICE OF AGRICULTURAL ECONOMICS MINISTRY, 2021) ผู้วิจัยแบ่งข้อมูลออกเป็น 2 ชุด ชุดที่ 1 คือ ข้อมูลการเรียนรู้ (Training dataset) ตั้งแต่เดือนมกราคม 2542 ถึงเดือน ธันวาคม 2562 จำนวน 252 ค่า ใช้สำหรับการสร้างตัวแบบการพยากรณ์ ชุดที่ 2 คือ ข้อมูลการทดสอบ (Testing dataset) ตั้งแต่เดือน มกราคม 2563 ถึงเดือนพฤศจิกายน 2564 จำนวน 23 ค่า เกณฑ์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพการพยากรณ์ (Cross -Validation) ดังกล่าวใช้เกณฑ์ร้อยละ 91 และ ร้อยละ 9 ทั้งนี้เนื่องจากผู้วิจัยเก็บข้อมูลเป็นรายเดือนจำนวนคาบเวลาใน 1 ฤดูกาลเท่ากับ 12 เดือน ผู้วิจัยจึงแบ่งข้อมูลตามสัดส่วนดังกล่าว โดยมีวัตถุประสงค์เพื่อให้ครอบคลุมจำนวนคาบเวลาของฤดูกาล และจากการศึกษางานวิจัยของ Sujjaviriyasup (2018) ได้สร้างตัวแบบโครงข่ายประสาทเทียมสำหรับพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ใช้ข้อมูลรายเดือนจำนวน 227 ค่า แบ่งออกเป็น 2 ชุด คือ ชุดที่ 1 ข้อมูลจำนวน 216 ค่า สำหรับสร้างตัวแบบ และ ข้อมูลชุดที่ 2 จำนวน 11 ค่า ใช้สำหรับทดสอบตัวแบบ และการศึกษาของ Riansut & Thongrit (2017) การพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ของประเทศไทย ใช้ข้อมูลรายเดือนจำนวน 227 เดือน โดยแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดที่ 1 จำนวน 216 เดือน ใช้สำหรับสร้างตัวแบบการพยากรณ์ ชุดที่ 2 จำนวน 11 เดือน ใช้สำหรับเปรียบเทียบความแม่นยำของตัวแบบการพยากรณ์ เพื่อให้ใกล้เคียงกับงานวิจัยที่ศึกษา ผู้วิจัยจึงแบ่งข้อมูลตามสัดส่วนดังกล่าวข้างต้น วิธีการสร้างตัวแบบและวิธีการเปรียบเทียบตัวแบบการพยากรณ์ มีรายละเอียดดังนี้

1. ตัวแบบอารีมาแบบมีฤดูกาล (Seasonal autoregressive integrated moving average: SARIMA)

ตัวแบบ SARIMA เป็นตัวแบบการพยากรณ์ที่มีความแม่นยำสูงพัฒนาขึ้นโดยบ็อกซ์ –เจนกินส์ (Box *et al*.,2008) มีสมการทางคณิตศาสตร์ดังสมการที่ (1) – (6)

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (1)$$

$$\delta = \mu\phi_p(B)\Theta_P(B^s) \quad (2)$$

$$\phi_p(B) = 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p \quad (3)$$

$$\Phi_P(B^s) = 1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps} \quad (4)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q \quad (5)$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - ... - \Theta_Q B^{Qs} \quad (6)$$

เมื่อ Y_t แทน อนุกรมเวลา ณ เวลา t , ε_t แทน ความคลาดเคลื่อนมีการแจกแจงปรกติ มีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่และเป็นอิสระกัน $(\varepsilon_t \sim NID(0, \sigma^2))$, t แทน คาบเวลาซึ่งมีค่าตั้งแต่ 1 ถึง n , s แทน ความยาวของคาบฤดูกาลใน 1 ฤดูกาล, d และ D แทน ลำดับที่ของการหาผลต่างภายในฤดูกาลและผลต่างระหว่างฤดูกาล, B แทน ตัวดำเนินการถอยหลัง (Backward operator) โดยที่ $B^s Y_t = Y_{t-s}$ และ $BY_t = Y_{t-1}$, $\phi_p(B)$ แทน ตัวดำเนินการสหสัมพันธ์ในตนเองแบบไม่มีฤดูกาลอันดับที่ p , $\Phi_P(B^s)$ แทน ตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาลอันดับที่ P , $\theta_q(B)$ แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่ภายในฤดูกาลอันดับที่ q และ $\Theta_Q(B^s)$ แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่ระหว่างฤดูกาลอันดับที่ Q

การนำข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ไปสร้างตัวแบบอารีมาแบบมีฤดูกาลจะต้องทำการกำหนดรูปแบบ SARIMA (p,d,q)(P,D,Q)$_s$ การดำเนินการผู้วิจัยเขียนคำสั่งโดยใช้โปรแกรมภาษาอาร์มีฟังก์ชัน auto.arima() ซึ่งอยู่ใน Package "forecast" ที่ใช้ในการระบุรูปแบบ ARIMA(p,d,q)(P,D,Q)s ที่เหมาะสม ขั้นตอนการสร้างตัวแบบ มี 4 ขั้นตอน ได้แก่ คือ

ขั้นตอนที่ 1 พิจารณาอนุกรมเวลาว่าคงที่หรือไม่ โดยพิจารณาจากกราฟของอนุกรมเวลาชุดนั้น หากกราฟอนุกรมเวลาชุดนั้นมีแนวโน้มเป็นส่วนประกอบ กล่าวคือ ข้อมูลลักษณะเพิ่มขึ้นหรือลดลงอย่างเป็นสัดส่วนกับเวลาและมีการกระจายไม่สม่ำเสมอมีลักษณะบานออกเป็นปากแตร แสดงว่า ข้อมูลอนุกรมเวลาชุดนั้นมีสภาวะไม่คงที่ งานวิจัยนี้ทดสอบความไม่คงที่ของข้อมูลด้วย Augmented Dickey – Fuller (ADF) หากผลการทดสอบไม่มีนัยสำคัญทางสถิติแสดงว่าข้อมูลอนุกรมเวลาชุดนั้นมีลักษณะไม่คงที่ เมื่อทราบว่าข้อมูลมีลักษณะไม่คงที่ต้องทำการแปลงข้อมูลให้มีลักษณะคงที่ โดยการหาลอการิทึมของข้อมูลและหาผลต่างของข้อมูลชุดดังกล่าว เพื่อทำให้ข้อมูลอนุกรมเวลามีลักษณะคงที่ก่อนนำไปกำหนดตัวแบบ

ขั้นตอนที่ 2 กำหนดตัวแบบการพยากรณ์

ขั้นตอนที่ 3 ประมาณค่าพารามิเตอร์และตรวจสอบความเหมาะสมของตัวแบบ เมื่อได้ตัวแบบที่เหมาะสมก่อนที่จะนำตัวแบบไปใช้ในการพยากรณ์ต้องมีการตรวจสอบตัวแบบ (Diagnostics Checking) ซึ่งมีทั้งหมด 4 ข้อ คือ 1) ทดสอบความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ โดยใช้สถิติทดสอบ t-test 2) ทดสอบความคลาดเคลื่อนมีการแจกแจงปรกติโดยใช้สถิติทดสอบ Kolmogorov – Smirnov 3) ทดสอบความเป็นอิสระกันของความคลาดเคลื่อนโดยใช้สถิติทดสอบ Ljung – Box Q- statistics และ 4) ความคลาดเคลื่อนมีความแปรปรวนคงที่โดยพิจารณาจากการพล็อตกราฟดูลักษณะการกระจายระหว่างความคลาดเคลื่อนกับค่าพยากรณ์หากกราฟมีการกระจายไม่มีรูปแบบ แสดงว่าความคลาดเคลื่อนมีความแปรปรวนคงที่

ขั้นตอนที่ 4 นำตัวแบบที่เหมาะสมไปใช้ในการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยต่อไป

2. ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support vector Regression: SVR)

วิธีการพยากรณ์ด้วยตัวแบบ SVR เป็นตัวแบบการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised learning) เป็นตัวแบบที่มีความยืดหยุ่นสูงใช้ได้ดีกับข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง หลักการพยากรณ์ คือ จะทำการแมปข้อมูลให้อยู่ในปริภูมิอันดับสูงและทำการวิเคราะห์รีเกรสชันแบบเชิงเส้น ซึ่งมีฟังก์ชันการประมาณค่า ดังสมการที่ (7) โดย $\phi(x)$ คือ การแมปข้อมูลอินพุตสเปซไม่เป็นเชิงเส้น (Singchai & Keeratiwintakorn, 2014)

$$f(x) = \langle w, \phi(x) \rangle + b \tag{7}$$

เมื่อ w น้ำหนักของเวกเตอร์ และ b คือ ความคลาดเคลื่อนของเส้นถดถอย หลักการในการหาระนาบเกินที่ใช้แทนกลุ่มของข้อมูลจะทำการค้นหาค่าของ w และ b ซึ่งสามารถกำหนดได้โดย Regression Risk ดังสมการที่ (8), (9)

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} w^T w + c \sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{8}$$

$$subject\ to \begin{cases} w^T \phi(x_i) + b - y_i \le \varepsilon + \xi_i \\ y_i - w^T \phi(x_i) - b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0,\ i = 1, 2, ..., N \end{cases} \tag{9}$$

เมื่อ $c > 0$ และ $\varepsilon > 0$ โดยที่ c เป็นค่าคงที่ที่กำหนดค่า Tradeoff ระหว่างสมการ $f(x)$ ที่ได้กับค่าความถูกต้องของข้อมูลการเรียนรู้ (Training data) ค่าของ x_i ส่วนใหญ่จะอยู่ในช่วงของ ε - tube ส่วนค่า ξ_i คือ ค่าของความผิดพลาดในการฝึกสอนข้อมูลที่ต่ำไป และ ξ_i^* คือ ค่าความผิดพลาดในการฝึกสอนข้อมูลที่สูงไป การแก้ปัญหาของสมการที่ (8) โดยมีเงื่อนไขตามสมการที่ (9) ปัญหานี้สามารถแก้ได้โดยแปลงเป็นปัญหาควบคู่ (Dual Problem) ด้วยตัวคูณลากรองจ์ (Lagrange multipliers) แสดงดังสมการที่ (10) และ เงื่อนไขข้อบังคับดังสมการที่ (11)

$$\min_{\alpha_i,\alpha_i^*} \frac{1}{2} \sum_{i,j=1}^{N} y_i(\alpha_i, \alpha_i^*)(\alpha_j, \alpha_j^*) k(x_i, x_j) + \sum_{i=1}^{N} ((\varepsilon - y_i)\alpha_i + (\varepsilon + y_i)\alpha_i^*) \tag{10}$$

$$subject\ to \begin{cases} \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) = 0 \\ 0 \le \alpha_i \le c, i = 1, 2, ..., N \\ 0 \le \alpha_i^* \le c, i = 1, 2, ..., N \end{cases} \tag{11}$$

เมื่อ x_i, x_j คือ ข้อมูลอินพุต α_i, α_i^* คือ ตัวคูณลากรองจ์ C คือ ค่าคงที่ N คือ จำนวนซัพพอร์ตเวกเตอร์ซึ่งมีอินพุตเป็นซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* > 0$ ส่วนอินพุตเวกเตอร์ที่มิใช่ซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* = 0$ และหลังจากที่คำนวณค่า α_i และ α_i^* จากชุดข้อมูลการเรียนรู้ ดังนั้น เราสามารถสร้างสมการ SVR เพื่อใช้ทำนายค่าเอาท์พุตจากอินพุตเวกเตอร์แสดงดังสมการที่ (12)

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{12}$$

เมื่อ α_i และ α_i^* คือ ตัวคูณลากรองจ์, $k(x_i, x)$ คือ เคอร์เนลฟังก์ชันที่ใช้ในการแปลงข้อมูลให้มีมิติที่สูงขึ้น โดยเคอร์เนลฟังก์ชันที่นิยมใช้ในตัวแบบ SVR มีหลากหลายวิธี ในงานวิจัยนี้จะประยุกต์ใช้เกาส์เซียนเคอร์เนล (Gaussian(RBF) kernel) : $k(x_i, x) = \exp(-\frac{\|x_i.x\|^2}{2\sigma^2})$ เนื่องจากทดลองใช้กับข้อมูลแล้ว พบว่า มีความแม่นยำมากที่สุด

3. วิธีเชิงพันธุกรรม (Genetic Algorithm: GA)

วิธี GA เป็นส่วนหนึ่งของวิธีเมต้าฮิวริสติกส์ที่ใช้ในการค้นหาพารามิเตอร์ที่เหมาะสมสำหรับตัวแบบการพยากรณ์ โดยเฉพาะวิธีพยากรณ์ด้วยวิธีการเรียนรู้โดยใช้เครื่อง (Kao *et al.,* 2020) เพื่อเพิ่มความแม่นยำให้ค่าพยากรณ์ วิธี GA เป็นแนวคิดที่เสนอโดย Holland (1975) การหาคำตอบที่เหมาะสมที่สุดเริ่มต้นจากการสุ่มกลุ่มของคำตอบ (Solution) มาเป็นสมาชิกของประชากรรุ่นแรก (Initial population) แต่ละคำตอบจะถูกเข้ารหัสและ เรียกว่า โครโมโซม (Chromosome) แต่ละโครโมโซมประกอบไปด้วยยีน (Gene) ซึ่งเป็นพารามิเตอร์ที่ต้องการหาค่าที่เหมาะสมที่สุด จากนั้นคำตอบจากประชากรรุ่นก่อนหน้าหรือ เรียกว่า รุ่นพ่อแม่ (Parent) จะถูกนำไปสร้างเป็นประชากรรุ่นถัดไปหรือเรียกว่า รุ่นลูกหลาน (Offspring) การดำเนินการแบบจีเนติก (Genetic operator) สามารถแบ่งเป็น 2 ประเภท คือ การข้ามสายพันธุ์ (Crossover) และการกลายพันธุ์ (Mutation) จากนั้นจะทำการแทนที่ (Replacement) ประชากรรุ่นพ่อแม่ด้วยประชากร รุ่นลูกหลาน และกระบวนการจะถูกกระทำซ้ำจนกว่าจะเข้าสู่เงื่อนไขสิ้นสุดการทำงาน การสร้างสมการพยากรณ์ด้วยวิธี SVR มีพารามิเตอร์ที่สำคัญอยู่ 3 ตัว ซึ่งมีผลต่อความแม่นยำของการพยากรณ์ ได้แก่ พารามิเตอร์ C พารามิเตอร์ σ และพารามิเตอร์ ε ซึ่งกระบวนการค้นหาพารามิเตอร์ที่เหมาะสมของวิธี SVR ด้วยวิธี GA สามารถดำเนินการ ได้ดังนี้

3.1 กำหนดโครโมโซม (Chromosome representation) เป็นขั้นตอนของการระบุรูปแบบของยีน (Gene) ภายใต้โครโมโซม เมื่อนำมาใช้กับการค้นหาพารามิเตอร์ที่เหมาะสมของวิธี SVR แต่ละโคโมโซมประกอบด้วยยีนที่เป็นตัวแทนของพารามิเตอร์ โดยมีการเข้ารหัสโครโมโซมเป็นค่าจำนวนจริง (Floats chromosome) จากนั้นสร้างประชากรรุ่นแรก (Initial population) แบบสุ่มเพื่อนำไปประเมินความเหมาะสมในขั้นตอนต่อไป

3.2 ประเมินค่าความเหมาะสม (Fitness evaluation) เป็นการประเมินค่าความเหมาะสมของแต่ละโครโมโซม เพื่อใช้ในการคัดเลือกโครโมโซมลูกหลานที่จะไปเป็นประชากรในรุ่นถัดไป

3.3 ดำเนินการคัดเลือก (Selection operation) เป็นขั้นตอนในการคัดเลือกโครโมโซมลูกหลาน เพื่อไปเป็นประชากรรุ่นถัดไป

3.4 การข้ามสายพันธุ์ (Crossover) เป็นการสร้างโครโมโซมลูกหลาน ด้วยการนำโคโมโซมพ่อแม่จำนวน 2 โครโมโซม มาดำเนินการข้ามสายพันธุ์ด้วยอัตราการข้ามสายพันธุ์ที่กำหนด

3.5 การกลายพันธุ์ (Mutation) เป็นการสร้างโครโมโซมลูกหลาน ด้วยการนำโครโมโซมพ่อแม่จำนวน 1 โครโมโซม มาดำเนินการกลายพันธุ์

3.6 การแทนที่ประชากรเดิม (Replacement) เป็นการแทนที่ประชากรเดิมด้วยประชากรใหม่ที่มีความเหมาะสมดีขึ้น ทำซ้ำ *ขั้นตอนที่ 3.2 ถึง 3.6 จนกว่าจะเข้าสู่เงื่อนไขสิ้นสุดการทำงานของวิธี GA*

4. ตัวแบบผสม (Hybrid model)

การสร้างตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบ SVR และใช้วิธี GA ในการกำหนดกำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับตัวแบบ เพื่อให้เข้าใจตรงกันงานวิจัยนี้ใช้ชื่อเรียกตัวแบบผสมว่า HybridSARIMAGASVR ขั้นตอนและวิธีการสร้างตัวแบบแสดงดังภาพที่ 1 มีรายละเอียดดังนี้

ขั้นตอนที่ 1 แบ่งข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ออกเป็น 2 ชุด คือ ชุดข้อมูลการเรียนรู้ และชุดข้อมูลการทดสอบ

ขั้นตอนที่ 2 นำชุดข้อมูลการเรียนรู้มาสร้างตัวแบบการพยากรณ์ ด้วยตัวแบบ SARIMA ดังสมการที่ (1) – (6) เพื่ออธิบายลักษณะข้อมูลแบบไม่เป็นเชิงเส้นตรง เมื่อได้ตัวแบบที่เหมาะสมทำการพยากรณ์ล่วงหน้าหนึ่งช่วงเวลา ดังสมการที่ (13)

$$\hat{L}_{t+1} = f(y_t, y_{t-1}, y_{t-2}, ..., y_t) \tag{13}$$

เมื่อ $\hat{L}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลอนุกรมเวลาที่มีลักษณะความสัมพันธ์เชิงเส้นตรง

ขั้นตอนที่ 3 หาค่าส่วนเหลือ จากสมการที่ 13 โดยนำค่าจริง (y_t) ลบด้วยค่าพยากรณ์ ($\hat{L}_{t+1}$) เพื่อสร้างฟังก์ชันความสัมพันธ์ไม่เป็นเชิงเส้นตรง สำหรับอธิบายลักษณะความสัมพันธ์ไม่เป็นเชิงเส้นตรงของส่วนเหลือที่ได้จากตัวแบบ SARIMA ที่เหมาะสม ดังสมการที่ 14

$$N_t = f(y_t - \hat{L}_{t+1}) \tag{14}$$

เมื่อ N_t คือ ค่าข้อมูลที่มีลักษณะความสัมพันธ์ไม่เป็นเชิงเส้นตรง เมื่อได้ข้อมูล N_t จากนั้นทำการแปลงข้อมูลให้เป็นชุดข้อมูล $D = \{(x_t, y_t)\}_{i=1}^{n}$ การแปลงข้อมูลจะทำการจัดเรียงข้อมูลอนุกรมเวลาให้อยู่ในรูปค่าสังเกตล้าหลังไป m คอลัมน์ ดังเมทริกซ์ (Matrix) ดังนี้

$$y_1, y_2, y_3, ..., y_{t-3}, y_{t-2}, y_{t-1}, y_t$$

$$\downarrow$$

$$\begin{bmatrix} y_1, & y_2, & y_3, & \cdots & y_{m-2}, & y_{m-1}, & y_m \\ y_2, & y_3, & y_4, & \cdots & y_{m-1}, & y_m, & y_{m+1} \\ y_3, & y_4, & y_5, & \cdots & y_m, & y_{m+1}, & y_{m+2} \\ \vdots & & & \vdots & & & \vdots \\ y_{t-m+1}, & y_{t-m+2}, & y_{t-m+3}, & \cdots & y_{t-2}, & y_{t-1}, & y_t \end{bmatrix}$$

จากเมทริกซ์ แสดงข้อมูลย้อนหลังตั้งแต่คอลัมน์ที่ 1 ถึง คอลัมน์ที่ $m-1$ ใช้เป็นข้อมูลป้อนเข้า (Input data) และคอลัมน์ที่ m เป็นคอลัมน์เป้าหมาย (Target data) เพื่อให้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเรียนรู้และสร้างฟังก์ชันการทำนายลักษณะของข้อมูล งานวิจัยนี้เลือกใช้จำนวนคอลัมน์ตั้งแต่ 2 ถึง 12 คอลัมน์ เนื่องจากผู้วิจัยเก็บข้อมูลมาเป็นรายเดือน

จำนวนคาบเวลาใน 1 ฤดูกาลเท่ากับ 12 เดือน ตามรอบปฏิทินหากใช้ข้อมูลย้อนหลังหลายช่วงเวลามากเกินไปค่าพยากรณ์ที่ได้อาจจะไม่เหมาะสม (Sujjaviriyasup, 2018)

ขั้นตอนที่ 4 สร้างตัวแบบ SVR จะดำเนินการเขียนโปรแกรมภาษาอาร์ซึ่งมีฟังก์ชัน SVM( ) ในแพกเกจชื่อ e1072 โดยใช้วิธี GA ค้นหาค่าของพารามิเตอร์ที่เหมาะสม มีขั้นตอนดังนี้ (ดังภาพที่ 1)

4.1 กำหนดประชากรรุ่นแรก (Initial population) ด้วยการเลือกโครโมโซม (Chromosome) เข้ามาเป็นสมาชิกของประชากรแบบสุ่ม จำนวน 200 โครโมโซม แต่ละโครโมโซมประกอบด้วยยีน (Gene) จำนวน 3 ตัว ยีนแต่ละตัวหมายถึง ค่าพารามิเตอร์ของตัวแบบ SVR (C , σ^2 และ ε) มีชนิดเป็นตัวเลขจำนวนจริง เรียกชื่อว่า โฟล์ทโครโมโซม (Floast chromosome)

4.2 ประเมินค่าความเหมาะสม (Fitness evaluation) ของโครโมโซมเพื่อใช้ในการคัดเลือกโครโมโซมลูกหลาน (Offspring) ที่จะไปเป็นประชากรรุ่นถัดไป กำหนดฟังก์ชันการประเมิน (Evaluation function) ความเหมาะสมด้วยมาตรวัดความคลาดเคลื่อน MAPE

4.3 ตรวจสอบเงื่อนไขสิ้นสุดการทำงานของวิธี GA งานวิจัยนี้กำหนดเงื่อนไขสิ้นสุดการทำงาน เมื่อสร้างประชากรได้ 100 รุ่น

4.4 ถ้ายังไม่เข้าเงื่อนไขสิ้นสุดการทำงานของวิธี GA ให้ไปดำเนินการคัดเลือก (Selection) ซึ่งเป็นขั้นตอนในการคัดเลือกโครโมโซมลูกหลานเพื่อไปเป็นประชากรในรุ่นถัดไป โดยการคัดเลือกใช้วิธีแบบวงล้อรูเล็ท (Roulette wheel) ซึ่งเป็นวิธีการคัดเลือกที่เป็นค่าเริ่มต้นของฟังก์ชัน rgba( ) ที่ใช้ในการค้นหาคำตอบตามกระบวนการของวิธี GA

4.5 ดำเนินการข้ามสายพันธุ์ (Crossover) เป็นการสร้างโครโมโซมลูกหลาน โดยนำโครโมโซมพ่อแม่ จำนวน 2 โครโมโซม มาดำเนินการข้ามสายพันธุ์

4.6 ดำเนินการกลายพันธุ์ (Mutation) ซึ่งเป็นการสร้างโครโมโซมลูกหลาน ด้วยการนำโครโมโซมพ่อแม่จำนวน 1 โครโมโซม มาดำเนินการกลายพันธุ์ กำหนดอัตราการกลายพันธุ์เท่ากับ 0.01

4.7 ดำเนินการแทนที่ประชากร (Replacement) ซึ่งเป็นการแทนที่ประชากรเดิมด้วยประชากรใหม่ที่มีความเหมาะสมที่ดีขึ้น โดยดำเนินการแทนที่โครโมโซมพ่อแม่ด้วยโครโมโซมลูกหลาน ตามอัตราการข้ามสายพันธุ์เท่ากับ 0.80

4.8 ดำเนินการสร้างตัวแบบ SVR ด้วยประชากรรุ่นถัดไปตามวิธี GA และทำซ้ำขั้นตอนที่ 2 ถึง 7 จนกระทั่งเข้าสู่เงื่อนไข สิ้นสุดการทำงานของวิธี GA

ขั้นตอนที่ 5 ได้ค่าพามิเตอร์ของวิธี SVR ที่เหมาะสมที่สุดจากวิธี GA จากนั้นนำไปสร้างตัวแบบ SVR โดยใช้ข้อมูลชุดการเรียนรู้

ขั้นตอนที่ 6 พยากรณ์ค่า N_t จำนวน 23 ค่า จากตัวแบบ SVR จากขั้นตอนที่ 5

ขั้นตอนที่ 7 นำผลการพยากรณ์ค่าข้อมูลเชิงเส้นตรงและไม่เป็นเชิงเส้นตรงมารวมกันแบบความสัมพันธ์เชิงบวก เพื่อหาค่าพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ดังสมการที่ (15)

$$\hat{y}_{t+1} = \hat{L}_{t+1} + \hat{N}_{t+1} + \varepsilon_{t+1} \tag{15}$$

โดยที่ $\hat{y}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลอนุกรมเวลา ณ เวลาที่ t+1, $\hat{L}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบเชิงเส้นตรง ณ เวลาที่ t+1, $\hat{N}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบไม่เชิงเส้นตรง ณ เวลาที่ t+1, ε_{t+1} คือ ความคลาดเคลื่อน ณ เวลาที่ t+1 จากนั้นประเมินประสิทธิภาพของตัวแบบเพื่อนำไปเปรียบเทียบกับตัวแบบดั้งเดิมต่อไป

5. เกณฑ์การประเมินประสิทธิภาพของตัวแบบ

การประเมินประสิทธิภาพของตัวแบบเพื่อวัดความผิดพลาดเชิงปริมาณ เพื่ออธิบายความสามารถในการพยากรณ์ที่ถูกต้องและเหมาะสมของข้อมูลอนุกรมเวลา ผู้วิจัยเลือกใช้เกณฑ์ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย เนื่องจากเป็นเกณฑ์ที่คิดความคลาดเคลื่อนของตัวแบบการพยากรณ์เป็นเปอร์เซ็นต์และมีการประยุกต์ใช้ในงานวิจัยอนุกรมเวลาอย่างแพร่หลาย โดยมีสูตรการคำนวณ แสดงดังสมการที่ (16)

$$MAPE = 100 \times \sum_{t=1}^{n} \left| 1 - \hat{Y}_t / Y_t \right| / n \qquad (16)$$

โดยที่ $MAPE$ ยิ่งน้อยก็ยิ่ง แสดงว่า ตัวแบบมีประสิทธิภาพสูง

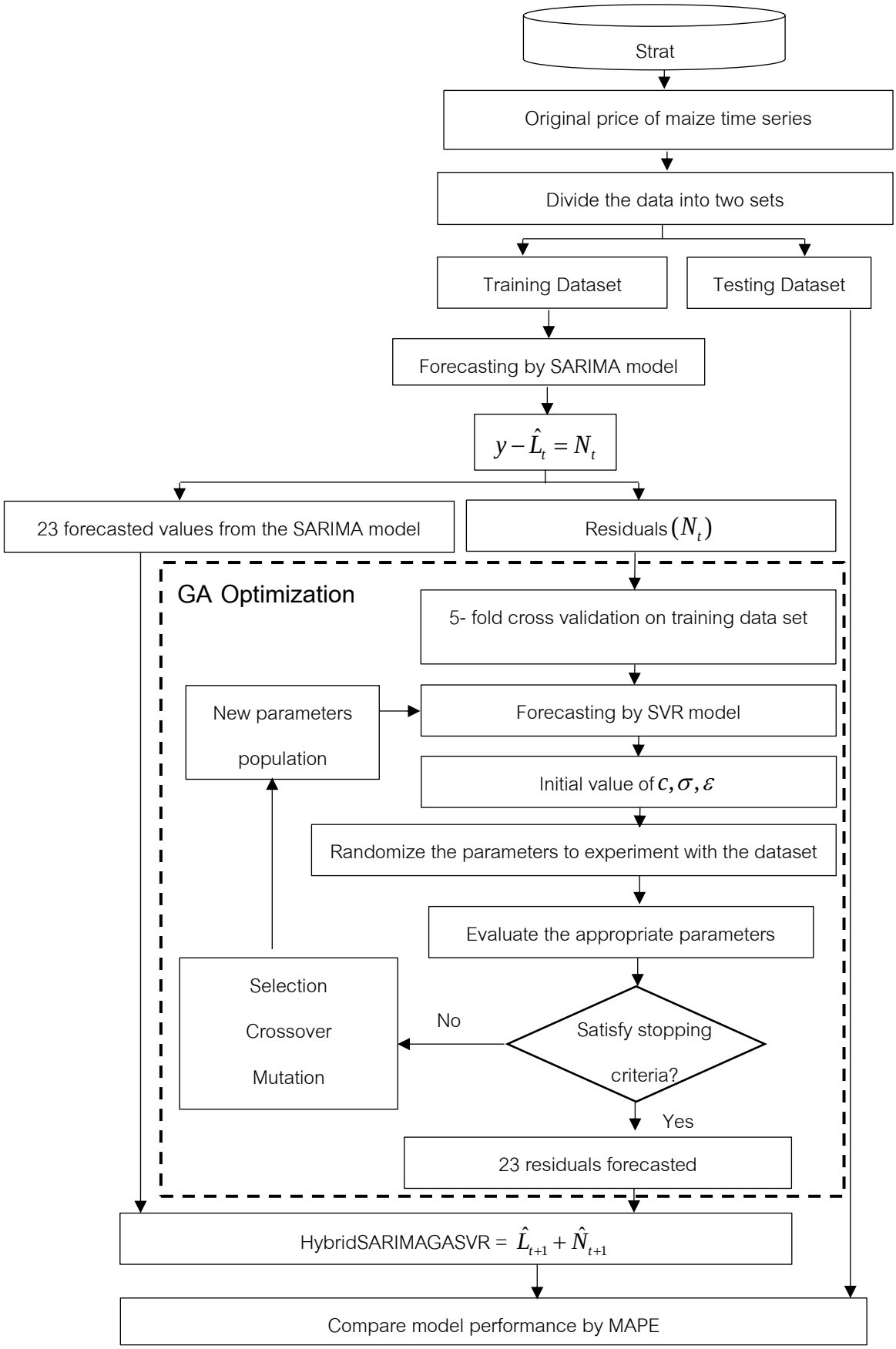

Figure 1 The Flowchart of the proposed HybridSARIMAGASVR forecasting system.

ผลการวิจัย

1. ผลการศึกษาความเคลื่อนไหวของข้อมูลอนุกรมเวลา จากการพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย (บาทต่อกิโลกรัม) ตั้งแต่เดือนมกราคม 2542 ถึงเดือน พฤศจิกายน 2564 จำนวน 275 ค่า ดังภาพที่ 3 พบว่า อนุกรมเวลาทั้ง 4 ชุด มีแนวโน้มและความผันแปรตามฤดูกาลเป็นส่วนประกอบ กล่าวคือ ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยมีแนวโน้มเพิ่มขึ้น และลดลงเมื่อเวลาเปลี่ยนแปลงไป ซึ่งเป็นผลจากฤดูกาล และมีความสอดคล้องสัมพันธ์กับกราฟแสดงความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ในตัวและระยะห่าง k (ภาพที่ 2)

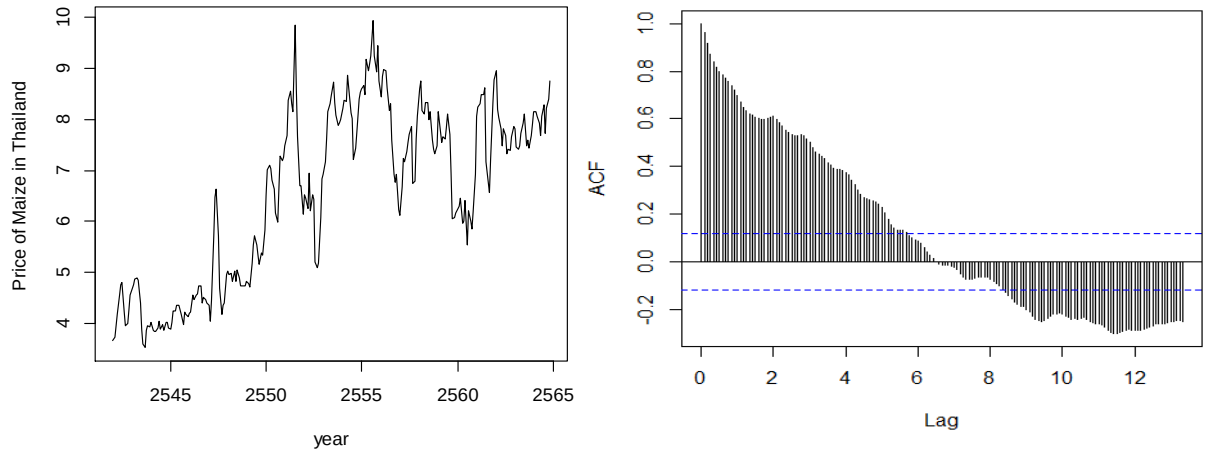

Figure 2 Data time series monthly price of maize in Thailand and autocorrelation function.

2. ผลการวิเคราะห์ข้อมูลค่าสถิติพื้นฐาน พบว่า ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย มีค่าเฉลี่ยเท่ากับ 6.531 บาทต่อกิโลกรัม ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.696 บาทต่อกิโลกรัม โดยมีราคาสูงสุดเท่ากับ 9.950 บาทต่อกิโลกรัม และต่ำสุดเท่ากับ 3.520 บาทต่อกิโลกรัม รายละเอียดดังตารางที่ 1

Table 1 Descriptive summary monthly price of maize in Thailand (Unit: Bath/ kilogram) .

Dataset	Number	Mean	Std.	Max.	Min.	Median
All sample	275	6.531	1.696	9.950	3.520	6.830
Training Dataset	252	6.407	1.716	9.950	3.520	6.595
Testing Dataset	23	7.891	0.351	8.760	7.420	7.900

3. ผลพยากรณ์โดยใช้ตัวแบบ SARIMA

3.1 จากการพิจารณาข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย มีลักษณะไม่คงที่ โดยพิจารณาจากรูปที่ 2 พบว่า กราฟมีแนวโน้มเป็นส่วนประกอบมีลักษณะการกระจายไม่สม่ำเสมอลักษณะของกราฟบานออกเป็นปากแตร แสดงว่า ข้อมูลมีลักษณะไม่คงที่ สอดคล้องกับผลการทดสอบความคงที่ของข้อมูลด้วย Augmented Dickey – Fuller (ADF = -2.4722, p-value = 0.377) สรุปได้ว่า ข้อมูลชุดนี้มีลักษณะไม่คงที่ นั่นคือ

มีค่าเฉลี่ยไม่คงที่และความแปรปรวนไม่คงที่ เมื่อทราบว่าข้อมูลมีลักษณะไม่คงที่ ผู้วิจัยจึงทำการแปลงข้อมูลโดยการหาลอการิทึมก่อนจากนั้นหาผลต่างของชุดข้อมูลอนุกรมเวลาก่อนนำไปสร้างตัวแบบ SARIMA และทำการทดสอบความคงที่ของข้อมูลด้วย Augmented Dickey – Fuller (-7.6486, p-value = 0.01) ผลการทดสอบ พบว่า ข้อมูลอนุกรมเวลามีลักษณะคงที่และนำไปใช้ในการกำหนดตัวแบบในขั้นตอนต่อไป

3.2 กำหนดรูปแบบของตัวแบบ SARIMA ต้องระบุอันดับ (p,d,q)และอันดับ $(P,D,Q)_s$ ซึ่งโปรแกรมภาษาอาร์มีฟังก์ชัน auto.arima() ซึ่งอยู่ใน Package “forecast” สำหรับกำหนดอันดับ (p,d,q) และอันดับ $(P,D,Q)_s$ ที่เหมาะสม ผลจากการใช้ฟังก์ชัน auto.arima( ) กำหนดรูปแบบของตัวแบบของ SARIMA ผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย คือ $SARIMA(2,1,2)(0,0,2)_{12}$ ไม่มีพจน์ค่าคงที่ โดยมีคาบเวลา 1 ฤดูกาลเท่ากับ 12 ($s=12$) สามารถประมาค่าสัมประสิทธิ์ $AR_1(\hat{\phi}_1)$ เท่ากับ 1.2329 (z =8.6398**,p-value<0.01), $AR_2(\hat{\phi}_2)$ เท่ากับ -0.6490 (z =-4.9064**,p-value<0.01) , $MA_1(\hat{\theta}_1)$ เท่ากับ -1.0140 (z=-5.7423**, p-value<0.01), $MA_2(\hat{\theta}_2)$ เท่ากับ 0.3673 (z= 2.2854*,p-value<0.01), $SAR_1(\hat{\Phi}_1)$ เท่ากับ 0.1011 (z =2.8281**, p= < 0.01) และ $SAR_2(\hat{\Phi}_2)$ เท่ากับ 0.1778 (z = 2.7128**, p= < 0.01) ตามลำดับ

3.3 การตรวจสอบความเหมาะสมของตัวแบบ โดยพิจารณาตามข้อตกลงเบื้องต้นของตัวแบบ SARIMA ซึ่งมีทั้งหมด 4 ข้อ ได้แก่ 1) ความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ ผลการตรวจสอบ พบว่า ค่าเฉลี่ยความคลาดเคลื่อนมีค่าเท่ากับ 0 (t-test = 0.76563, p-value = 0.4446) 2) ความคลาดเคลื่อนมีการแจกแจงปรกติ พบว่า ความคลาดเคลื่อนมีการแจกแจงปรกติ (KS –value = 0.4703, p-value = 0.153) 3) ความคลาดเคลื่อนเป็นอิสระต่อกัน โดยใช้ Ljung – Box Q-statistics พบว่า ความคลาดเคลื่อนเป็นอิสระต่อกัน (Q-statistics = 21.035, p-value = 0.3349) และ 4) ความคลาดเคลื่อนมีค่าความแปรปรวนคงที่ (ดังภาพที่ 3) ซึ่งพบว่า ความคลาดเคลื่อนมีค่าความแปรปรวนคงที่ (พิจารณาจากกราฟ ACF) เนื่องจากความคลาดเคลื่อนมีการกระจายคงที่รอบค่าใดค่าหนึ่ง ดังนั้นสามารถสรุปได้ว่า ตัวแบบมีความเหมาะสมที่จะนำไปพยากรณ์

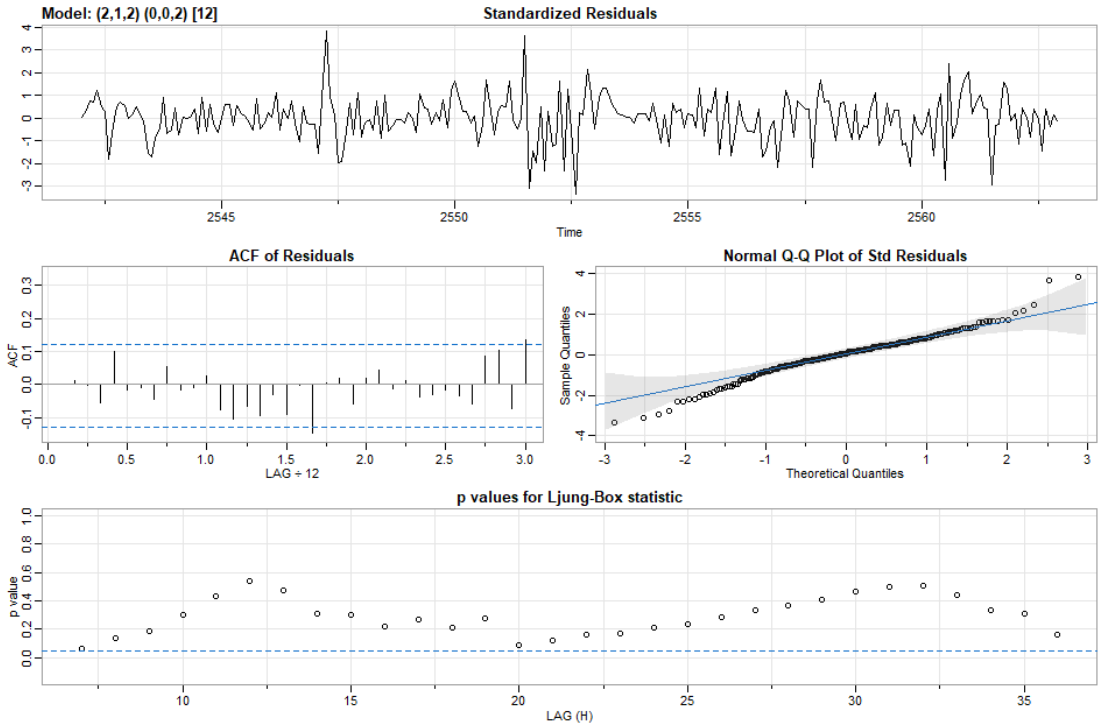

Figure 3 Examining the basic assumption of the SARIMA $(2,1,2)(0,0,2)_{12}$ model.

4. ผลพยากรณ์โดยใช้ตัวแบบ HybridSARIMAGASVR

การวิเคราะห์ข้อมูลด้วยตัวแบบ SARIMA จากขั้นตอนที่ 3 ได้ตัวแบบที่เหมาะสมคือ $SARIMA(2,1,2)(0,0,2)_{12}$ ไม่มีพจน์ค่าคงที่ ผู้วิจัยนำตัวแบบดังกล่าวพยากรณ์เพื่อหาค่าพยากรณ์ที่มีลักษณะความสัมพันธ์เชิงเส้นตรงที่เวลา t+1 ตามขั้นตอนการสร้างตัวแบบผสม จากนั้นนำค่าจริงลบด้วยค่าพยากรณ์จะได้ส่วนเหลือซึ่งเป็นส่วนที่มีลักษณะของความสัมพันธ์ไม่เป็นเชิงเส้นตรง นำค่าส่วนเหลือพยากรณ์ด้วยตัวแบบ SVR การพยากรณ์ส่วนเหลือ ผู้วิจัยจะทำการแปลงข้อมูลให้อยู่ในรูปของอินพุตเวกเตอร์และทาร์เก็ต $D = \{(x_t, y_t)\}_{i=1}^{n}$ (ตามหัวข้อการสร้างตัวแบบผสม) และแบ่งข้อมูลออกเป็น 5 ส่วน (5- fold cross validation) ใช้สำหรับสร้างและตรวจสอบตัวแบบ ขั้นตอนนี้จะทำการเขียนคำสั่งโดยใช้โปรแกรมภาษาอาร์ สุ่มข้อมูลแต่ละส่วนมาสร้างและทดสอบตัวแบบตามวิธีของตัวแบบ SVR กำหนดเคอร์เนลฟังก์ชันของตัวแบบประยุกต์ใช้เกาส์เซียนเคอร์เนล (RBF) เนื่องจากเป็นเคอร์เนลที่นิยมใช้และมีประสิทธิภาพ การใช้เกาส์เซียนเคอร์เนล (RBF) จะต้องมีการกำหนดพารามิเตอร์ 3 ตัว ได้แก่ C , σ และ ε งานวิจัยนี้จะใช้วิธี GA กำหนดค่าพารามิเตอร์ที่เหมาะสมให้กับตัวแบบ SVR โดยกำหนดค่า C อยู่ระหว่าง 1 - 3 ค่า σ อยู่ระหว่าง 0.001-0.100 และค่า ε อยู่ระหว่าง 0.001-0.100 หลักการในการหาค่าพารามิเตอร์ที่เหมาะสม จะค่อย ๆ เพิ่มค่าพารามิเตอร์ทีละตัว จากนั้นคำนวณค่าความคลาดของตัวแบบการพยากรณ์ ค่าพารามิเตอร์ที่ให้ค่าความคลาดเคลื่อนต่ำสุด คือ ค่าพารามิเตอร์ที่มีความเหมาะสมและมีประสิทธิภาพในการพยากรณ์มากที่สุด งานวิจัยนี้กำหนดการทำซ้ำทั้งหมด 1,000 ครั้ง พบว่า ได้ค่า C เท่ากับ 1.439208, σ เท่ากับ 0.001288846 และ ε เท่ากับ 0.07866982 เมื่อได้ค่าพารามิเตอร์ที่เหมาะสม ผู้วิจัยดำเนินการนำค่าพารามิเตอร์ดังกล่าวไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อน จากนั้นนำไปรวมกับค่าพยากรณ์จากตัวแบบ $SARIMA(2,1,2)(0,0,2)_{12}$ จะได้ค่าพยากรณ์

จากตัวแบบผสมตามลำดับ ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ พบว่า ตัวแบบ SARIMA$(2,1,2)(0,0,2)_{12}$ มีค่า $MAPE$ เท่ากับ 0.0382668 และตัวแบบ HybridSARIMAGASVR มีค่า $MAPE$ เท่ากับ 0.02765259 ซึ่งมีค่าน้อยกว่าตัวแบบ SARIMA ตามลำดับ ผู้วิจัยนำค่าจริงกับค่าพยากรณ์มาพล๊อตกราฟแสดงดังภาพที่ 4 จากนั้นนำตัวแบบ HybridSARIMAGASVR พยากรณ์ล่วงหน้าจำนวน 13 เดือน ตั้งแต่เดือนธันวาคม 2564 ถึงเดือนธันวาคม 2565 ดังตารางที่ 2

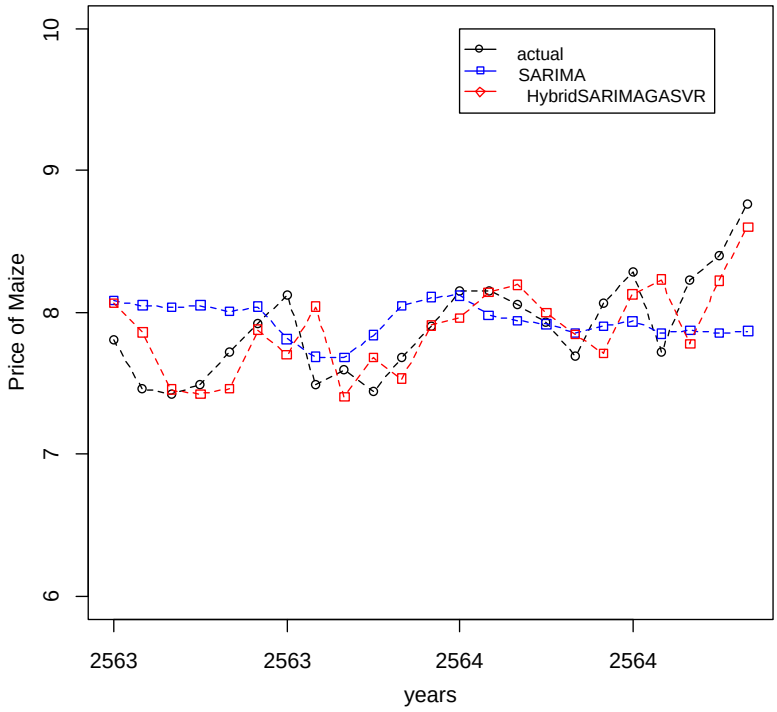

Figure 4 The forecasting results using SARIMA and HyrbridSARIMAGASVR model for monthly price of maize in Thailand.

Table 2 Thailand's monthly maize price forecasting.

Months /Year	value	Months /Year	value	Months /Year	value	Months /Year	value
Dec. 2021	8.84	Apr. 2022	7.58	Aug. 2022	7.93	Dec.2022	7.98
Jan. 2022	7.89	May. 2022	7.60	Sep. 2022	7.39		
Feb. 2022	7.76	Jun. 2022	8.15	Oct. 2022	7.71		
Mar. 2022	7.48	Jul. 2022	7.80	Nov. 2022	7.62		

วิจารณ์ผลการวิจัย

การศึกษาครั้งนี้ พบว่า วิธีการพยากรณ์ด้วยตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบ SVR มีค่าความแม่นยำมากที่สุดในการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย หรืออาจกล่าวได้ว่าวิธีการพยากรณ์แบบผสมมีประสิทธิภาพสูงที่สุด สอดคล้องกับการศึกษาของ Riansut & Thongrit (2017) วิธีการพยากรณ์รวมมีประสิทธิภาพสูงสุดในการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย นอกจากนี้ยังมีนักวิชาการหลายท่านที่ประยุกต์ใช้ตัวแบบผสมในการพยากรณ์ข้อมูลอนุกรมเวลา ผลการวิจัยให้ผลสอดคล้องกันว่า ตัวแบบผสมมีความถูกต้องแม่นยำมากกว่าตัวอนุกรมเวลาเชิงเดี่ยว (Sutthison, 2019; Xu *et al.*, 2019; Fan *et al.*, 2021; Liu *et al.*, 2021) ทั้งนี้อาจเป็นเพราะตัวแบบผสมได้รวมจุดเด่นของตัวแบบอนุกรมเวลาเชิงเดี่ยวที่พยากรณ์ข้อมูลอนุกรมเวลาเชิงเส้นตรงได้ดี และตัวแบบการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised learning) ได้แก่ ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression : SVR) เป็นตัวแบบการพยากรณ์ข้อมูลอนุกรมเวลาไม่เป็นเชิงเส้นตรงได้ดี ดังนั้น เมื่อนำตัวแบบทั้งสองมารวมกันจึงทำให้ตัวแบบมีความแม่นยำสูง (Zhang , 2003) เมื่อใช้ตัวแบบ HybridSARIMAGASVR พยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ตั้งแต่เดือน ธันวาคม 2564 ถึงเดือนธันวาคม 2565 พบว่า ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย ยังคงมีแนวโน้มลดลงและเพิ่มขึ้นเล็กน้อยตามช่วงเวลาที่เปลี่ยนแปลงไป (ตารางที่ 3) เพราะว่าราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยมีการเปลี่ยนแปลงอยู่เสมอประกอบกับข้อมูลมีแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้อง การวิจัยนี้จะเป็นประโยชน์อย่างยิ่งต่อหน่วยงานที่เกี่ยวข้องสามารถนำตัวแบบที่ผู้วิจัยพัฒนาขึ้นพัฒนาต่อยอดหรือใช้เป็นแนวทางสำหรับเป็นตัวแบบพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทยในอนาคต และสามารถนำค่าพยากรณ์ที่ได้จากตัวแบบดังกล่าวนี้ ใช้สำหรับเป็นข้อมูลเพื่อวางแผนหรือหากลยุทธวิธีการในการให้คำแนะนำเกษตรกรผู้ปลูกข้าวโพดเลี้ยงสัตว์ในอนาคต ข้อเสนอแนะในการทำงานวิจัยครั้งต่อไป เนื่องจากข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงมีความผันผวนค่อนข้างสูงหากสามารถลดความผันผวนของข้อมูลอนุกรมเวลาดังกล่าวลง และนำตัวแบบผสมที่พัฒนาขึ้นไปพยากรณ์ข้อมูลอนุกรมเวลาที่ทำการศึกษาจะทำให้ได้ค่าพยากรณ์มีความแม่นยำมากขึ้น ปัจจุบันการลดความผันผวนหรือสัญญาณแกว่งไวของข้อมูลอนุกรมเวลามีหลายวิธี เช่น Empirical Mode Decomposition, Wavelet transformation เป็นต้น

สรุปผลการวิจัย

การศึกษาครั้งนี้ ผู้วิจัยนำเสนอวิธีการสร้างและคัดเลือกตัวแบบการพยากรณ์ที่เหมาะสมกับข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์ของประเทศไทย โดยใช้ข้อมูลจากสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์ ตั้งแต่เดือนมกราคม 2542 ถึงเดือนพฤศจิกายน 2564 จำนวน 275 ค่า ด้วยตัวแบบ SARIMA และตัวแบบ HybridSARIMAGASVR ผลการวิจัยพบว่า ตัวแบบ Hybrid SARIMAGASVR เป็นตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบ SVR มีค่าความแม่นยำมากกว่าตัวแบบ SARIMA โดยมีค่า MAPE เท่ากับ 0.02765259 ดังนั้น สรุปได้ว่า ตัวแบบผสมที่นำเสนอมีความเหมาะสมที่จะนำไปพยากรณ์ข้อมูลอนุกรมเวลาราคาข้าวโพดเลี้ยงสัตว์รายเดือนของประเทศไทย

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณสาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี ที่สนับสนุนทุนการวิจัย และขอขอบคุณสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์ ที่อนุเคราะห์ข้อมูลที่ใช้ในการทดลองวิจัยครั้งนี้

เอกสารอ้างอิง

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2008). Time Series Analysis: Forecasting and Control, 4th ed. John Wiley and Sons, Inc., New Jersey.

Chen, R., Liang, C.-Y., Hong, W.-C., & Gu, D.-X. (2015). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing*, *26*, 435 – 443.

Cheng, Y., Zhu, Q., Peng, Y., Huang, X.-F., & He, L.-Y. (2021). Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models. *Journal of Cleaner Production*, *326*, 1-15.

de Oliveira, J. F. L., Pacífico, L. D. S., de Mattos Neto, P. S. G., Barreiros, E. F. S., Rodrigues, C. M. de O., & Filho, A. T. de A. (2020). A hybrid optimized error correction system for time series forecasting. *Applied Soft Computing*, *87*, 1-9.

Fan, G.-F., Yu, M., Dong, S.-Q., Yeh, Y.-H., & Hong, W.-C. (2021). Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy*, *73*,1-18.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence; University of Michigan Press: Ann Arbor, MI, USA.

Huang, X., Wang, J., & Huang, B. (2021). Two novel hybrid linear and nonlinear models for wind speed forecasting. *Energy Conversion and Management*, *238*, 1-19.

Kao, Y.-S., Nawata, K., & Huang, C.-Y. (2020). Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition. *Mathematics*, *8*(10), 1 - 19.

Li, R., Hu, Y., Heng, J., & Chen, X. (2021). A novel multiscale forecasting model for crude oil price time series. *Technological Forecasting and Social Change*, *173*, 1-15.

Liu, M.-D., Ding, L., & Bai, Y.-L. (2021). Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. *Energy Conversion and Management*, *233*, 1 - 19.

Liu, Z., Jiang, P., Zhang, L., & Niu, X. (2020). A combined forecasting model for time series: Application to short-term wind speed forecasting. *Applied Energy*, *259,* 1 - 25.

OFFICE OF AGRICULTURAL ECONOMICS MINISTRY OF AGRICULTURE AND COOPERATIVES. (2021). *Agricultural statistics of Thailand 2020*. BANGKOK: OFFICE OF AGRICULTURAL ECONOMICS.

Riansut, W., & Thongrit, K. (2017). Forecasting the Price of Field Corn in Thailand. *RUMTP Research Journal*, *11*(1), 1-14. (in Thai)

Singchai, P., & Keeratiwintakorn, P. (2014). Electricity Demand Forecast for Thailand Demand Side Management Center. *Information Technology Journal*, *10*(2), 32 - 42. (in Thai)

Sujjaviriyasup, T. (2018). Artificial Neural Network Model for Forecasting Monthly Price of Maize in Thailand. *SWU Sci.J*, *34*(1), 91-107. (in Thai)

Sutthison, T. (2019). Appropriate Models for Forecasting of Water Supply Consumption in Ubonratchathani Province. *Burapha Science Journal*, *25*(1), 28 – 50. (in Thai)

Xu, S., Chan, H. K., & Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transportation Research Part E: Logistics and Transportation Review*, *122*, 169 –180.

Yang, S., Chen, D., Li, S., & Wang, W. (2020). Carbon price forecasting based on modified ensemble empirical mode decomposition and long short-term memory optimized by improved whale optimization algorithm. *Science of The Total Environment, 716,* 1-13.

Zhang, P.G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*,159 -175.

Zhu, Q., Zhang, F., Liu, S., Wu, Y., & Wang, L. (2019). A hybrid VMD–BiGRU model for rubber futures time series forecasting. *Applied Soft Computing*, *84*, 1-12.