

Research Article

การพยากรณ์ราคาข้าวเปลือกหอมมะลิ ด้วยเทคนิคการทำเหมืองข้อมูล

Forecast of the price of jasmine rice with Data Mining Techniques

เฉลิมวุฒิ คำเมือง^{1*} และ Wachirarak Orosram¹

Chalermwut Comemuang^{1*} and Wachirarak Orosram¹

¹สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ อ.เมือง จ.บุรีรัมย์ 31000 ประเทศไทย

¹Department of Mathematics, Faculty of Science, Buriram Rajabhat University, Muang, Buriram 31000, Thailand

*E-mail: chalermwut.cm@bru.ac.th

Received: 30/03/2021; Revised: 23/09/2021; Accepted: 22/09/2021

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาแบบจำลองทางคณิตศาสตร์เพื่อทำนายราคาข้าวเปลือกหอมมะลิ ข้อมูลที่ใช้ในการศึกษาราคาข้าวเปลือกหอมมะลิเฉลี่ยรายเดือน ตั้งแต่ มกราคม พ.ศ. 2553 - ธันวาคม พ.ศ. 2563 จำนวน 132 ข้อมูล โดยแบ่งข้อมูลออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่เดือน มกราคม พ.ศ. 2553 - ธันวาคม พ.ศ. 2562 จำนวน 120 ข้อมูล สำหรับศึกษาตัวแบบพยากรณ์ โดยวิธีบ็อกซ์เจนกินส์ (Box-Jenkins) วิธีแบ็กกิ้ง (Bagging) และ วิธีแรนดอมฟอเรสต์ (Random Forest) ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคม พ.ศ. 2563 - ธันวาคม พ.ศ. 2563 จำนวน 12 ค่า สำหรับการเปรียบเทียบความแม่นยำของค่าพยากรณ์ โดยใช้เกณฑ์รากที่สองของความคลาดเคลื่อนเฉลี่ย (Root Mean Square Error: RMSE) และเกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ผลการศึกษาพบว่าจากวิธีการพยากรณ์ที่เหมาะสมที่สุดคือ วิธีแรนดอมฟอเรสต์

คำสำคัญ: พยากรณ์, บ็อกซ์เจนกินส์, แรนดอมฟอเรสต์, แบ็กกิ้ง

Abstract

The purpose of this research was to construct model for predicting the price of Jasmine Paddy price. Data used to study the average monthly price of jasmine rice from January 2010 to December 2020, 132 values, divided into two sets. The first set from January 2010 to December 2019, 120 values were used for the modeling by of box jenkins method, bagging and random forest. Another set from January 2020 to December 2020, 12 values were used for checking the accuracy of the forecasting models via the determination of the lowest root mean square error and mean absolute percentage error. The results showed that all forecasting methods studied, the random forest was the most appropriate method.

Keywords: forecast, box-jenkins, random forest, bagging

บทนำ

ข้าวเป็นธัญญาหารหลักของชาวโลก จัดเป็นพืชสายพันธุ์เดียวกับหญ้า ซึ่งนับได้ว่าเป็นหญ้าที่มีขนาดใหญ่ที่สุดในโลก และมีความหลากหลายทางชีวภาพ สามารถปลูกขึ้นได้ง่ายมีความทนทานต่อทุกสภาพภูมิประเทศในโลก ไม่ว่าจะเป็นที่ดินแห้งแล้งแบบทะเลทราย พื้นที่ราบลุ่มน้ำท่วมถึง หรือแม้กระทั่งบนเทือกเขาที่หนาวเย็น ข้าวก็ยังสามารถงอกงามขึ้นมาได้อย่างทรหดอดทน ในส่วนของประเทศไทยข้าวเป็นพืชที่เป็นอาหารประจำชาติและมีตำนานประวัติศาสตร์มายาวนาน ปรากฏเป็นร่องรอยพร้อมกับอารยธรรมไทยมาไม่น้อยกว่า 5,500 ปี และข้าวหอมมะลิก็เป็นพันธุ์ข้าวที่ได้ถือกำเนิดขึ้นในประเทศไทย โดยมีจุดเด่นคือเมล็ดมีคุณภาพดี ใส และมีกลิ่นหอมคล้ายใบเตย เวลารับประทานจะมีความนุ่มและหอมอร่อยมากกว่าข้าวทั่วไป และประเทศไทยเป็นประเทศกสิกรรม ประชาชนส่วนใหญ่เป็นกสิกร ทำการเพาะปลูกพืชสวน และพืชไร่ โดยที่ข้าวมีพื้นที่ปลูกมากกว่าพืชชนิดอื่น ๆ คิดเป็นพื้นที่ประมาณ 11.3% ของพื้นที่ทั่วประเทศและมีแนวโน้มการปลูกข้าวเพิ่มมากขึ้น ภาคกลาง และภาคตะวันออกเฉียงเหนือ มีพื้นที่ทำนามากที่สุด รองลงมา ได้แก่ ภาคเหนือ และภาคใต้ตามลำดับ

การส่งออกข้าวในช่วง มกราคม พ.ศ. 2563 – พฤศจิกายน พ.ศ. 2563 มีปริมาณ 5.24 ล้านตัน ลดลง 26.3% จากช่วงเดียวกันของปี 2562 และมีมูลค่า 106,656 ล้านบาท หรือ 3,418.8 ล้านดอลลาร์สหรัฐ ลดลง 12.1% จากช่วงเดียวกันของปีก่อน ซึ่งไทยถือเป็นผู้ส่งออกอันดับที่ 3 โดยประเทศอินเดียที่ส่งออกได้เป็นอันดับที่ 1 จำนวน 13.05 ล้านตัน และเวียดนามส่งออกได้เป็นอันดับที่ 2 จำนวน 5.79 ล้านตัน แต่ผลจากค่าเงินบาทแข็ง ทำให้ราคาข้าวไทยสูงกว่าคู่แข่ง 150 เหรียญสหรัฐต่อตัน โดยเมื่อวันที่ 23 ธันวาคม พ.ศ. 2563 ราคาข้าวขาว 5% ราคาตันละ 529 เหรียญสหรัฐ ขณะที่ราคาข้าวของเวียดนามราคาตันละ 498-502 เหรียญสหรัฐ และราคาข้าวของอินเดียราคาตันละ 368-372 เหรียญสหรัฐ (The Thai Rice Exporters Association, 2020)

ปัญหาของราคาข้าวเกิดจากหลายสาเหตุ ปัญหาด้านต้นทุนการผลิตและผลผลิตต่อไร่ซึ่งต้นทุนการผลิตต่อไร่สูงแต่ผลผลิตต่อไร่ต่ำ ระบบชลประทานในพื้นที่เพาะปลูกมีจำกัด ราคาส่งออกข้าวไทยสูงกว่าประเทศเพื่อนบ้าน ซึ่งในปัจจุบันรัฐบาลใช้วิธีการจำหน่ายข้าวหอมมะลิในราคาสูง เพื่อช่วยเหลือเกษตรกรที่ปลูกข้าวหอมมะลิ ในขณะที่เดียวกันก็ทำให้เกิดปัญหาการปลอมปนข้าว เพราะมีผู้นำข้าวบางส่วนสั่งซื้อข้าวหอมมะลิจากไทยไป แล้วนำไปผสมในต่างประเทศ และอ้างว่าเป็นข้าวหอมมะลิจากไทย ซึ่งสร้างผลเสียต่อภาพลักษณ์การส่งออกข้าวหอมมะลิไทย ดังนั้นการพยากรณ์ราคาจึงสามารถช่วยทำให้ประเมินราคาลินค้าในอนาคตได้ การศึกษาแนวโน้มและพยากรณ์ราคาข้าวในประเทศไทยจึงมีความจำเป็นอย่างมาก เพื่อที่จะนำผลการศึกษาที่ได้จากการพยากรณ์มาเป็นแนวทางในการตัดสินใจสำหรับผู้ผลิตและผู้ส่งออก ในการวางแผนการผลิตและการส่งออกให้มีประสิทธิภาพดียิ่งขึ้นและสอดคล้องกับความต้องการของตลาด

Kodosueb & Boonlha (2016) ได้ศึกษาตัวแบบที่เหมาะสมและเพื่อเปรียบเทียบตัวแบบสำหรับการพยากรณ์ราคาข้าวหอมมะลิ 105 จากข้อมูลรายเดือนราคาข้าวหอมมะลิ 105 ตั้งแต่เดือนมกราคม พ.ศ.2540 ถึงเดือนธันวาคม พ.ศ.2558 รวม 18 ปี (หรือ 216 เดือน) ซึ่งเก็บรวบรวมโดยสำนักงานเศรษฐกิจการเกษตร ศึกษาตัวแบบการพยากรณ์โดยวิธีการปรับให้เรียบแบบเอกซ์โปเนนเชียลซ้ำสองครั้ง และวิธีบ็อกซ์และเจนกินส์ และเปรียบเทียบตัวแบบสำหรับการพยากรณ์โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) น้อยที่สุด ผลการศึกษาพบว่า วิธีการพยากรณ์ที่เหมาะสมที่สุดคือ วิธีบ็อกซ์และเจนกินส์ ซึ่งได้ตัวแบบพยากรณ์ คือ ARIMA (1,1,2) Riansu (2016) การสร้างตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาราคาข้าวเปลือกเจ้าความชื้น 15% โดยใช้อนุกรมเวลาราคาข้าวเปลือกเจ้าความชื้น 15% จากเว็บไซต์ของสำนักงานเศรษฐกิจการเกษตร ตั้งแต่เดือนมกราคม 2548 ถึงเดือนกันยายน 2558 จำนวน 129 ค่าซึ่งข้อมูลถูกแบ่งออกเป็น 2 ชุด ข้อมูล ชุดที่ 1 ตั้งแต่เดือนมกราคม 2548 ถึงเดือนธันวาคม 2557 จำนวน 120 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธี บ็อกซ์เจนกินส์วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของโฮลด์วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีแนวโน้มแบบแฉก และวิธีการพยากรณ์รวม ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคมถึงเดือนกันยายน 2558 จำนวน 9 ค่านำมาใช้สำหรับการเปรียบเทียบความแม่นยำของค่าพยากรณ์โดยใช้เกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด ผลการวิจัยพบว่า จากวิธีการพยากรณ์ทั้งหมดที่ได้ศึกษา วิธีที่มีความแม่นยำมากที่สุด คือวิธีการพยากรณ์รวม Yoyram et al. (2020) ได้สร้างแบบจำลองโดยใช้เรแกนคอมโพเรสต์ และทำการเปรียบเทียบแบบจำลองโดยทำการคัดเลือกคุณลักษณะของข้อมูล 4 วิธี ได้แก่ IR IG CFS และ SU โดยใช้โปรแกรม WEKA ผลการศึกษาพบว่า โมเดลการพยากรณ์แบบเรแกนคอมโพเรสต์ เมื่อ ใช้ร่วมกับเทคนิคการลดคุณลักษณะของข้อมูลแบบ SU ทำให้ประสิทธิภาพในการพยากรณ์เพิ่มขึ้น โดยวัดค่าความถูกต้องได้ค่าร้อยละ 98.74 Wanon et al. (2018) ได้ศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาในระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูลที่เหมาะสม ทดลองวัดความแม่นยำด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีเรแกนคอมโพเรสต์ และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิ้ง ผลการวิจัยพบว่า ความแม่นยำในการจำแนกประเภทข้อมูล จาก 3 เทคนิค 1) เทคนิคการจำแนกข้อมูล

ด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 81.91% 2) เทคนิคการจำแนกข้อมูลด้วยวิธีเรนดอมฟอเรสต์ เท่ากับ 84.29% และ 3) เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิ้ง เท่ากับ 81.71% พบว่าเทคนิคเรนดอมฟอเรสต์ ให้ความถูกต้องในการจำแนกประเภทข้อมูลสูงที่สุด

เพื่อให้ได้ตัวแบบที่เหมาะสมสำหรับการพยากรณ์ราคาข้าวในประเทศไทยและสามารถนำแนวคิดในการหาตัวแบบของการพยากรณ์ไปใช้ในการพยากรณ์ราคาข้าวพันธุ์อื่นได้ เพื่อเป็นแนวทางในการตัดสินใจสำหรับผู้ผลิตและผู้ส่งออกข้าวในประเทศไทย ผู้วิจัยจึงนำข้อมูลราคาข้าวเปลือกหอมมะลิเฉลี่ยรายเดือน ตั้งแต่ มกราคม พ.ศ. 2553 - ธันวาคม พ.ศ. 2563 จำนวน 132 ข้อมูล โดยแบ่งข้อมูลออกเป็น 2 ชุด ข้อมูลชุดที่ 1 ตั้งแต่เดือน มกราคม พ.ศ. 2553 - ธันวาคม พ.ศ. 2562 จำนวน 120 ข้อมูล สำหรับศึกษาตัวแบบพยากรณ์ ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคม พ.ศ. 2563 - ธันวาคม พ.ศ. 2563 จำนวน 12 ค่า สำหรับการเปรียบเทียบความแม่นยำของค่าพยากรณ์

จากข้อมูลดังกล่าวพบว่าราคาข้าวเปลือกหอมมะลิเฉลี่ยรายเดือนมีการเปลี่ยนแปลงตามฤดูกาลซึ่งเป็นส่วนประกอบหนึ่งของอนุกรมเวลา ผู้วิจัยจึงใช้วิธีบ็อกซ์-เจนกินส์ ซึ่งเป็นการพยากรณ์โดยอาศัยแนวคิดที่ว่าอนุกรมเวลาที่พิจารณาอาจมีลักษณะของสหสัมพันธ์ในตัวเอง (ACF) และสหสัมพันธ์ในตัวเองบางส่วน (PACF) ซึ่งเป็นวิธีที่นิยมในปัจจุบันแต่การพยากรณ์ในวิธีการดังกล่าวยังพบความคลาดเคลื่อนยังมีค่าสูง ดังนั้นผู้วิจัยได้พยากรณ์ด้วยวิธีแบ็กกิ้ง วิธีเรนดอมฟอเรสต์ เนื่องจากเป็นการวิเคราะห์พยากรณ์ที่เป็นที่นิยมของการทำเหมืองข้อมูล (Data mining) ซึ่งเป็นวิธีเป็นการค้นหาหรือสกัดความรู้จากฐานข้อมูลขนาดใหญ่ เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัย หลักการทางคณิตศาสตร์ สถิติ การรู้จำ การเรียนรู้ ทางเครื่องจักร เพื่อนำความรู้ ที่ได้้นั้นมาใช้ในการ แก้ปัญหา วางแผน หรือดำเนินกลยุทธ์ขององค์กรให้ ประสบความสำเร็จสูงสุด และสามารถดึงข้อมูลมาใช้ โดยการคัดเลือกข้อมูลออกมาใช้งานในส่วนที่ต้องการ

วิธีการทดลอง

1. วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins method)

การพยากรณ์ด้วยวิธีของบ็อกซ์-เจนกินส์เป็นการพยากรณ์เชิงปริมาณวิธีหนึ่งที่มีแนวคิดว่าพฤติกรรมในอดีตของสิ่งที่ต้องการพยากรณ์นั้นเพียงพอที่จะพยากรณ์พฤติกรรมในอนาคตของตัวเองได้ โดยในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีของบ็อกซ์-เจนกินส์นี้จะแตกต่างจากการพยากรณ์โดยวิธีอื่นซึ่งผู้ที่สร้างตัวแบบพยากรณ์นั้นต้องกำหนดรูปแบบของความสัมพันธ์ก่อนที่จะทำการวิเคราะห์ โดยเฉพาะเมื่ออนุกรมเวลาไม่มีแนวโน้มวัฏจักรหรือฤดูกาลที่ชัดเจน ทำให้ยากในการกำหนดรูปแบบหรือการวิเคราะห์การถดถอยที่เหมาะสมได้ ซึ่งจะต้องทำการกำหนดรูปแบบของความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามก่อน แต่วิธีพยากรณ์ของบ็อกซ์-เจนกินส์สามารถแก้ปัญหาดังกล่าวได้ เพราะวิธีพยากรณ์ของบ็อกซ์-เจนกินส์นั้นไม่มีการกำหนดรูปแบบที่ตายตัวขึ้นก่อนทำการวิเคราะห์ โดยในระหว่างการวิเคราะห์รูปแบบจะถูกกำหนดขึ้นมาเองซึ่งสามารถทำตามขั้นตอนของบ็อกซ์-เจนกินส์ได้ดังนี้

1. กำหนดค่าของฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF)

2. กำหนดรูปแบบการพยากรณ์ $SARIMA(p,d,q)(P,D,Q)_s$ ที่เหมาะสมโดยพิจารณาจากค่า ACF และ PACF แสดงดังสมการที่ (1) (Jenkins & Reinsel, 1994)

$$f_p(B)F_p(B^s)(1-B)^d(1-B^s)^D Y_t = d + q_q(B)Q_Q(B^s)e_t \quad (1)$$

โดยที่ Y_t แทนอนุกรมเวลา ณ เวลา t
 t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n
 n แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1
 s แทนจำนวนคาบของฤดูกาล
 d และ D แทนลำดับที่ของการหาผลต่าง และผลต่างฤดูกาล ตามลำดับ
 B แทนตัวดำเนินการถอยหลัง (Backward Operator) โดยที่ $B^s Y_t = Y_{t-s}$
 e_t แทนอนุกรมเวลาของความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน
 ด้วยค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ทุกช่วงเวลา

$d = mf_p(B)F_p(B^s)$ แทนค่าคงตัว โดยที่ m แทนค่าเฉลี่ยของอนุกรมเวลาที่คงที่

$f_p(B) = 1 - f_1 B - f_2 B^2 - \dots - f_p B^p$ แทนตัวดำเนินการสหสัมพันธ์ในตัวเองแบบไม่มีฤดูกาลอันดับที่ p

(Non-Seasonal Autoregressive Operator of Order p : $AR(p)$)

$F_p(B^s) = 1 - F_1 B^s - F_2 B^{2s} - \dots - F_p B^{ps}$ แทนตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาลอันดับที่ P

(Seasonal Autoregressive Operator of Order P : $SAR(P)$)

$q_q(B) = 1 - q_1 B - q_2 B^2 - \dots - q_p B^p$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับที่ q

(Non-Seasonal Moving Average Operator of Order q : $MA(q)$)

$Q_Q(B^s) = 1 - Q_1 B^s - Q_2 B^{2s} - \dots - Q_Q B^{Qs}$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับที่ Q

(Seasonal Moving Average Operator of Order Q : $SMA(Q)$)

2. วิธีแบ็กกิ้ง (Bagging Method)

วิธีแบ็กกิ้งเป็นเทคนิคย่อยหนึ่งในเทคนิคการรวมกลุ่มตัวจำแนกประเภท (Ensemble Classify) โดยมีวิธีการสุ่มข้อมูลย่อยจากชุดข้อมูลฝึกสอนแบบสุ่มซ้ำได้ (Boostraps Replacement) จำนวนหลายชุดย่อย นำมาสร้างตัวจำแนก ที่แตกต่างกันแล้วจึงนำผลการทำนายของแต่ละตัวจำแนกย่อยมาพิจารณาร่วมกัน แล้วสร้างตัวแบบแต่ละตัวที่ใช้ชุดข้อมูลเรียนรู้ที่สุ่มมาจากชุดข้อมูลเรียนรู้ที่กำหนดให้ในแบบคืน (Sampling with Replacement) กล่าวคือเมื่อสุ่มข้อมูลขึ้นมาได้หนึ่งตัวจะคืนข้อมูลนั้นกลับเข้าไปยังชุดข้อมูลเดิม ทำให้ข้อมูลนั้นมีโอกาสถูกสุ่มซ้ำอีกในอนาคต

ข้อมูลแต่ละตัวในชุดข้อมูลเรียนรู้ จะมีโอกาสถูกสุ่มเพื่อใช้ในการสร้างตัวแบบเท่ากับ $1 - \frac{1}{n}$ โดยที่ n เป็นจำนวน ข้อมูลในชุดข้อมูลเรียนรู้ที่กำหนดให้ ถ้าทำการสุ่มข้อมูลด้วยวิธีนี้ n ครั้ง เพื่อสร้างชุดข้อมูล สำหรับสร้างตัวแบบ 1 ตัว จะได้ชุดข้อมูลที่เป็นเซตย่อย (subset) ของชุดข้อมูลเรียนรู้ที่กำหนดให้ ทั้งนี้เนื่องจากข้อมูลเดียวกันอาจถูกสุ่มซ้ำ ทำให้ชุดข้อมูลจากการสุ่มมีจำนวนข้อมูลไม่ถึง n ตัว โดยเฉลี่ยแล้วจะมีขนาดประมาณ 63.2% ของ n (Gorad et al., 2017)

3. วิธีเรณดอมฟอเรสต์ (Random Forest Method)

เทคนิคที่ทำการสุ่มเลือกคุณสมบัติ ออกมาจากชุดข้อมูลหลาย ๆ ชุด จากนั้น นำเอาชุดของคุณสมบัติเหล่านั้นมาสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจหลาย ๆ ต้น ลักษณะของต้นไม้ที่อยู่ภายในป่าของเทคนิควิธีเรณดอมฟอเรสต์ จะถูกควบคุมด้วย 3 ปัจจัยคือ

1. ต้นไม้แต่ละต้นจะถูกสอน (Train) โดยการใช้เซตย่อยจากข้อมูลตัวอย่าง
2. เมื่อต้นไม้โตขึ้น จะสามารถค้นหาโนด (Node) แต่ละ โนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้โดยใช้การสุ่มเลือกคุณสมบัติจาก N คุณสมบัติ
3. ต้นไม้แต่ละต้นจะไม่มีการตัดออก แต่จะปล่อยให้ต้นไม้โตขึ้นไปเรื่อย ๆ จนได้ผลลัพธ์ที่ดีที่สุดหลังจากการสร้างป่า แล้วทำการให้คะแนนโหนด โดยต้นไม้ภายในป่า หากต้นไม้ต้นใดได้คะแนน โหนดสูงสุดก็จะนำเอาต้นไม้ต้นนั้นออกมาสร้างเป็น โมเดล (Sujatha & Devi, 2013)

4. การวัดความคลาดเคลื่อนของการพยากรณ์ (Forecasting Error)

การวัดความคลาดเคลื่อนของการพยากรณ์จะอาศัยหลักการ การเปรียบเทียบระหว่างค่า พยากรณ์ที่คำนวณได้กับข้อมูลจริงในช่วงเวลา t หากค่าพยากรณ์มีค่าคลาดเคลื่อนมากอาจหมายถึงวิธีการที่ใช้พยากรณ์นั้นไม่เหมาะสม หรืออาจจำเป็นต้องเปลี่ยนค่าพารามิเตอร์บางค่าให้เหมาะสม สำหรับงานวิจัยนี้จะใช้วิธีวัดความคลาดเคลื่อนของการพยากรณ์ 2 วิธีคือ

- 1) เกณฑ์รากที่สองของความคลาดเคลื่อนเฉลี่ย (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (\text{Chai \& Draxler, 2014})$$

- 2) เปอร์เซนต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

$$MAPE = \frac{\sum_{i=1}^n |(e_i / Y_i)|}{n} \cdot 100 \quad (\text{Isserman, 1977})$$

โดยที่ Y_i แทนค่าของข้อมูลจริงที่เวลา i

F_i แทนค่าพยากรณ์ที่เวลา i

e_i แทนค่าความคลาดเคลื่อนที่เวลา i

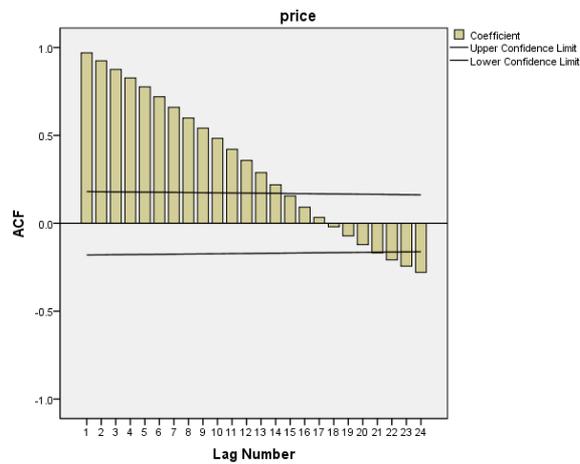
$$e_i = Y_i - F_i$$

ผลและวิจารณ์ผลการทดลอง

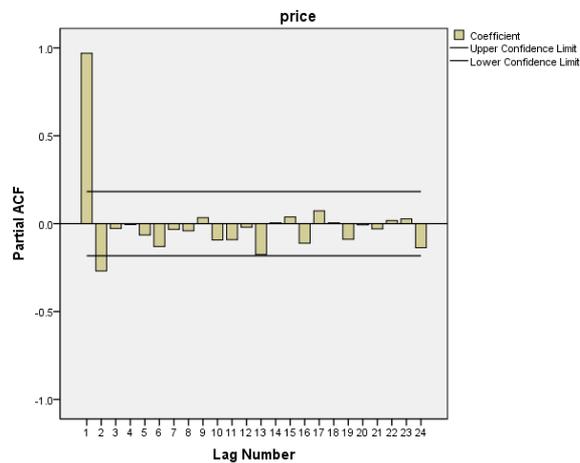
1. วิธีบอกซ์เจนนินส์ (Box-Jenkins method)

นำข้อมูลราคาข้าวเปลือกหอมมะลิรายเดือน ไปวิเคราะห์โดยใช้โปรแกรม Statistics Package for Social Sciences: SPSS ตามขั้นตอนต่อไปนี้

ตรวจสอบข้อมูลจากกราฟของฟังก์ชันสหสัมพันธ์ในตัวเอง ACF และการฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน PACF ของราคาข้าวเปลือกหอมมะลิดังรูปที่ 1 และรูปที่ 2

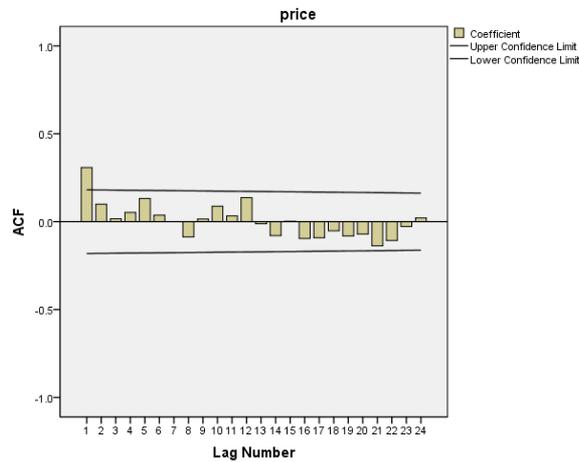


รูปที่ 1 กราฟแสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF)

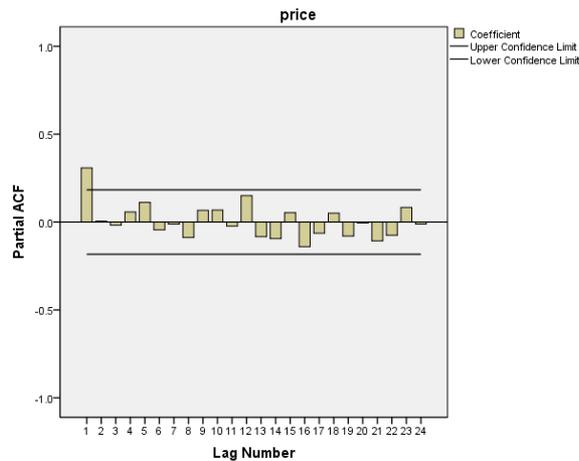


รูปที่ 2 กราฟแสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF)

จากรูปที่ 1 พบว่าการฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ลดลงเข้าสู่ศูนย์อย่างช้า ๆ แสดงว่าอนุกรมเวลาไม่นิ่ง จึงทำการแปลงอนุกรมเวลาโดยการหาผลต่างอันดับที่ 1 ($d = 1$) ของราคาข้าวเปลือกหอมมะลิทั้ง 120 เดือน ได้กราฟ ACF และ PACF ดังรูปที่ 3 และรูปที่ 4



รูปที่ 3 กราฟแสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของผลต่างอันดับที่ 1 ($d = 1$)



รูปที่ 4 กราฟแสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของผลต่างอันดับที่ 1 ($d = 1$)

จากรูปที่ 3 พบว่า ACF มีค่าแตกต่างไปจากศูนย์อย่างมีนัยสำคัญ และลดลงเข้าสู่ศูนย์อย่างรวดเร็วแสดงว่าอนุกรมเวลานิ่ง จึงกำหนดตัวแบบพยากรณ์ที่เป็นไปได้เริ่มต้นคือตัวแบบ SARIMA(1,1,0)(1,0,0)₁₂ แบบมีพจน์ค่าคงตัว พร้อมกับประมาณค่าพารามิเตอร์ โดยตัวแบบพยากรณ์มีพารามิเตอร์ทุกตัวมีนัยสำคัญทางสถิติที่ระดับ 0.05 มีค่า

BIC ต่ำที่สุด (BIC = 0.144) และมีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญที่ระดับ 0.05 (Ljung-Box Q ณ lag 18 = 12.391 p-value = 0.873) เมื่อตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ที่ระดับนัยสำคัญ 0.05 พบว่าความคลาดเคลื่อนมีการแจกแจงปรกติ มีการเคลื่อนไหวเป็นอิสระกัน และจากรูปที่ 5 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนของความคลาดเคลื่อนอยู่ในขอบเขตความเชื่อมั่นร้อยละ 95 ดังนั้นตัวแบบดังกล่าวจึงมีความเหมาะสม จากสมการที่ (1) สามารถเขียนเป็นตัวแบบได้ดังนี้

$$(1 - f_1 B)(1 - F_1 B^{12})(1 - B)^1 Y_t = d + e_t$$

แทนค่าประมาณพารามิเตอร์ $d = 43.289$, $f_1 = 0.344$ และ $F_1 = 0.261$ จะได้

$$(-0.089784B^{14} + 0.350784B^{13} - 0.261B^{12} + 0.344B^2 - 1.344B + 1)Y_t = 43.289 + e_t$$

$$Y_t = 43.289 + 0.089784Y_{t-14} - 0.350784Y_{t-13} + 0.261Y_{t-12} - 0.344Y_{t-2} + 1.344Y_{t-1} - 1 + e_t$$

ดังนั้นตัวแบบพยากรณ์คือ

$$\hat{Y}_t = 43.289 + 0.089784Y_{t-14} - 0.350784Y_{t-13} + 0.261Y_{t-12} - 0.344Y_{t-2} + 1.344Y_{t-1} - 1 + e_t$$

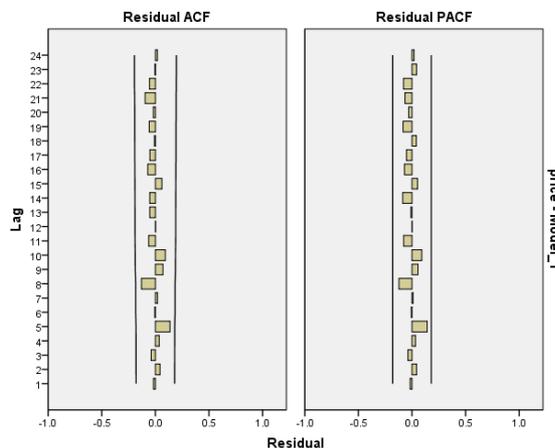
โดยที่ \hat{Y}_t แทนค่าพยากรณ์ ณ เวลา t

Y_{t-j} แทนอนุกรมเวลา ณ เวลา $t - j$

e_t ค่าความคลาดเคลื่อนจากการพยากรณ์ ณ เวลา t

นำสมการดังกล่าวไปพยากรณ์ทำนาราคาข้าวเปลือกหอมมะลิล่วงหน้า 12 เดือน คือ เดือนมกราคม พ.ศ.

2563 - ธันวาคม พ.ศ. 2563 แสดงได้ดังตารางที่ 1



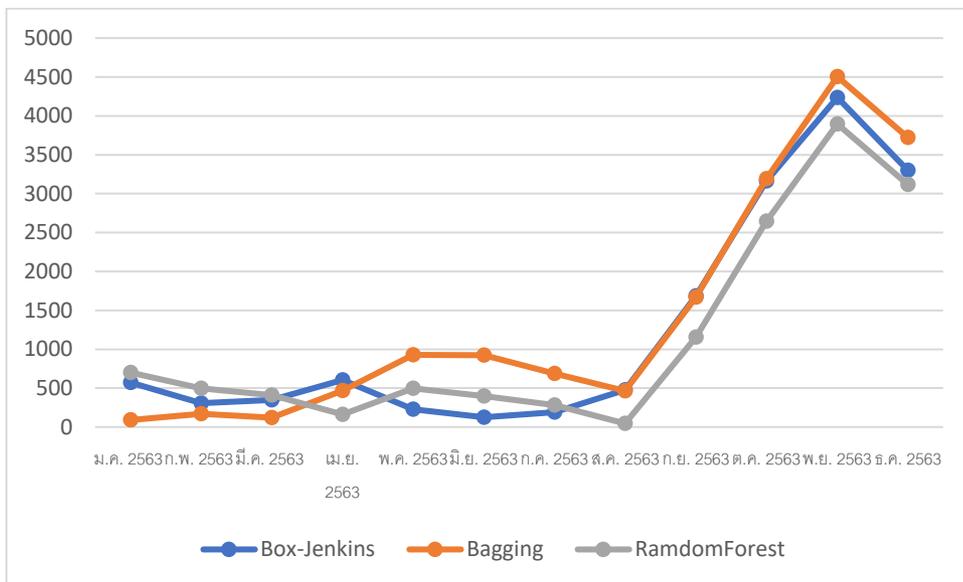
รูปที่ 5 กราฟ ACF และ PACFของความคลาดเคลื่อนจากการพยากรณ์โดยวิธีบอซกซ์เงินกินส์

2. วิธีแบ็กกิ้ง (Bagging method) และ วิธีเรนดอมฟอเรสต์ (Random Forest method)

สำหรับการสร้างแบบจำลองโดยวิธีแบ็กกิ้งและวิธีเรนดอมฟอเรสต์นั้น ได้ใช้โปรแกรม Weka ในการคำนวณ โดยวิธีแบ็กกิ้ง ผู้วิจัยได้กำหนดค่าพารามิเตอร์ ดังนี้ Sample ratio มีค่าเท่ากับ 0.9, Iteration มีค่าเท่ากับ 7 และ Average Confidences มีค่าเป็น True และวิธีเรนดอมฟอเรสต์ ผู้วิจัยได้กำหนดค่าพารามิเตอร์ ดังนี้ Number of tree มีค่าเท่ากับ 50, Criterion มีค่าเป็น Information_gain, Maximal depth มีค่าเท่ากับ 9, Apply pruning มีค่าเป็น True และค่า Confidence มีค่าเท่ากับ 0.25 ซึ่งได้ผลค่าพยากรณ์ตามตารางที่ 1

ตารางที่ 1 ค่าจริงและค่าพยากรณ์ราคาข้าวเปลือกหอมมะลิ(บาท/เมตริกตัน) ตั้งแต่เดือนมกราคม พ.ศ. 2563 – เดือนธันวาคม พ.ศ. 2563

เดือน	ค่าจริง	ค่าพยากรณ์ล่วงหน้า		
		Box-Jenkins	Bagging	Ramdom Forest
มกราคม พ.ศ. 2563	13,756	13183	13846.38	14455.71
กุมภาพันธ์ พ.ศ. 2563	13,913	13607	14084.90	14408.86
มีนาคม พ.ศ. 2563	13,949	14298	14068.16	14357.77
เมษายน พ.ศ. 2563	14,471	13864	14003.06	14306.95
พฤษภาคม พ.ศ. 2563	14,871	14643	13942.96	14371.83
มิถุนายน พ.ศ. 2563	14,868	14994	13942.96	14471.17
กรกฎาคม พ.ศ. 2563	14,630	14820	13942.96	14349.66
สิงหาคม พ.ศ. 2563	14,409	14885	13942.96	14363.21
กันยายน พ.ศ. 2563	13,237	14921	14907.42	14393.36
ตุลาคม พ.ศ. 2563	11,715	14876	14907.42	14362.61
พฤศจิกายน พ.ศ. 2563	10,403	14640	14907.42	14298.41
ธันวาคม พ.ศ. 2563	11,185	14485	14907.42	14303.89
ความคลาดเคลื่อน	RMSE	1,891.05	2,037.26	1,699.64
ของการพยากรณ์	MAPE	10.83	11.90	9.80



รูปที่ 6 กราฟเปรียบเทียบค่าความคลาดเคลื่อน(บาท/เมตรกรัดัน) ของค่าพยากรณ์

จากผลการทำนายราคาข้าวเปลือกหอมมะลิ ผลการทดลองพบว่า วิธีบ็อกซ์เจนกินส์, วิธีแบ็กกิ้ง และ วิธี แรนดอมฟอเรสต์ ได้ค่า RMSE เป็น 1,891.05, 2,037.26 และ 1,699.64 ตามลำดับ และได้ค่า MAPE เป็น 10.83, 11.90 และ 9.80 ตามลำดับ ซึ่งค่าที่ได้มีความใกล้เคียงกัน แต่วิธีแรนดอมฟอเรสต์ เป็นวิธีที่ได้ค่าทั้งสองต่ำที่สุด เมื่อเทียบกับ วิธีบ็อกซ์เจนกินส์และวิธีแบ็กกิ้ง ดังรูปที่ 6

ผลการวิจัยนี้สอดคล้องกับงานวิจัยของ Poonsawad & Sanrach (2019) ซึ่งศึกษาเทคนิคพยากรณ์การได้รับ บัณฑิตพื้นฐานนักเรียนยากจนของนักเรียนโรงเรียนวัดพระขาว (ประจวบคีรีขันธ์) ด้วยเทคนิคเหมืองข้อมูล โดยใช้ เทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ ต้นไม้ตัดสินใจ แรนดอมฟอเรสต์ และเทคนิคแบ็กกิ้ง ผลการวิจัย พบว่า เทคนิคการจำแนกข้อมูลด้วยเทคนิคแรนดอมฟอเรสต์ให้ค่าประสิทธิภาพมากที่สุด และสอดคล้องกับงานวิจัยของ Intarat & Sillaparat (2019) ได้ศึกษาการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนด้วยวิธีแรนดอมฟอเรสต์ และข้อมูลภาพถ่ายจากดาวเทียมรายละเอียดสูง ผลการวิจัยพบว่า การใช้วิธีแรนดอมฟอเรสต์ ให้ประสิทธิภาพในการจำแนกที่สูง กว่าวิธีการจำแนกเชิงคุณภาพแบบความน่าจะเป็นสูงสุด โดยมีค่าความถูกต้องโดยรวมที่ร้อยละ 78.00 และค่าสถิติ แคลปปาที่ 0.72 ในขณะที่ผลการจำแนกด้วยความน่าจะเป็นสูงสุดให้ค่าความถูกต้องโดยรวมที่ร้อยละ 56.00 และ ค่าสถิติแคลปปาที่ 0.44 ผลจากการทดสอบค่าสถิติ Z ($Z = 3.68$) ช่วยยืนยันถึงความแตกต่างระหว่างทั้งสองวิธีการ จำแนก อย่างมีนัยสำคัญที่ค่าความเชื่อมั่นร้อยละ 95.00

สรุปผลการทดลอง

แบบจำลองทางคณิตศาสตร์ทำนายราคาข้าวเปลือกหอมมะลิ โดยใช้ตัวแบบพยากรณ์ วิธีบอกซ์-เจนกินส์ วิธีแบ็กกิ้ง และ วิธีเรนคอมโพเรสต์พบว่า วิธีเรนคอมโพเรสต์ได้ทำนายราคาข้าวเปลือกหอมมะลิได้ใกล้เคียงกับค่าจริงที่สุด และเมื่อคำนวณค่า RMSE ได้ 1,699.64 และได้ค่า MAPE เป็น 9.80

เอกสารอ้างอิง

- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3), 1247-1250.
- Gorad, N., Zalte, I., Nandi, A., & Nayak, D. (2017). Career counselling using data mining. *International Journal of Engineering Science and Computing*, 7(4), 10271-10274.
- Intarat, K., & Sillaparat, S. (2019). Tropical mangrove species classification using random forest algorithm and very high-resolution satellite imagery. *Burapha Science Journal*, 24(2), 742-753. (in Thai)
- Isserman, A. M. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute for Planners*, 43(3), 247-259.
- Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis; forecasting and control* (3rd ed.). Prentice Hall, Englewood Cliff, New Jersey.
- Kodosueb, S., & Boonlha, K. (2016). Construction of model for the price of Thai jasmine rice 105. *Science and technology Nakhon Sawan Rajabhat University Journal*, 8(8), 49-60. (in Thai)
- Poonsawad, A., & Sanrach, C. (2019). A study of techniques in predicting of a fund receiving for poor students of Phrakhao school students by using data mining technique. *The Sci of Phetchaburi Rajabhat University*, 16(2), 1-10. (in Thai)
- Riansu, W. (2016). Forecasting the prices of paddy rice at 15% moisture content. *The Sci J of Phetchaburi Rajabhat University*, 13(1), 14-25. (in Thai)
- Sujatha, M., & Devi, L. (2013). Feature selection techniques using for high dimensional data in machine learning. *International Journal of Engineering Research & Technology*, 2(9), 97-102.
- Thai Rice Exporters Association. (2020). the 53rd Annual General Assembly Meeting for the Year 2020. (in Thai)
- Wanon, S., Arreerard, T., & Sanrach, C. (2018). A study of techniques in predicting career counseling for undergraduate students of the computer program by using data mining technique. *Journal of Innovation Technology Management*, 5(1), 164-171. (in Thai)

Yoyram, J., Suntaralak, S., & Somchai, K. (2020). Analysis of consumer factors towards the purchase of silk pupa in Ban Hua Saphan, Ban Yang, Phutthaisong District, Buriram Province with data mining techniques. *The 6th National Conference on Technology and Innovation Management NCTIM 2020* (pp. 1-7). Rajabhat Maha Sarakham University, Maha Sarakham. (in Thai)