

เปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องสำหรับทำนายราคาพริกชี้หนู ในจังหวัดนครศรีธรรมราช

Comparison of the Efficiency of Machine Learning Model for Predicting Bird's Eye Chili Prices in Nakhon Si Thammarat Province

โสภี แก้วชะภา¹ สมพร เรืองอ่อน^{*1} และ อุทัย คูหาพงศ์²

Sopee Kaewchada¹, Somporn Ruang-on^{*1}, & Uthai Kuhapong²

¹สาขาวิชาวิศวกรรมวิทยาศาสตร์และเทคโนโลยีสร้างสรรค์ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยราชภัฏนครศรีธรรมราช

²สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครศรีธรรมราช

¹Creative Innovation in Science and Technology Program, Faculty of Science and Technology,
Nakhon Si Thammarat Rajabhat University

²Information Technology Program, Faculty of Science and Technology,
Nakhon Si Thammarat Rajabhat University

Submitted 18/2/2022 ; Revised 18/4/2022 ; Accepted 3/5/2022

บทคัดย่อ

จังหวัดนครศรีธรรมราชเป็นแหล่งผลิตพริกชี้หนูส่งออกแต่ประสบปัญหาภัยแล้งและราคาผลผลิตตกต่ำ ประกอบกับเกษตรกรไม่สามารถกำหนดราคาได้เอง งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาตัวแบบการเรียนรู้ของเครื่องที่เหมาะสมสำหรับการทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช เพื่อใช้ประกอบการพัฒนาตัวแบบสำหรับทำนายราคาพริกชี้หนู และวางแผนการปลูกผักในพื้นที่ลุ่มน้ำปากพนัง โดยใช้ข้อมูลย้อนหลังของสภาพภูมิอากาศจากสถานีอุตุนิยมวิทยานครศรีธรรมราช และราคาพริกชี้หนูจากสำนักงานพาณิชย์จังหวัดนครศรีธรรมราช พ.ศ. 2553 – 2563 ตัวแบบการเรียนรู้ของเครื่องที่นำมาพิจารณา ได้แก่ การถดถอยเชิงเส้นเพื่อบ้านใกล้เคียง และต้นไม้ป่าสุ่ม ผลการวิจัยพบว่า จากการประเมินประสิทธิภาพของตัวแบบโดยหาค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 11.71, 11.43 และ 9.68 ตามลำดับ สรุปได้ว่า ต้นไม้ป่าสุ่มเป็นตัวแบบที่เหมาะสมสำหรับนำไปประกอบการพัฒนาตัวแบบสำหรับทำนายราคาพริกชี้หนู ในพื้นที่จังหวัดนครศรีธรรมราช และผลการทำนายราคาพริกชี้หนูโดยใช้ต้นไม้ป่าสุ่มเปรียบเทียบกับราคาเฉลี่ยรายเดือนจริงของพริกชี้หนู พบว่าราคามีความใกล้เคียงกับราคาจริง

คำสำคัญ: ตัวแบบ การเรียนรู้ของเครื่อง การทำนาย ราคาพริกชี้หนู

***ผู้ประสานงานหลัก (Corresponding Author)**

E-mail: somporn_rua@nstru.ac.th

เปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องสำหรับทำนายราคาพริกชี้หนู
ในจังหวัดนครศรีธรรมราช

Abstract

Nakhon Si Thammarat Province is a source of bird's eye chili for export but faces drought and low prices. In addition, farmers are unable to set their prices. Therefore, the purpose of this research was to examine appropriate machine learning models for predicting the price of bird's eye chili in Nakhon Si Thammarat province. The appropriate model can be used in the development of a model for predicting the price of bird's eye chili and plant planning in the Pak Phanang River Basin. The data used in this research was historical data on the climate and the price of bird's eye chili between 2010 and 2020, provided by the Meteorological Station and the Provincial Commercial Office, Nakhon Si Thammarat province, respectively. The machine learning models used in this research were linear regression, K-nearest neighbor, and random forest. The results of the model's efficiency evaluation by estimating the mean absolute error (MAE) were 11.71, 11.43, and 9.68, respectively. As a result, the random forest is a suitable model for predicting bird's eye chili prices in the Nakhon Si Thammarat region. Moreover, the prediction price from the random forest is found to be closed to the actual price of the bird's eye chili monthly average price.

Keywords: model, machine learning, prediction, price

บทนำ

พริกชี้หนูจัดเป็นไม้ล้มลุกหรือไม้พุ่มขนาดเล็ก ที่มีความสูงของต้นประมาณ 30-90 เซนติเมตร มีอายุประมาณ 1-3 ปี พื้นที่ลุ่มน้ำปากพนังครอบคลุม 5 อำเภอของจังหวัดนครศรีธรรมราช ได้แก่ หัวไทร ปากพนัง เขียวใหญ่ เฉลิมพระเกียรติ และชะอวด เป็นแหล่งผลิตพริกชี้หนูเพื่อส่งออกที่สำคัญของภาคใต้ โดยเฉพาะการส่งออกไปยังตลาดหลักที่สำคัญในประเทศมาเลเซียและสิงคโปร์ [1] ปัญหาที่พบในการปลูกพริกชี้หนูในจังหวัดนครศรีธรรมราช คือ การเกิดภัยแล้ง จากสถิติปี พ.ศ. 2556 - 2562 จังหวัดนครศรีธรรมราชประสบปัญหาภัยแล้งทั้งหมด 5 ครั้ง โดยในปี พ.ศ. 2559 มีพื้นที่ประสบภัยแล้งสูงสุด 12 อำเภอ และมีพื้นที่เกษตรกรรมได้รับความเสียหาย 552,212 ไร่ [2] นอกจากปัญหาสภาพภูมิอากาศที่ไม่เอื้ออำนวยแล้ว เกษตรกรในพื้นที่จังหวัดนครศรีธรรมราชยังประสบปัญหาโรคพืช แมลงศัตรูพืชระบาด ราคาผลผลิตตกต่ำ และเกษตรกรไม่สามารถกำหนดราคาผลผลิตทางการเกษตรได้เอง [1]

ด้วยวิทยาการทางด้านเทคโนโลยีดิจิทัลมีความก้าวหน้าอย่างรวดเร็วและมีการพัฒนาแบบก้าวกระโดด ต้องค์กรหรืออาชีพต่าง ๆ ขณะที่ประเทศไทยมีเป้าหมายพัฒนาสู่การเป็นประเทศไทย 4.0 (Thailand 4.0) โดยเน้นการขับเคลื่อนเศรษฐกิจด้วยนวัตกรรม การนำเทคโนโลยีดิจิทัลมาผนวกกับการเกษตรเป็นการเกษตรยุคใหม่ที่มุ่งเพิ่มความสามารถในการผลิต และนำไปช่วยในการบริหารจัดการ การวางแผนการเพาะปลูก การตลาด โลจิสติกส์ รวมถึงการคาดการณ์แนวโน้มในอนาคตให้เหมาะสม เป็นต้น [3] ประกอบกับในปัจจุบันเทคนิค “การเรียนรู้ของเครื่อง (machine learning)” นิยมนำมาประยุกต์ใช้อย่างแพร่หลาย เพื่อรวบรวมข้อมูลและสารสนเทศที่สำคัญไปวิเคราะห์และสังเคราะห์สำหรับอธิบายและจำแนกรูปแบบ รวมถึงการนำไปใช้ในการทำนายหรือพยากรณ์ โดยการทดสอบเปรียบเทียบความสามารถของอัลกอริทึมแต่ละประเภท เพื่อค้นหาอัลกอริทึมที่เหมาะสมที่สุด ผลจากการทดสอบส่วนใหญ่ปรากฏว่าไม่มีอัลกอริทึมใดที่ทำงานได้ดีที่สุดในข้อมูลทุกประเภท เนื่องจากข้อมูลแต่ละประเภทมีลักษณะเฉพาะที่แตกต่างกัน [4]

จากการศึกษางานวิจัยอื่นผ่านมาพบว่า เทคนิคการเรียนรู้ของเครื่องถูกนำมาใช้ในพัฒนาตัวแบบสำหรับทำการวิเคราะห์ข้อมูลและทำนายผลด้วยการเปรียบเทียบตัวแบบต่างอัลกอริทึมเพื่อหาแบบการอธิบายข้อมูล เช่น การเปรียบเทียบเทคนิคการจำแนกประเภทข้อมูลสำหรับทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียน [4] การประเมินประสิทธิภาพของต้นไม้ป่าสุ่ม (random forest) กับการถดถอยพหุคูณ (multiple regression) สำหรับการคาดการณ์ราคาของอพาร์ทเมนต์ [5] แบบจำลองการพยากรณ์ราคามันสำปะหลังโดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น [6] เป็นต้น

จากเหตุผลดังกล่าวผู้วิจัยจึงเปรียบเทียบตัวแบบการเรียนรู้ของเครื่องเพื่อหาตัวแบบที่เหมาะสมในการทำนายราคาพริกชี้หนูสำหรับวางแผนการปลูกพริกชี้หนูในพื้นที่ลุ่มน้ำปากพนัง จังหวัดนครศรีธรรมราช โดยใช้ข้อมูลสภาพภูมิอากาศจากสถานีอุตุนิยมวิทยานครศรีธรรมราช (บางจาก) และราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช จากสำนักงานพาณิชย์จังหวัดนครศรีธรรมราช ระหว่างปี พ.ศ. 2553-2563

วัตถุประสงค์

1. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องที่เหมาะสมสำหรับทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช
2. เพื่อทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช สำหรับใช้วางแผนการเพาะปลูกพริกชี้หนูในพื้นที่ลุ่มน้ำปากพนัง จังหวัดนครศรีธรรมราช

วิธีดำเนินการวิจัย

การดำเนินการวิจัยประกอบด้วย 2 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 เปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องที่เหมาะสมสำหรับทำนายราคาพริกขี้หนูในจังหวัดนครศรีธรรมราช มีรายละเอียดดังนี้

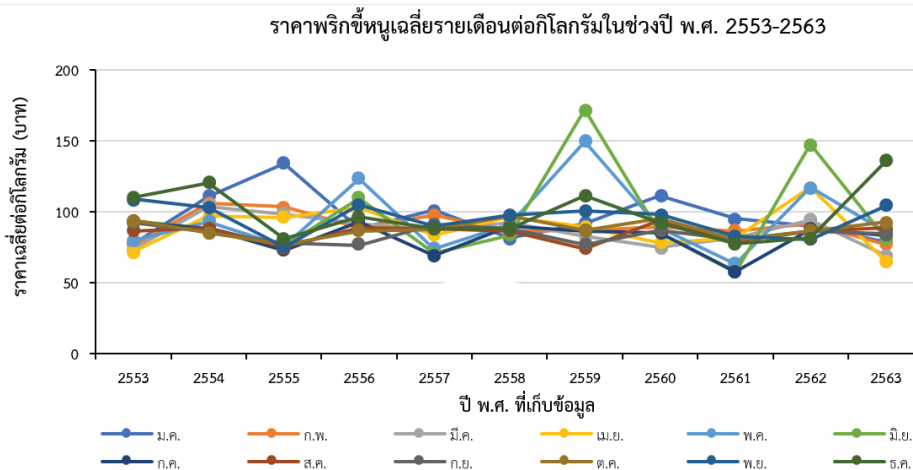
1. พื้นที่การทำวิจัย

งานวิจัยครั้งนี้ใช้พื้นที่ลุ่มน้ำปากพนัง จังหวัดนครศรีธรรมราชเป็นพื้นที่ศึกษาวิจัย ข้อมูลสภาพภูมิอากาศจากสถานีอุตุนิยมวิทยานครศรีธรรมราช และราคาพริกขี้หนูในจังหวัดนครศรีธรรมราชจากสำนักงานพาณิชย์ จังหวัดนครศรีธรรมราช

2. การรวบรวมและเตรียมข้อมูล

กระบวนการรวบรวมและเตรียมข้อมูลประกอบด้วย 3 กระบวนการย่อยดังนี้

2.1 การคัดเลือกข้อมูล โดยข้อมูลที่ใช้ในการวิจัย ได้แก่ (1) เดือนที่เก็บข้อมูลราคา (2) ปริมาณน้ำฝนเฉลี่ยรายเดือน (3) ความชื้นเฉลี่ยรายเดือน (4) อุณหภูมิเฉลี่ยรายเดือน (5) ฤดูกาลที่เก็บข้อมูลราคา (6) ราคาพริกขี้หนูเฉลี่ยรายเดือนต่อกิโลกรัม และ (7) ราคาพริกขี้หนูเฉลี่ยรายปีต่อกิโลกรัม โดยราคาพริกขี้หนูเฉลี่ยรายเดือนจากสำนักงานพาณิชย์จังหวัดนครศรีธรรมราช ระหว่างปี พ.ศ. 2553-2563 แสดงดังภาพที่ 1



ภาพที่ 1 ราคาพริกขี้หนูเฉลี่ยรายเดือนในช่วงปี พ.ศ. 2553-2563

ที่มา: สำนักงานพาณิชย์จังหวัดนครศรีธรรมราช

2.2 การทำความสะอาดข้อมูล

2.2.1 การแทนค่าข้อมูลสูญหาย แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยอนุกรม (series mean) ดังสมการที่ 1 ซึ่งเป็นการแทนที่ค่าสูญหายด้วยค่าเฉลี่ยทั้งชุดข้อมูล [7] โดยใช้โปรแกรมสำเร็จรูปทางสถิติ

$$\hat{X} = \frac{\sum_{i=1}^k X_i}{k} \quad (1)$$

โดยที่ X_i คือข้อมูลที่ไม่สูญหาย; k คือ จำนวนข้อมูลที่ไม่สูญหาย และ $\hat{X}$ คือ ค่าประมาณข้อมูลที่สูญหายที่ถูกประมาณด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย

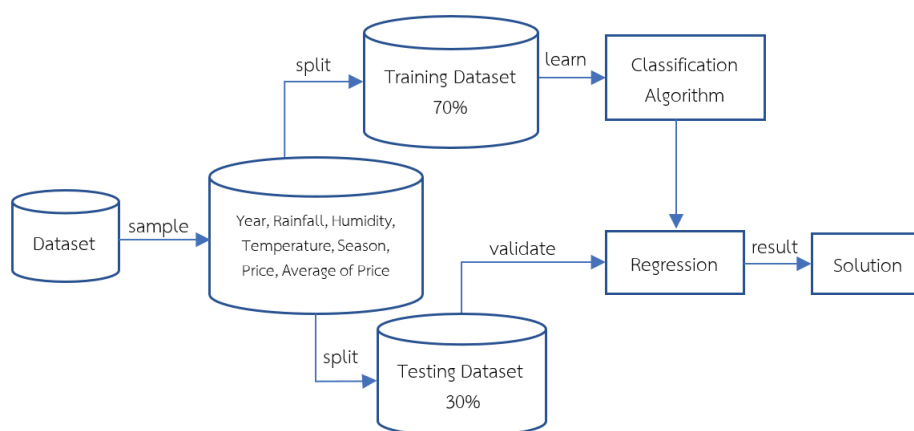
เปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องสำหรับทำนายราคาพริกขี้หนูในจังหวัดนครศรีธรรมราช

2.2.2 การจัดการข้อมูลผิดปกติ สำหรับสำรวจข้อมูลที่แสดงค่าผิดปกติ โดยใช้ boxplot ซึ่งเป็นวิธีมาตรฐานสำหรับแสดงข้อมูล (data visualization) ที่แสดงลักษณะการกระจายตัวและแสดงค่าความผิดปกติของข้อมูลได้ [8]

2.3 การแปลงรูปแบบข้อมูล แปลงแฟ้มข้อมูลจากรูปแบบตารางทำการ (spreadsheet file) ให้อยู่ในรูปแบบไฟล์ซีเอสวี (comma separated values: CSV) โดยใช้โปรแกรมไมโครซอฟต์เอกเซล (Microsoft Excel) เพื่อนำไปใช้พัฒนาแบบจำลองการทำนาย

3. การพัฒนาตัวแบบการทำนาย

3.1 การออกแบบตัวแบบการทำนาย งานวิจัยนี้ออกแบบตัวแบบโดยใช้หลักการทดสอบประสิทธิภาพของตัวแบบด้วยวิธีการแยกทดสอบ (split test) [9] โดยแบ่งข้อมูลออกเป็น 2 ชุด ข้อมูลชุดแรกใช้สำหรับสร้างตัวแบบการทำนาย (training dataset) จำนวน 92 รายการ คิดเป็นร้อยละ 70 ของข้อมูลทั้งหมด และข้อมูลชุดที่ 2 ใช้สำหรับทดสอบผลการทำนายของตัวแบบการทำนายแต่ละตัวแบบที่สร้างขึ้น (testing dataset) จำนวน 40 รายการ คิดเป็นร้อยละ 30 ของข้อมูลทั้งหมด โดยตัวแบบการเรียนรู้สำหรับทำนายเป็นตัวแบบการเรียนรู้แบบมีผู้สอนประเภทการจำแนกประเภทข้อมูล (classification) ซึ่งเลือกวิธีการจำแนกแบบถดถอย (regression) เพื่อต้องการทำนายคลาสที่เป็นตัวเลข และนำตัวแบบที่สร้างขึ้นมาทดสอบกับชุดข้อมูลสำหรับทดสอบเพื่อประเมินประสิทธิภาพของตัวแบบ ดังภาพที่ 2



ภาพที่ 2 กรอบแนวคิดการออกแบบตัวแบบการทำนาย (ดัดแปลงจาก [8])

3.2 การพัฒนาตัวแบบการทำนาย ในการพัฒนาตัวแบบการเรียนรู้ของเครื่องโดยใช้ scikit-learn ซึ่งเป็นไลบรารีฟรีในภาษาไพธอน (Python) ตัวแบบการเรียนรู้ของเครื่อง จำนวน 3 ตัวแบบ ได้แก่ การถดถอยเชิงเส้น (linear regression) เพื่อนบ้านใกล้เคียง (K-nearest neighbor: KNN) และต้นไม้ป่าสุ่ม (random forest) ซึ่งทั้ง 3 ตัวแบบ เป็นตัวแบบการทำนายในการจำแนกประเภทข้อมูลและเป็นตัวแบบการเรียนรู้แบบมีผู้สอน (supervised learning)

3.2.1 การถดถอยเชิงเส้น เป็นการเรียนรู้ของเครื่องแบบมีผู้สอน โดยใช้ความสัมพันธ์เชิงเส้นที่เกี่ยวข้องกับการศึกษาและวิเคราะห์ตัวแปรอิสระที่มีอิทธิพลต่อการเปลี่ยนแปลงของตัวแปรตาม [6] ซึ่งสามารถอธิบายรูปแบบการถดถอย ดังสมการที่ 2

$$Y = \alpha + \beta x \quad (2)$$

โดย Y คือ ตัวแปรตาม; x คือ ตัวแปรอิสระหรือค่าคงที่; α คือ สัมประสิทธิ์ของความถดถอยที่เป็นจุดตัดบนแกน Y และ β คือ สัมประสิทธิ์ของความถดถอยที่เป็นค่าความชันของเส้นตรง

3.2.2 เพื่อนบ้านใกล้เคียง เป็นวิธีการจำแนกประเภทแบบมีผู้สอน โดยใช้วิธีแบ่งกลุ่มซึ่งมีการแบ่งออกเป็นชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบ โดยกระบวนการทำงานในแต่ละชุดข้อมูลจะหาระยะห่างของแต่ละแอททริบิวต์ในข้อมูลที่เรียกว่าระยะทางแบบยุคลิดกับชุดข้อมูลฝึกฝน [10] ดังสมการที่ 3

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

โดย $\text{dist}(x, y)$ คือ ระยะห่างระหว่างข้อมูลตัวอย่าง x กับตัวอย่าง y ; n คือ จำนวนคุณลักษณะทั้งหมดของข้อมูลตัวอย่าง; x คือ คุณลักษณะตัวที่ i ของข้อมูลตัวอย่าง x และ y คือ คุณลักษณะตัวที่ i ของข้อมูลตัวอย่าง y

การพัฒนาตัวแบบการทำนายโดยใช้เพื่อนบ้านใกล้เคียงมีการกำหนดค่าพารามิเตอร์ 10 จำนวน โดยผลการพัฒนาตัวแบบการทำนาย ดังตารางที่ 1

ตารางที่ 1 ผลการพัฒนาตัวแบบโดยใช้เพื่อนบ้านใกล้เคียง

ที่	ค่าพารามิเตอร์	ค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสอง	ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน
1	1	4.05	16.40
2	5	3.60	12.96
3	9	3.46	11.99
4	13	3.42	11.75
5	17	3.43	11.78
6	21	3.38	11.43
7	25	3.38	11.45
8	29	3.39	11.50
9	33	3.40	11.59
10	37	3.41	11.66

3.2.3 ต้นไม้ป่าสุ่ม เป็นวิธีการจำแนกประเภทข้อมูลฝึกฝนตัวแบบที่เหมือนกันหลาย ๆ ครั้ง (instance) บนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการฝึกฝนจะเลือกส่วนของข้อมูลที่ฝึกฝนไม่เหมือนกัน และการตัดสินใจของตัวแบบมาจากคลาสที่ถูกเลือกมากที่สุด [11]

การพัฒนาตัวแบบการทำนายโดยใช้ต้นไม้ป่าสุ่มมีการกำหนดค่าพารามิเตอร์จำนวนต้นไม้ในตัวแบบการทำนาย ($n_estimators$) และจำนวนลำดับขั้นของต้นไม้ตัดสินใจ (max_depth) จำนวน 6 ตัวแปร โดยผลการพัฒนาตัวแบบการทำนาย ดังตารางที่ 2

ตารางที่ 2 ผลการพัฒนาตัวแบบการทำนายโดยใช้ต้นไม้ป่าส้ม

ที่	จำนวนต้นไม้ใน ตัวแบบการทำนาย	จำนวนลำดับชั้นของ ต้นไม้ตัดสินใจ	ค่าเฉลี่ยของรากที่สองของ ความคลาดเคลื่อนกำลังสอง	ค่าเฉลี่ยของค่าสัมบูรณ์ ของความคลาดเคลื่อน
1	50	5	3.11	9.68
2	100	5	3.17	10.09
3	150	5	3.11	9.71
4	50	10	3.17	10.05
5	100	10	3.13	9.82
6	150	10	3.14	9.88

4. การประเมินประสิทธิภาพตัวแบบ

ในการทำนายข้อมูลจะมีความคลาดเคลื่อนเกิดขึ้นทุกครั้ง ความถูกต้องของค่าทำนายจะมากหรือน้อยขึ้นอยู่กับค่าความคลาดเคลื่อนของการพยากรณ์ (error) ซึ่งมีผลต่างของค่าจริงกับค่าทำนาย [12] ในการวัดประสิทธิภาพของการทำนายทำได้โดยการหาค่าความคลาดเคลื่อนของการทำนาย ดังต่อไปนี้

4.1 ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน (mean absolute error: MAE)

เป็นวิธีการวัดค่าความคลาดเคลื่อนที่บอกถึงขนาดของความคลาดเคลื่อนรวมในการวัดค่าความแม่นยำ หากค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้มีความแม่นยำมาก ดังสมการที่ 4

$$\text{ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน} = \frac{1}{N} \sum_{i=1}^N |y_1 - \bar{y}_1| \quad (4)$$

โดย N คือ จำนวนข้อมูลที่ใช้; y คือ ค่าจริง และ $\bar{y}$ คือ ค่าที่ได้จากการทำนาย

4.2 ค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสอง (root mean squared error: RMSE) เป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐานในการวัดค่าความแม่นยำ หากค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสองมีค่าน้อยแสดงว่าตัวแบบที่ได้มีความแม่นยำมาก ดังสมการที่ 5

$$\text{ค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสอง} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \bar{y})^2} \quad (5)$$

โดย N คือ จำนวนข้อมูลที่ใช้; y คือ ค่าจริง และ $\bar{y}$ คือ ค่าที่ได้จากการทำนาย

เกณฑ์การแปลผลการประเมินประสิทธิภาพของตัวแบบการทำนายโดยใช้ค่าร้อยละของความคลาดเคลื่อนสัมบูรณ์ แสดงดังตารางที่ 3 [13]

ตารางที่ 3 เกณฑ์การแปลผลการประเมินประสิทธิภาพของตัวแบบการทำนาย

ที่	ค่าร้อยละเฉลี่ย	การแปลผล
1	0-10	ตัวแบบการทำนายมีประสิทธิภาพอยู่ในเกณฑ์ดีเยี่ยม
2	11-20	ตัวแบบการทำนายมีประสิทธิภาพอยู่ในเกณฑ์ดี
3	21-50	ตัวแบบการทำนายมีประสิทธิภาพอยู่ในเกณฑ์ดีปานกลาง
4	51-100	ตัวแบบการทำนายมีความผิดพลาดมาก

ขั้นตอนที่ 2 ทำนายราคาพริกขี้หนูในจังหวัดนครศรีธรรมราช สำหรับใช้ในการวางแผนการเพาะปลูกพริกขี้หนูในพื้นที่ลุ่มน้ำปากพนัง จังหวัดนครศรีธรรมราช โดยนำตัวแบบการเรียนรู้ของเครื่องทั้ง 3 ตัวแบบมาทำนายราคาพริกขี้หนู โดยใช้ scikit-learn

ผลการวิจัย

1. ผลการเปรียบเทียบตัวแบบการทำนาย

จากการเปรียบเทียบตัวแบบการทำนายราคาพริกขี้หนูพบว่า ต้นไม้ป่าสุ่มมีค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อนน้อยกว่าการถดถอยเชิงเส้น และเพื่อนบ้านใกล้เคียง อยู่ที่ 3.11 และ 9.68 แสดงให้เห็นว่าต้นไม้ป่าสุ่มมีผลการประเมินประสิทธิภาพของตัวแบบการทำนายอยู่ในเกณฑ์ดีเยี่ยม ดังตารางที่ 4 ด้วยต้นไม้ป่าสุ่มเป็นตัวแบบการเรียนรู้แบบกลุ่มที่สุ่มสร้างข้อมูลการทดสอบที่แตกต่างกัน และในการวิเคราะห์การถดถอยใช้วิธีการทำนายหาค่าเฉลี่ยโดยนำค่าที่ทำนายของทุกต้นไม้มัดตติสินใจมาคำนวณหาค่าเฉลี่ยเพื่อแสดงเป็นค่าการทำนาย

ตารางที่ 4 ผลการเปรียบเทียบตัวแบบการทำนายราคาพริกขี้หนู

ที่	ตัวแบบการทำนาย	ค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนกำลังสอง	ค่าเฉลี่ยของค่าสัมบูรณ์ของความคลาดเคลื่อน
1	การถดถอยเชิงเส้น	3.42	11.71
2	เพื่อนบ้านใกล้เคียง	3.38	11.43
3	ต้นไม้ป่าสุ่ม	3.11	9.68

2. ผลการทำนายราคาพริกขี้หนู

จากผลการทำนายราคาพริกขี้หนูเฉลี่ยรายเดือน 12 เดือน โดยใช้ต้นไม้ป่าสุ่มเปรียบเทียบกับราคาเฉลี่ยรายเดือนจริงของพริกขี้หนูพบว่า ราคามีความใกล้เคียงกับราคาจริง ดังตารางที่ 5

ตารางที่ 5 ผลการทำนายราคาพริกขี้หนูเฉลี่ยรายเดือนต่อกิโลกรัมในระยะเวลา 12 เดือน

เดือน	ค่าจริง	ค่าที่ทำนายโดยต้นไม้ป่าสุ่ม	ค่าที่ทำนายโดยการถดถอยเชิงเส้น	ค่าที่ทำนายโดยเพื่อนบ้านใกล้เคียง
ม.ค.	78.75	80.63	79.82	81.46
ก.พ.	76.88	78.68	79.26	81.46
มี.ค.	69.38	70.94	66.54	75.00
เม.ย.	65.00	71.36	75.01	77.29
พ.ค.	85.63	84.74	75.67	86.25
มิ.ย.	81.25	81.97	84.04	83.13
ก.ค.	83.75	84.18	92.12	83.13
ส.ค.	88.75	86.84	79.92	84.38
ก.ย.	84.38	85.17	87.09	87.61
ต.ค.	92.81	90.61	87.60	88.65
พ.ย.	104.38	104.33	115.00	107.92
ธ.ค.	135.63	121.46	124.53	107.92

เปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่องสำหรับทำนายราคาพริกขี้หนูในจังหวัดนครศรีธรรมราช

อภิปรายผลการวิจัย

งานวิจัยนี้ได้สร้างตัวแบบในการทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช ผลการวิจัยพบว่า ต้นไม้ป่าสุ่มเป็นตัวแทนที่มีความเหมาะสมสำหรับการทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราช เนื่องจากช่วยจำแนกข้อมูลได้อย่างต่อเนื่อง ผลการทำนายแม่นยำ และลดความซ้ำซ้อนของข้อมูล ทำให้สร้างตัวทำนายที่เหมาะสมและแม่นยำ ผลการศึกษาสอดคล้องกับการศึกษาการประเมินประสิทธิภาพของต้นไม้ป่าสุ่มกับการถดถอยพหุคูณสำหรับการคาดการณ์ราคาของอพาร์ทเมนต์ [5] การทำนายทิศทางราคาน้ำมันปาล์มโดยใช้ต้นไม้ป่าสุ่มที่สามารถทำนายราคาน้ำมันปาล์มได้แม่นยำ 91.11% [14] การคาดคะเนราคาตัวสำหรับการขนส่งสาธารณะโดยใช้การถดถอยเชิงเส้นและต้นไม้ป่าสุ่ม แนวทางปฏิบัติโดยใช้การเรียนรู้ของเครื่อง พบว่า ต้นไม้ป่าสุ่มให้ความแม่นยำในการคาดคะเนขึ้นทั้งชุดข้อมูลการฝึกและการทดสอบ 79.06% และ 80.10% ตามลำดับ [15] และการคาดการณ์ราคาขายรถยนต์โดยใช้ต้นไม้ป่าสุ่ม อัลกอริทึมการเรียนรู้ของเครื่องพบว่า ต้นไม้ป่าสุ่มและต้นไม้ตัดสินใจจำนวนมาก (extra tree regression) สามารถทำนายราคาขายรถมือสองได้อย่างแม่นยำโดยไม่คำนึงถึงขนาดของชุดข้อมูล [16]

สรุปผลการวิจัย

ต้นไม้ป่าสุ่มเป็นตัวแทนสำหรับการทำนายราคาพริกชี้หนูในจังหวัดนครศรีธรรมราชที่เหมาะสมที่สุดในงานวิจัยครั้งนี้ จากข้อมูลที่ใช้ในการเรียนรู้ จำนวน 7 ตัวแปร ได้แก่ เดือนที่เก็บข้อมูลราคา (2) ปริมาณน้ำฝนเฉลี่ยรายเดือน (3) ความชื้นเฉลี่ยรายเดือน (4) อุณหภูมิเฉลี่ยรายเดือน (5) ฤดูกาลที่เก็บข้อมูลราคา (6) ราคาพริกชี้หนูเฉลี่ยรายเดือนต่อกิโลกรัม และ (7) ราคาพริกชี้หนูเฉลี่ยรายปีต่อกิโลกรัม ให้ค่าประสิทธิภาพของตัวแบบได้ดีที่สุด คือ ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์เท่ากับ 9.68 ซึ่งสามารถนำมาประยุกต์ใช้เป็นตัวแทนในการทำนายราคาผักได้ อย่างไรก็ตามตัวแบบที่สร้างขึ้นในงานวิจัยครั้งนี้มีความเหมาะสมกับชุดข้อมูลสภาพภูมิอากาศและราคาผักของจังหวัดนครศรีธรรมราชเท่านั้น ผู้ใช้งานตัวแบบต้องพิจารณาปัจจัยต่าง ๆ เพิ่มเติม เช่น สภาพดิน สภาพแวดล้อม ศัตรูพืช โรคพืช เป็นต้น เพราะค่าที่ทำนายได้จากตัวแบบเป็นเพียงสารสนเทศส่วนหนึ่งที่ช่วยประกอบการตัดสินใจในการวางแผนการปลูกพริก นอกจากนี้ผู้วิจัยควรศึกษาตัวแบบการจำแนกประเภทข้อมูลอื่น เช่น ต้นไม้ตัดสินใจแบบเบย์ หรือโครงข่ายประสาทเทียม เป็นต้น เพื่อเพิ่มประสิทธิภาพในการจำแนกประเภทข้อมูล และนำตัวแบบที่เหมาะสมไปพัฒนาเว็บแอปพลิเคชันสำหรับอำนวยความสะดวกให้แก่เกษตรกรในการวางแผนการปลูกพริกต่อไป

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับความอนุเคราะห์ข้อมูลจากกลุ่มงานอากาศเกษตรนครศรีธรรมราช สถานีอุตุนิยมวิทยานครศรีธรรมราช (บางจาก) สำนักงานพาณิชย์จังหวัดนครศรีธรรมราช ขอขอบคุณคณะวิทยาศาสตร์และเทคโนโลยี และบัณฑิตวิทยาลัย มหาวิทยาลัยราชภัฏนครศรีธรรมราช ที่ให้การสนับสนุน

เอกสารอ้างอิง

- [1] สาธิต บัวขาว. (2564). รายงานกิจกรรมส่งเสริมและสนับสนุนการวิจัย การถ่ายทอดองค์ความรู้และเทคโนโลยีการปลูกพริกตามมาตรฐานการปฏิบัติทางการเกษตรที่ดี (GAP) ที่สอดคล้องกับความต้องการของตลาด. กรุงเทพฯ: สำนักงานการวิจัยแห่งชาติ (วช.).
- [2] ศูนย์ติดตามและแก้ไขปัญหาภัยพิบัติด้านการเกษตร. (2563). แผนป้องกันและบรรเทาสาธารณภัยด้านการเกษตรในช่วงฤดูแล้ง ปี 2562/63. นครศรีธรรมราช: ศูนย์ติดตามและแก้ไขปัญหาภัยพิบัติด้านการเกษตร จังหวัดนครศรีธรรมราช.
- [3] สำนักงานเศรษฐกิจการเกษตร. (2559). แผนการพัฒนาการเกษตรในช่วงแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 12 (พ.ศ. 2560 - 2564). กรุงเทพฯ: กระทรวงเกษตรและสหกรณ์.
- [4] ศิรินทรา เสือพิทักษ์, บุญมี นิสสัยดี, และวิภาวรรณ บัวทอง. (2564). การเปรียบเทียบเทคนิคการจำแนกประเภทข้อมูลสำหรับทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียน. *วารสารวิชาการชาชนันท์ มหาวิทยาลัยราชภัฏภูเก็ต*, 5(2), 42-50.
- [5] Čeh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS international journal of geo-information*, 7(5), 168.
- [6] จิตกานต์ จันทราช, มนทิราลัย ชัยมงคล, รัตนชัย แซ่ไฉ่, สายทิพย์ พลอยสัมฤทธิ์, และสายชล สีนสมบูรณ์ทอง. (2563). การเปรียบเทียบประสิทธิภาพการทำนายผลการจำแนกกรณีข้อมูลสูญหายด้วยเทคนิคการทำเหมืองข้อมูล. *Thai Journal of Science and Technology*, 9(1), 1-15.
- [7] สุรพงศ์ เอื้อวัฒนามงคล. (2561). *การทำเหมืองข้อมูล (พิมพ์ครั้งที่ 2)*. กรุงเทพฯ: สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- [8] เอกสิทธิ์ พชรวงศ์ศักดิ์. (2563). *A little Book of Big Data and Machine Learning*. นนทบุรี: ไอดีซี พรีเมียร์.
- [9] ผุสดี บุญรอด, และกรวัฒน์ พลเยี่ยม. (2560). แบบจำลองการพยากรณ์ราคามันสำปะหลังโดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น. *วารสารวิทยาศาสตร์และเทคโนโลยี*. 25(3), 533-543.
- [10] Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert systems with Applications*, 42(20), 7046-7056.
- [11] Ayub, U., & Moqurrah, S. A. (2018). Predicting crop diseases using data mining approaches: classification (pp. 1-6). In *2018 1st International Conference on Power, Energy and Smart Grid (Icpesg)*. IEEE.
- [12] เฉลิมชาติ อีระวิริยะ. (2560). การเปรียบเทียบวิธีการพยากรณ์สำหรับความต้องการใช้พลังงานไฟฟ้าในจังหวัดนครพนม. *Naresuan University Journal: Science and Technology*, 25(4), 124-137.
- [13] Moreno, J. J. M., Pol, A. P., Abad, A. S., & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25(4), 500-506.
- [14] Myat, A. K., & Tun, M. T. Z. (2019). Predicting palm oil price direction using random forest (pp. 1-6). In *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)*. IEEE.

- [15] Dutta, A., Dureja, A., Abrol, S., & Dureja, A. (2019.). Prediction of ticket prices for public transport using linear regression and random forest regression methods: A practical approach using machine learning (pp. 140-150). In *International Conference on Recent Developments in Science, Engineering and Technology*. Springer, Singapore.
- [16] Pandey, A., Rastogi, V., & Singh, S. (2020). Car's selling price prediction using random st machine learning algorithm. In *5th International Conference on Next Generation Computing Technologies (NGCT-2019)*. India.