

การวิเคราะห์ข้อมูลเพื่อพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ด้วยขั้นตอนวิธีถดถอย

Data Analytics for Maize Price Prediction using Regression Algorithms

ภวิดา จันทันแก้ว (Pawita Jantankaew)* นวลวรรณ สุนทรภิชัย (Nuanwan Soonthornphisaj)^{1**}

(Received: March 30, 2022; Revised: August 1, 2022; Accepted: August 3, 2022)

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อการศึกษาแบบจำลองที่มีประสิทธิภาพต่อการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ที่เกษตรกรขายได้ โดยใช้ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ ขั้นตอนวิธีถดถอยแบบริดจ์ และขั้นตอนวิธีถดถอยแบบลาโซ เพื่อเรียนรู้จากชุดข้อมูลที่สำนักงานเศรษฐกิจการเกษตรได้รวบรวม ในช่วงเดือน มกราคม พ.ศ. 2545 ถึง พฤษภาคม พ.ศ. 2562 ผู้วิจัยได้สร้างพีเจอร์ใหม่ จำนวน 2 พีเจอร์ ได้แก่ อัตราการเปลี่ยนแปลงของราคามันสำปะหลังย้อนหลัง 1 เดือน และอัตราการเปลี่ยนแปลงของราคาข้าวโพดเลี้ยงสัตว์ย้อนหลัง 1-4 เดือน การวิจัยครั้งนี้ได้ทำการวิเคราะห์ความสัมพันธ์ของข้อมูลและวัดประสิทธิภาพของขั้นตอนวิธีด้วยการเปรียบเทียบอาร์กำลังสอง ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ตามลำดับ ผลการวิจัยพบว่าการคัดเลือกพีเจอร์ก่อนการเรียนรู้ด้วยขั้นตอนวิธีถดถอย ส่งผลให้ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ มีค่าอาร์กำลังสองเท่ากับ 0.94 ขั้นตอนวิธีถดถอยแบบริดจ์มีค่าอาร์กำลังสองเท่ากับ 0.86 ขั้นตอนวิธีถดถอยลาโซมีค่าอาร์กำลังสองเท่ากับ 0.86 โดยทั้งสามขั้นตอนวิธีมีความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 0.31, 0.42, 0.43 ตามลำดับ และมีค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.50, 0.58 และ 0.58 ตามลำดับ

ABSTRACT

This research aims for finding the regression model that can predict maize price sold by Thai farmers. Three regression models are explored which are multiple linear regression, Ridge regression and Lasso regression. These algorithms learn from dataset collected by office of agricultural economics from Jan 2002-May 2019. We propose two new features which are the rate of cassava price change from 1 month and the rate of maize price change from 1-4 months. We do statistical analysis to see the relationship between features. Performance of regression algorithms are measured in terms of R-squared, Root mean square error and Mean absolute error. The experimental results reveal that feature selection play an important role for multiple linear regression with the R-squared = 0.94. We found that multiple linear regression outperforms Ridge regression (R- Squared = 0.86) and Lasso regression (R- Squared = 0.86). The mean absolute error of multiple linear regression, Ridge regression and Lasso algorithm are 0.31, 0.42 and 0.43, respectively. The root mean square error of these three regression algorithms are 0.50, 0.58 and 0.58, respectively.

คำสำคัญ: ขั้นตอนวิธีถดถอย ราคาข้าวโพดเลี้ยงสัตว์ การเรียนรู้ของเครื่อง วิทยาการข้อมูล

Keywords: Regression algorithm, Maize price, Machine learning, Data Science

¹Corresponding author: nuanwan.s@ku.th

*นิสิต หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

**รองศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

บทนำ

ราคาสินค้าเกษตรมีความสำคัญต่อการวางแผนและการกำหนดนโยบายของภาครัฐและภาคธุรกิจเป็นอย่างมาก เนื่องจากราคาเป็นหนึ่งในข้อมูลใช้อุปทานของสินค้าเกษตร ทั้งนี้ข้อมูลสินค้าเกษตรค่อนข้างมีปัจจัยที่ส่งผลกระทบต่อ การเปลี่ยนแปลงราคาสินค้าอยู่เป็นประจำตามรอบการผลิต ฤดูกาล หรือราคาและผลิตภัณฑ์สินค้าทดแทน โดยเฉพาะอย่างยิ่ง สินค้าที่สามารถทดแทนกันได้ เช่น ข้าวโพดเลี้ยงสัตว์และมันสำปะหลังโรงงาน ซึ่งทั้ง 2 สินค้าจะถูกนำไปแปรรูปเป็น อาหารสัตว์ ดังนั้นการมีเครื่องมือที่สามารถอำนวยความสะดวกในการทำนายข้อมูลราคาสินค้าเกษตรได้รวดเร็วทันต่อการ เปลี่ยนแปลงย่อมเกิดประโยชน์ต่อภาคการเกษตรเป็นอย่างมาก

สำนักงานเศรษฐกิจการเกษตร (สศก.) กระทรวงเกษตรและสหกรณ์ มีพันธกิจในการเสนอแนะนโยบาย และจัดทำ ยุทธศาสตร์ แผนพัฒนา และมาตรการทางการเกษตร รวมทั้งจัดทำทำที่และร่วมเจรจาการค้า สินค้าเกษตรและความ ร่วมมือด้านเศรษฐกิจการเกษตรระหว่างประเทศ ซึ่ง สศก. มีเพียงการจัดเก็บข้อมูลราคาข้าวโพดเลี้ยงสัตว์เท่านั้นแต่ไม่ สามารถพยากรณ์ราคาสินค้าเกษตรล่วงหน้าได้ งานวิจัยนี้จึงได้ทำการศึกษาขั้นตอนวิธีถดถอยแบบต่างๆ เพื่อพยากรณ์ ราคาข้าวโพดเลี้ยงสัตว์ล่วงหน้าเพื่อให้เป็นประโยชน์ต่อการตัดสินใจของ 3 ภาคส่วน ดังนี้

1. เกษตรกร สามารถนำข้อมูลไปใช้ในการตัดสินใจว่าจะขายหรือเก็บผลผลิตตากแห้งไว้ก่อนเพื่อรอการ จำหน่ายเมื่อคาดว่าราคาจะดีขึ้น หรือการเพาะปลูกในรอบต่อไปควรยังคงปลูกข้าวโพดเลี้ยงสัตว์ต่อหรือเปลี่ยนการ เพาะปลูกเป็นสินค้าอื่นทดแทน
2. ภาคธุรกิจ สามารถนำข้อมูลไปใช้ในการวางแผนด้านผลิตและการตลาด เช่นการปรับสูตรการผลิตผลิตภัณฑ์ จากข้าวโพด การเก็บสต็อกสินค้าที่เป็นปัจจัยการผลิตและการแปรรูป
3. ภาครัฐ สามารถนำข้อมูลไปใช้ในการวางแผนเพื่อการรับมือกับกรณีการคาดการณ์ว่าจะมีการเปลี่ยนแปลงของราคาได้ทัน ต่อสถานการณ์ เช่นการคาดการณ์ว่าจะเกิดเหตุการณ์ภาวะราคาสินค้าตกต่ำ

งานวิจัยนี้ใช้ข้อมูลจาก สศก. ที่จัดเก็บข้อมูลในช่วงเดือนมกราคม พ.ศ. 2545 ถึง เดือนพฤษภาคม พ.ศ. 2562 และ ได้ทำการศึกษางานวิจัยของ Ge Y และ Wu H เพื่อเป็นแนวทางในการวิจัย พบว่าได้ที่มีการประยุกต์ขั้นตอนวิธีการถดถอย เชิงเส้นพหุคูณเพื่อพยากรณ์ราคาข้าวโพด โดยอาศัยหลักการของอุปสงค์และอุปทานโดยงานวิจัย ได้ใช้ปริมาณการบริโภค ข้าวโพดในสาธารณรัฐประชาชนจีน ปริมาณการนำเข้าข้าวโพดและและปริมาณส่งออกข้าวโพดเป็นตัวแปรต้นในการสร้าง โมเดล โดยใช้ข้อมูลในช่วงปี ค.ศ. 2005 - 2016 ผลการวิจัยพบว่าการใช้อุปสงค์และอุปทานสำหรับสินค้าเกษตรเพียงชนิด เดียวนั้น ไม่สามารถทำให้การพยากรณ์ราคาข้าวโพดมีความแม่นยำได้ [1] นอกจากนั้นการใช้ข้อมูลสภาพอากาศเพื่อ นำไปสู่การทำนายจำนวนผลผลิตซึ่งควรจะส่งผลกระทบต่อราคาขาย ได้ถูกนำมาศึกษาในงานวิจัยของ Ayush J. และคณะ ซึ่งได้ ประยุกต์โครงข่ายประสาทเทียมชนิด Long short term memory ซึ่งเป็น Recurrent Neural Networks เพื่อพยากรณ์ ราคาข้าวโพดในตลาดค้าส่งข้าวโพดจำนวน 7 แห่งในรัฐ Karnataka ประเทศอินเดีย โดยใช้ข้อมูลสภาพภูมิอากาศ ได้แก่ อุณหภูมิ ความชื้น และปริมาณน้ำฝน ประกอบกับราคาข้าวโพดย้อนหลังจำนวน 997 วัน โดยใช้ข้อมูลช่วงปี ค.ศ. 2016 - กุมภาพันธ์ ค.ศ. 2019 เป็นข้อมูลฝึก และทดสอบด้วยข้อมูลเดือนกุมภาพันธ์ ค.ศ. 2019 ถึง กรกฎาคม ค.ศ. 2019 เป็น เวลาจริงรวม 128 วัน ผลการทดลองพบว่าได้ค่าเฉลี่ยของรากที่สองของค่าคลาดเคลื่อนที่ดีที่สุดเท่ากับ 277.79 รูปี ต่อ 100 กิโลกรัม [2]

การทำนายราคาข้าวโพดโดยอาศัยเฉพาะราคาข้าวโพดเป็นฟีเจอร์เพียงตัวเดียวได้มีการศึกษาวิจัยโดย Roberto F และคณะ ซึ่งประยุกต์ขั้นตอนวิธีเอตาบูสต์ โดยใช้ซัพพอร์ตเวกเตอร์รีเกรสชัน เป็นขั้นตอนวิธีพื้นฐานภายในของเซมเบล ในการพยากรณ์ราคาข้าวโพดรายวันในประเทศบราซิลโดยใช้ข้อมูลราคาข้าวโพดในช่วงปี ค.ศ. 2003 - 2019 ผลการทดลอง พบว่า ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของขั้นตอนวิธีเอตาบูสต์ มีค่าเท่ากับ 0.335 เรียลต่อตัน [3]

นอกจากนี้ Ouyang H และคณะ ได้ทำการทดลองเปรียบเทียบประสิทธิภาพการพยากรณ์ราคาซื้อขายสินค้าเกษตรล่วงหน้า (future price) จำนวน 12 ชนิดในปี ค.ศ. 2006 ด้วยโครงข่ายประสาทเทียมชนิด Long short term memory ผลการทดลองพบว่าขั้นตอนวิธี Long short term memory ให้ค่าประสิทธิภาพดีที่สุด [4] นอกจากนี้ขั้นตอนวิธีถดถอยลาซโซ ได้ถูกนำมาประยุกต์เพื่อพยากรณ์ผลผลิตเฉลี่ยข้าวโพดในสามรัฐ ของประเทศสหรัฐอเมริกา ได้แก่ รัฐอิลลินอยส์ รัฐอินเดียนา และรัฐโอไฮโอ จากการเรียนรู้ด้วยชุดข้อมูลที่เก็บในช่วงปี ค.ศ. 2000-2018 ซึ่งประกอบด้วย จำนวนต้นข้าวโพดต่อพื้นที่เพาะปลูก 1 เอเคอร์ อัตราการปลูกข้าวโพด ปริมาณน้ำฝน อุณหภูมิ ผลการทดลองพบว่าขั้นตอนวิธีถดถอยลาซโซ สามารถพยากรณ์ปริมาณผลผลิตได้ดีที่สุด [5]

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่านักวิจัยศึกษาข้อมูลจากตัวแปรเดียว เช่น ราคาย้อนหลัง ยังไม่ได้มีการศึกษาความสัมพันธ์หรืออิทธิพลระหว่างสินค้าเกษตรต่างชนิดกัน ดังนั้นงานวิจัยนี้จึงมีจุดมุ่งหมายในการศึกษาความสัมพันธ์ระหว่างราคาข้าวโพดและสินค้าเกษตรอื่นในรูปวัตถุดิบและสินค้าแปรรูป ได้แก่ แป้งมันสำปะหลัง มันเส้น มันอัดเม็ด สาकुขาว รวมถึงน้ำสภาพภูมิอากาศคือ ปริมาณน้ำฝน มาพิจารณาเพื่อสร้างโมเดลด้วยขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ (Linear regression) ขั้นตอนวิธีถดถอยลาซโซ (Lasso regression) และขั้นตอนวิธีถดถอยแบบบริดจ์ (Ridge regression) ตามลำดับ

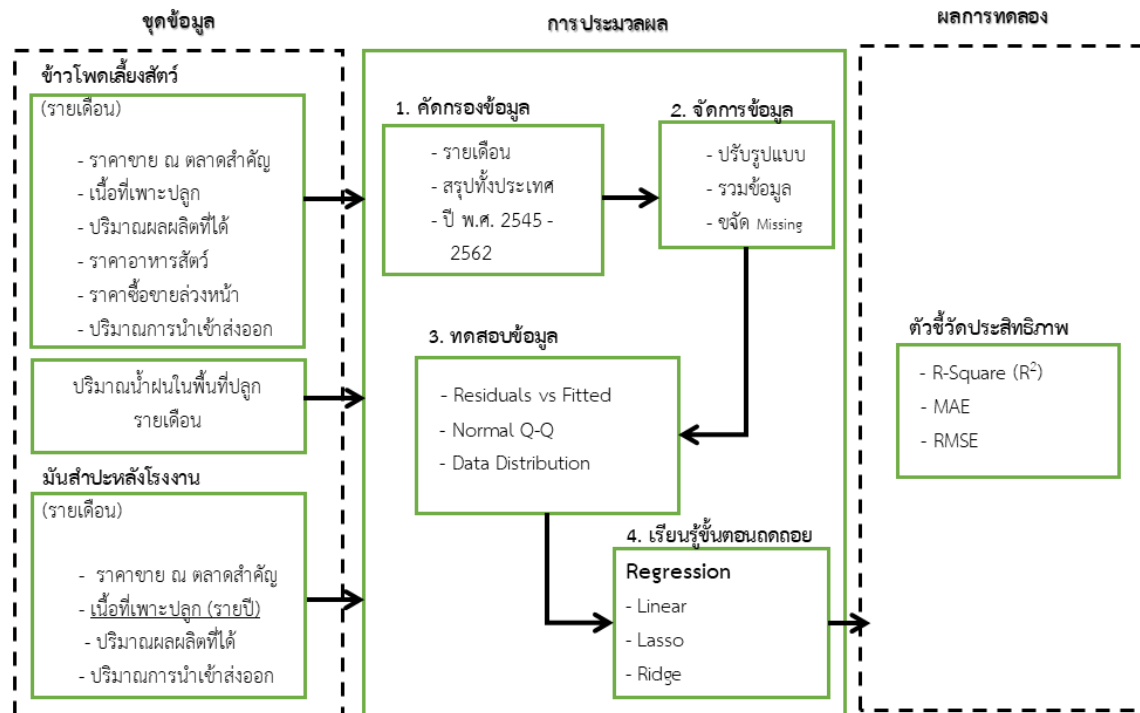
วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างโมเดลราคาข้าวโพดเลี้ยงสัตว์ด้วยขั้นตอนวิธีถดถอย โดยได้ศึกษาขั้นตอนวิธีถดถอย 3 วิธี ได้แก่ ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ ขั้นตอนวิธีถดถอยลาซโซ และขั้นตอนวิธีถดถอยแบบบริดจ์ นอกจากนี้ยังได้ศึกษาผลกระทบของการคัดเลือกฟีเจอร์ว่ามีความจำเป็นหรือไม่สำหรับขั้นตอนวิธีถดถอยดังกล่าว

วิธีดำเนินการวิจัย

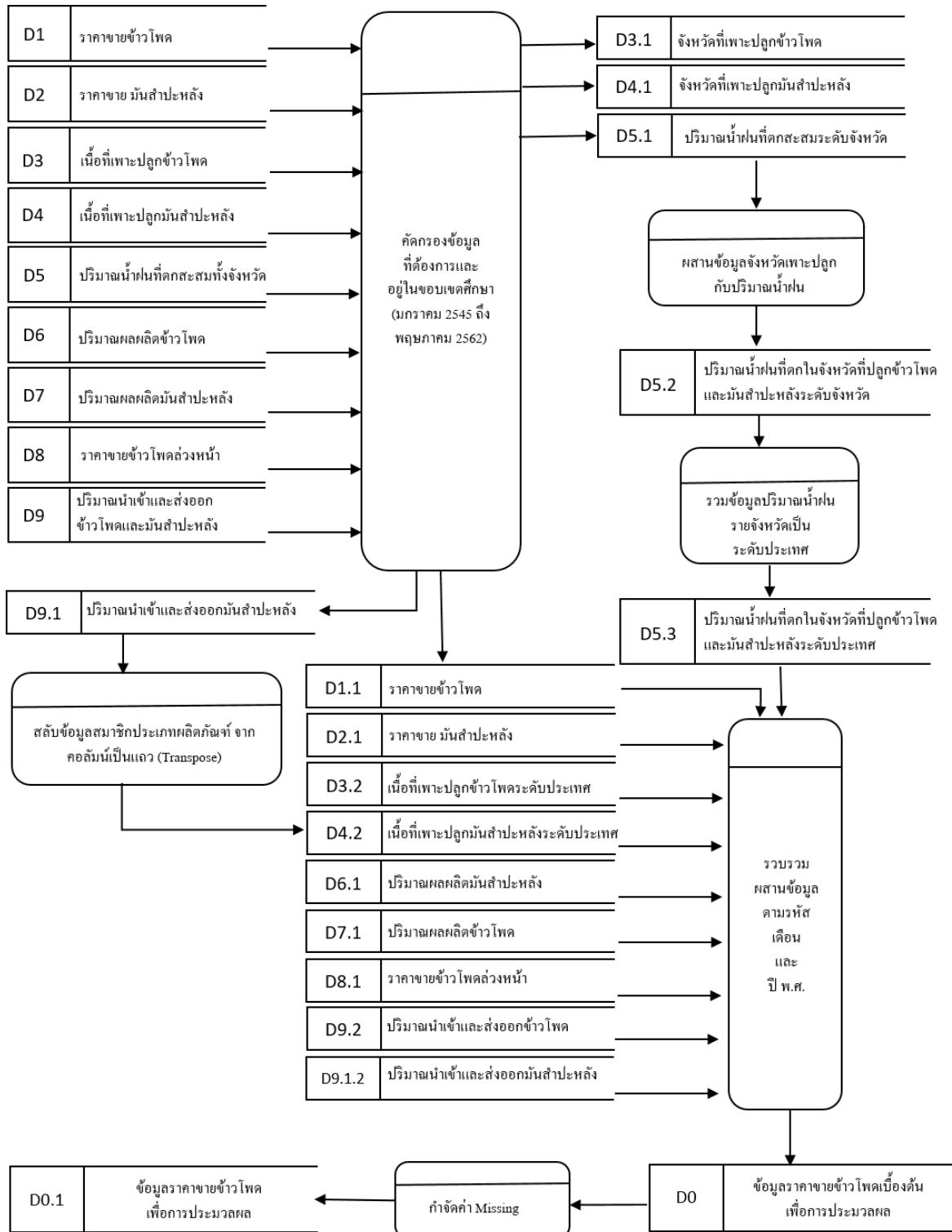
1. กรอบแนวคิด

ขั้นตอนการวิจัยแบ่งออกเป็น 6 ขั้นตอน คือ รวบรวมข้อมูล เตรียมข้อมูล วิเคราะห์ข้อมูล สร้างโมเดลจากขั้นตอนวิธีที่เหมาะสม วัดประสิทธิภาพโมเดล สุดท้ายคือสรุปผลและข้อเสนอแนะโดยมีการวางกรอบการดำเนินงานดังภาพที่ 1



ภาพที่ 1 กรอบการดำเนินงานวิจัย

ผู้วิจัยได้สร้างชุดข้อมูลโดยการรวบรวมข้อมูลซึ่งได้จัดเก็บในหลายรูปแบบประกอบด้วยข้อมูลจากฐานข้อมูล และ ตารางสรุปรายงานผลเป็นแฟ้ม Excel ข้อมูลจากฐานข้อมูลที่ได้จัดเตรียมแฟ้มในรูปแบบ csv หลังจากนั้นทำการคัดกรองข้อมูลเฉพาะส่วนที่อยู่ในขอบเขตของการศึกษา เช่น ปริมาณ และมูลค่า ในช่วงเวลาขอบเขตการศึกษา และปรับข้อมูลให้สามารถนำมาใช้ในการวิเคราะห์และการประมวลผลรวมทั้งสามารถนำมาใช้เป็นตัวแปรในการสร้างโมเดลได้เช่น ตารางที่มีข้อมูลคร่อมปีจัดให้อยู่ในปีเดียว ข้อมูลที่ใช้ชื่อจังหวัดได้เปลี่ยนเป็นรหัสจังหวัดเพื่อนำไปผสมเข้ากับข้อมูลปริมาณน้ำฝนเพื่อคำนวณหาปริมาณน้ำฝนที่ตกในพื้นที่ปลูก และผลิตภัณฑ์ที่ใช้เป็นรหัสทำการปรับให้อยู่ในรูปคอลัมน์ตามประเภทของผลิตภัณฑ์ รายละเอียดแสดงในภาพที่ 2 จากนั้นทำการทดสอบข้อมูลเบื้องต้นด้วยวิธีการ Residual vs Fitted, วิธี Normal Q-Q และดูการกระจายของข้อมูล (Data distribution) หลังจากนั้นทำการเรียนรู้ด้วยขั้นตอนวิธีถดถอย ได้แก่ ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ ขั้นตอนวิธีถดถอยลาซโซ และ ขั้นตอนวิธีถดถอยแบบริดจ์ และทำการวัดประสิทธิภาพของโมเดลด้วยค่าอาร์กำลังสอง ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ตามลำดับ



ภาพที่ 2 แผนภาพการไหลของข้อมูล (Data Flow Diagram)

2. ฟังก์ชันในชุดข้อมูล

ผู้วิจัยได้สร้างฟังก์ชันใหม่จำนวน 2 ฟังก์ชัน ได้แก่ $MaiChg_n$ และ $CasChg$ เพื่อแสดงค่าอัตราการเปลี่ยนแปลงของราคาข้าวโพดเลี้ยงสัตว์ย้อนหลังไป n เดือน และราคามันสำปะหลังย้อนหลังไป 1 เดือน โดยการคำนวณค่าการเปลี่ยนแปลงราคาข้าวโพดย้อนหลัง n เดือน ($MaiChg_n$) เป็นไปตามสมการที่ 1 ดังนี้

$$MaiChg_n = \frac{(MaiPrice_0 - MaiPrice_n)}{MaiPrice_0} \quad (1)$$

โดยที่ $MaiPrice_0$ คือราคาขายข้าวโพดเลี้ยงสัตว์ ณ เวลาปัจจุบัน

$MaiPrice_n$ คือราคาขายข้าวโพดเลี้ยงสัตว์ย้อนหลังไป n เดือน ในงานวิจัยนี้จะใช้ราคาย้อนหลัง 1-4 เดือน

พีเจอร์ $CasChg$ หมายถึงอัตราการเปลี่ยนแปลงของราคามันสำปะหลังย้อนหลังไป 1 เดือน เป็นไปตามสมการที่ 2 ดังนี้

$$CasChg = \frac{(CasPrice_0 - CasPrice_1)}{CasPrice_0} \quad (2)$$

โดยที่ $CasPrice_0$ คือราคามันสำปะหลัง ณ เดือนที่คำนวณ

$CasPrice_1$ คือราคามันสำปะหลังย้อนหลังไป 1 เดือน

ผู้วิจัยพบว่าพีเจอร์ $CasChg$ มีค่า P-value เป็น 0.151 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.27 ในขณะที่พีเจอร์ $MaiChg_n$ ย้อนหลังไปเดือนที่ 4 มีความสำคัญต่อราคาขายข้าวโพดเลี้ยงสัตว์ค่อนข้างมาก คือมีค่า P-value เป็น 5.20e-05 และค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็น 0.261 ตามลำดับ

3. การวิเคราะห์ข้อมูล

ผู้วิจัยได้ทำการศึกษาลักษณะของพีเจอร์ในชุดข้อมูล เพื่อให้สามารถเลือกใช้ขั้นตอนวิธีที่เหมาะสม โดยได้ใช้วิธีการทางสถิติ จากชุดข้อมูลทั้งหมดจำนวน 209 ตัวอย่าง ประกอบด้วย 46 พีเจอร์ ดังแสดงในตารางที่ 1

ตารางที่ 1 พีเจอร์ในชุดข้อมูล

(*หมายถึงพีเจอร์ที่งานวิจัยนี้เสนอ)

#	ชื่อพีเจอร์	ความหมาย	#	ชื่อพีเจอร์	ความหมาย
1	MaiPlant	พื้นที่เพาะปลูกข้าวโพด	24	CasImpVal4	มูลค่าการนำเข้าสาคุ
2	MaiPrice	ราคาข้าวโพดที่เกษตรกรขายได้	25	CasImpOut5	ปริมาณการนำเข้าหัวมันสด
3	MaiOut	ปริมาณผลผลิตข้าวโพดที่เก็บเกี่ยวได้	26	CasImpVal5	มูลค่าการนำเข้าหัวมันสด
4	MaiRain	ปริมาณน้ำฝนในพื้นที่เพาะปลูกข้าวโพด	27	CasImpOut6	ปริมาณการนำเข้ากาก
5	MaiRainPlant	ปริมาณน้ำฝนเฉลี่ยต่อพื้นที่ปลูกข้าวโพด	28	CasImpVal6	มูลค่าการนำเข้ากาก
6	MaiFutPrice	ราคาขายข้าวโพดตลาดล่วงหน้า(ชิคาโก)	29	CasImpOut7	ปริมาณการนำเข้าแป้งมัน สำปะหลังดัดแปร
7	MailmpOut	ปริมาณการนำเข้าข้าวโพด	30	CasImpVal7	มูลค่าการนำเข้าแป้งมัน สำปะหลังดัดแปร
8	MailmpVal	มูลค่าการนำเข้าข้าวโพด	31	CasExpOut1	ปริมาณการส่งออกแป้งมัน สำปะหลัง
9	MaiExpOut	ปริมาณการส่งออกข้าวโพด	32	CasExpVal1	มูลค่าการส่งออกแป้งมัน สำปะหลัง
10	MaiExpVal	มูลค่าการส่งออกข้าวโพด	33	CasExpOut2	ปริมาณการส่งออกมันเส้น
11	CasPlant	พื้นที่เพาะปลูกมันสำปะหลัง	34	CasExpVal2	มูลค่าการส่งออกมันเส้น
12	CasPrice	ราคามันสำปะหลังที่เกษตรกรขายได้รายเดือน	35	CasExpOut3	ปริมาณการส่งออกมันอัดเม็ด
13	CasChg	อัตราการเปลี่ยนแปลงราคามันสำปะหลัง จากเดือนที่แล้ว	36	CasExpVal3	มูลค่าการส่งออกมันอัดเม็ด

ตารางที่ 1 พีเจอร์ในชุดข้อมูล (ต่อ)

(*หมายถึงพีเจอร์ที่งานวิจัยนี้เสนอ)

#	ชื่อพีเจอร์	ความหมาย	#	ชื่อพีเจอร์	ความหมาย
14	CasQut	ปริมาณผลผลิตมันสำปะหลังที่เก็บเกี่ยวได้	37	CasExpQut4	ปริมาณการส่งออกสาธู
15	CasRain	ปริมาณน้ำฝนในพื้นที่เพาะปลูกมันสำปะหลัง	38	CasExpVal4	มูลค่าการส่งออกสาธู
16	CasRainPPlant	ปริมาณน้ำฝนเฉลี่ยต่อพื้นที่ปลูกมันสำปะหลัง	39	CasExpQut5	ปริมาณการส่งออกหัวมันสด
17	CasImpQut1	ปริมาณการนำเข้าแป้งมันสำปะหลัง	40	CasExpVal5	มูลค่าการส่งออกหัวมันสด
18	CasImpVal1	มูลค่าการนำเข้าแป้งมันสำปะหลัง	41	CasExpQut6	ปริมาณการส่งออกกาว
19	CasImpQut2	ปริมาณการนำเข้ามันเส้น	42	CasExpVal6	มูลค่าการส่งออกกาว
20	CasImpVal2	มูลค่าการนำเข้ามันเส้น	43	CasExpQut7	ปริมาณการส่งออกแป้งมันสำปะหลังดัดแปร
21	CasImpQut3	ปริมาณการนำเข้ามันอัดเม็ด	44	CasExpVal7	มูลค่าการส่งออกแป้งมันสำปะหลังดัดแปร
22	CasImpVal3	มูลค่าการนำเข้ามันอัดเม็ด	45*	Maichgn	การเปลี่ยนแปลงของราคาข้าวโพดเลี้ยงสัตว์ย้อนหลัง
23	CasImpQut4	ปริมาณการนำเข้าสาธู	46*	Caschg	การเปลี่ยนแปลงของราคามันสำปะหลังย้อนหลัง 1 เดือน

1) การทดสอบความสัมพันธ์ระหว่างพีเจอร์ใดๆ (f_i) กับค่าเป้าหมายที่ต้องการพยากรณ์ (ราคาข้าวโพดเลี้ยงสัตว์: $MaiPrice_i$) ด้วยวิธีสหสัมพันธ์ (Correlation) โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (r) แสดงดังสมการที่ 3

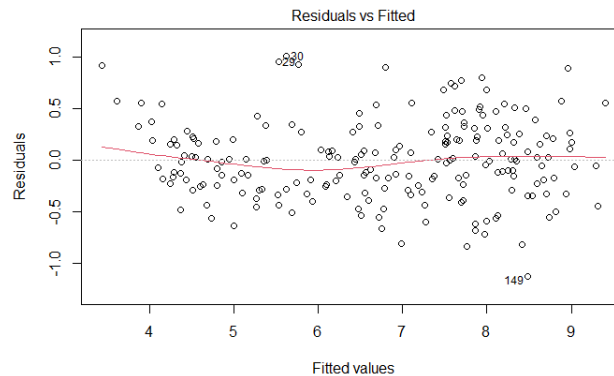
$$r = \frac{\sum_{i=1}^n (f_i - \bar{f})(MaiPrice_i - \bar{MaiPrice})}{\sqrt{\sum_{i=1}^n (f_i - \bar{f})^2 \sum_{i=1}^n (MaiPrice_i - \bar{MaiPrice})^2}} \quad (3)$$

ผลการวิเคราะห์พบว่า ราคาขายข้าวโพดตลาดล่วงหน้าฟิวเจอร์ (MaiFutPrice) และ ราคาขายมันสำปะหลัง (CasPrice) มีความสัมพันธ์ในทิศทางเดียวกับราคาข้าวโพดที่เกษตรกรขายได้ในประเทศด้วยสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.854 และ 0.8133 ตามลำดับ

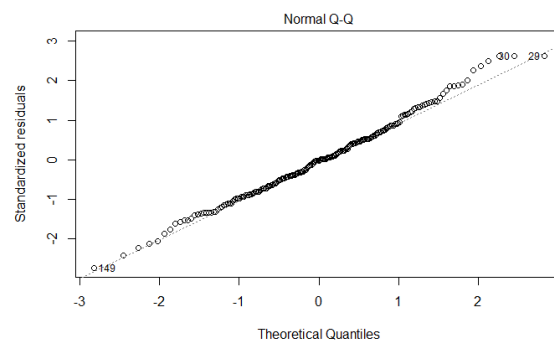
นอกจากนี้ยังพบว่าพีเจอร์พื้นที่ปลูกมันสำปะหลัง (CasPlant) มีความสัมพันธ์กับพีเจอร์ 3 ตัวได้แก่ มูลค่าการส่งออกแป้งมันสำปะหลัง (CasExpVal1) ปริมาณการส่งออกแป้งมันสำปะหลัง (CasExpQut1) และมูลค่าการนำเข้าแป้งมันสำปะหลังดัดแปร (CasImpVal7) ด้วยสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.8121 , 0.7803 และ 0.7750 ตามลำดับ

2) การวิเคราะห์ข้อมูล ความคลาดเคลื่อนของตัวแบบ โดยได้ทำการสร้างกราฟเพื่อทดสอบลักษณะความคลาดเคลื่อนเมื่อใช้ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณซึ่งเรียนรู้จากชุดข้อมูลโดยใช้พีเจอร์จำนวนทั้งสิ้น 46 ตัว เพื่อทดสอบลักษณะความคลาดเคลื่อน ดังภาพที่ 3 แสดงการทำนายตัวแบบราคาข้าวโพดเลี้ยงสัตว์กับพีเจอร์ทั้งหมดพบว่า เส้นพยากรณ์สีแดงมีการไต่อยู่ใกล้ที่เส้นกึ่งกลาง (เส้นประ) และภาพที่ 4 แสดงกราฟการกระจายของราคาข้าวโพดเลี้ยงสัตว์กับส่วนเหลือ (Normal Q-Q) มีลักษณะค่อนข้างไปทางเส้นตรง และได้ทำการพล็อตการกระจายของข้อมูลราคาข้าวโพด

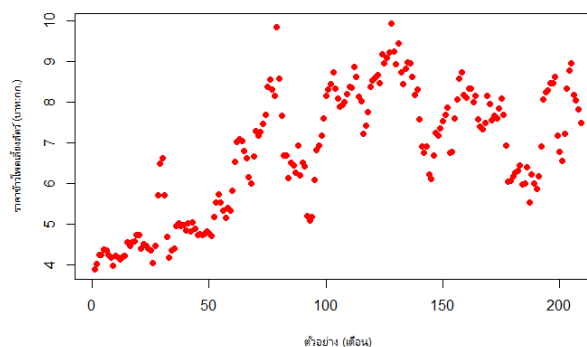
เลี้ยงสัตว์ด้วยการสร้างกราฟ ดังภาพที่ 5 กราฟแสดงการกระจายของข้อมูลราคาข้าวโพดเลี้ยงสัตว์ จากภาพที่ 5 พบว่าการกระจายข้อมูลมีลักษณะค่อนข้างไปทางเชิงเส้นตรง



ภาพที่ 3 กราฟแสดงการทำนายตัวแบบราคาข้าวโพดเลี้ยงสัตว์กับพีเจอร์ทั้งหมด

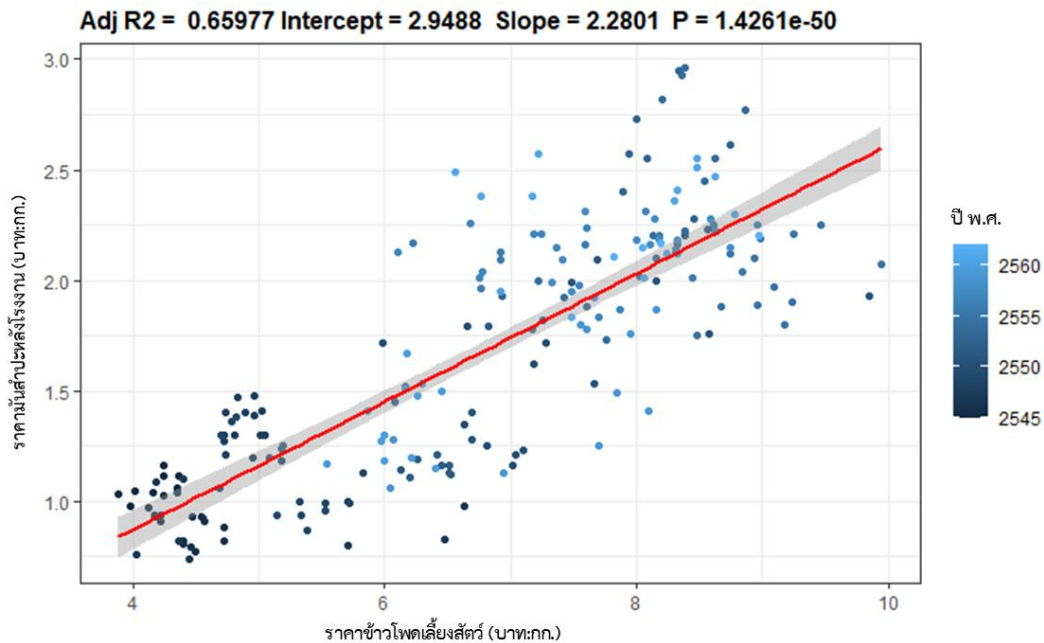


ภาพที่ 4 กราฟแสดงการกระจายของราคาข้าวโพดเลี้ยงสัตว์กับส่วนเหลือ



ภาพที่ 5 กราฟแสดงการกระจายของข้อมูลราคาข้าวโพดเลี้ยงสัตว์

จากภาพที่ 5 จะเห็นได้ว่าการกระจายของข้อมูลราคาข้าวโพดเลี้ยงสัตว์ก็มีลักษณะค่อนข้างไปในทิศทางเส้นตรงเช่นกัน จากค่าอาร์กำลังสอง และ Adjusted R-squared ประกอบกับภาพกราฟแสดงการกระจายข้อมูลในภาพที่ 3 - 5 ทำให้ผู้วิจัยเลือกใช้ขั้นตอนวิธีถดถอยเชิงเส้นเพื่อพยากรณ์ราคาข้าวโพด นอกจากนั้นยังได้ให้ความสนใจในเรื่องการเป็นสินค้าทดแทนกัน จึงได้ทำการทดสอบสินค้าทดแทนด้วยการนำข้อมูลราคาข้าวโพดเลี้ยงสัตว์และราคามันสำปะหลังมาทำการสร้างกราฟด้วยขั้นตอนถดถอยเชิงเส้นพหุคูณ เพื่อการวิเคราะห์ความสัมพันธ์ระหว่าง 2 ตัวแปร ดังปรากฏในภาพที่ 6 กราฟแสดงความสัมพันธ์ระหว่างข้อมูลราคาข้าวโพดเลี้ยงสัตว์กับราคามันสำปะหลัง



ภาพที่ 6 กราฟแสดงความสัมพันธ์ระหว่างข้อมูลราคาข้าวโพดเลี้ยงสัตว์กับราคามันสำปะหลัง

จากภาพที่ 6 พบว่า ในช่วงก่อนปี พ.ศ. 2550 ราคาข้าวโพดเลี้ยงสัตว์และมันสำปะหลังยังมีการเกาะกลุ่มใกล้เคียงกัน ซึ่งแสดงถึงความสัมพันธ์ที่มีต่อกัน แต่หลังจากปี พ.ศ. 2550 ข้อมูลเริ่มมีการกระจายตัวมากขึ้นเรื่อยๆ จึงได้ทำการค้นคว้าข้อมูลเพิ่มเติมพบว่า ข้าวโพดเลี้ยงสัตว์สามารถนำไปแปรรูปได้ไม่มากนัก เช่น แคลปซูลยา ขณะที่มันสำปะหลังโรงงานสามารถแปรรูปได้หลากหลายประเภทโดยเฉพาะการแปรรูปเป็นเอทานอล ซึ่งเป็นพลังงานทดแทนที่ประเทศไทยนิยมใช้ผลิตภายในประเทศในระดับต้นๆ รองจากกากน้ำตาล

การออกแบบการทดลอง

จากการวิเคราะห์ข้อมูลเบื้องต้นพบว่าการกระจายข้อมูลแบบเชิงเส้นตรง ผู้วิจัยจึงได้พิจารณาใช้ขั้นตอนวิธีถดถอยเชิงเส้น ดังต่อไปนี้

1. ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ

ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ [6] มีวัตถุประสงค์เพื่อสร้างเส้นตรงที่เป็นตัวแทนของข้อมูลฝึก โดยกำหนดให้ตัวอย่างในชุดข้อมูลประกอบด้วยฟีเจอร์ ($f_1, f_2, \dots, f_k$) การทำงานของขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณจะสร้างสมการเส้นตรงเพื่อพยากรณ์ค่าเป้าหมาย ($Maiprice_{predicted}$) ดังนี้

$$Maiprice_{predicted} = w_1f_1 + w_2f_2 + \dots w_kf_k + w_0 \quad (4)$$

โดยที่ $w_1, \dots, w_k$ เป็นสัมประสิทธิ์ที่ขั้นตอนวิธีจะต้องคำนวณหาค่าที่เหมาะสมที่สุดที่ทำให้ค่า SSE ต่ำที่สุด

$$SSE = \sum_{i=1}^n (Maiprice_{label} - Maiprice_{predicted})^2 \quad (5)$$

2. ขั้นตอนวิธีถดถอยลาซโซ

ขั้นตอนวิธีถดถอยลาซโซ ได้ปรับปรุงฟังก์ชันการหาสัมประสิทธิ์ที่เหมาะสมที่สุดด้วยการเพิ่มพจน์ penalty เพื่อค้นหาตัวแบบที่เหมาะสมที่สุดสำหรับชุดข้อมูลฝึก ตัวแบบที่ได้จากขั้นตอนวิธีถดถอยลาซโซเป็นไปตามสมการ

$$Maiprice_{predicted} = w_1f_1 + w_2f_2 + \dots w_nf_k + w_0 \quad (6)$$

$$\hat{w} = \operatorname{argmin}_w (\|y - \sum_{j=1}^k f_j w_j\|^2 + \text{penalty}(w_j)) \quad (7)$$

$$\text{penalty}(w_j) = \lambda \sum_{j=1}^k |w_j| \quad (8)$$

ผลลัพธ์ของวิธีการวิเคราะห์การถดถอยแบบลาซโซจะได้ค่าสัมประสิทธิ์ ($\hat{w}$) ของพีเจอร์ส่วนใหญ่เป็นศูนย์ซึ่งหมายความว่าพีเจอร์เหล่านั้นไม่ถูกนำมาใช้ในการพยากรณ์ค่าเป้าหมายนั่นเอง ส่วนพารามิเตอร์ λ โดยทั่วไปจะใช้วิธีการตรวจสอบไขว้กัน (cross validation) ในการหาค่าที่เหมาะสมสำหรับข้อมูลที่ต้องการวิเคราะห์ [7]

3. ขั้นตอนวิธีถดถอยแบบบริดจ์

ขั้นตอนวิธีถดถอยแบบบริดจ์ [8] ใช้ฟังก์ชัน penalty ดังสมการที่ 9 ต่อไปนี้

$$\text{penalty}(w_j) = \lambda \sum_{j=1}^k w_j^2 \quad (9)$$

โดยขั้นตอนวิธีถดถอยแบบบริดจ์มีวัตถุประสงค์เพื่อหาค่า ($\hat{w}$) ที่ทำให้สมการต่อไปนี้มีค่าน้อยที่สุด

$$\hat{w} = \operatorname{argmin}_w (\|y - \sum_{j=1}^k f_j w_j\|^2 + \text{penalty}(w_j)) \quad (10)$$

การใช้ฟังก์ชัน penalty ดังกล่าวจะเป็นการดึงค่าประมาณของสัมประสิทธิ์เข้าสู่ศูนย์ ทำให้ได้ค่า ($\hat{w}$) ทุกตัวที่ค่าสัมประสิทธิ์ต่ำ ดังนั้นขั้นตอนวิธีถดถอยแบบบริดจ์จึงเหมาะสำหรับการวิเคราะห์ข้อมูลที่มีสัมประสิทธิ์ขนาดเล็กที่ไม่เท่ากับศูนย์จำนวนมาก

ถ้า λ มีค่าเป็นศูนย์ จะส่งผลให้การทำงานของขั้นตอนวิธีถดถอยแบบบริดจ์ ได้โมเดลที่เหมือนกับโมเดลถดถอยเชิงเส้นพหุนามที่มีวัตถุประสงค์ให้ค่า SSE ต่ำสุด

ถ้า λ มีค่ามากจะส่งผลให้ค่าสัมประสิทธิ์ ($\hat{w}$) ของพีเจอร์ที่เป็นผลลัพธ์ของการทำงานของ มีค่าเข้าใกล้ศูนย์ [9]

การสร้างตัวแบบหรือการเรียนรู้ของขั้นตอนวิธีและประเมินประสิทธิภาพจะใช้ข้อมูลที่ได้ทำการแบ่งออกเป็น 2 ชุด ดังนี้

- ชุดข้อมูลฝึก ประกอบด้วยข้อมูลจำนวน 167 ตัวอย่าง คิดเป็น 80% ของข้อมูลทั้งหมด สำหรับใช้ในการเรียนรู้เพื่อสร้างตัวแบบ
- ชุดข้อมูลทดสอบ ประกอบด้วยข้อมูล 42 ตัวอย่าง คิดเป็น 20% สำหรับการทดสอบประสิทธิภาพ

หมายเหตุ ในการทดลองได้มีการนำวิธีการสกัดลักษณะเด่นเพื่อคัดเลือกพีเจอร์ก่อนการเรียนรู้ด้วยขั้นตอนวิธีถดถอย

วิธีวัดประสิทธิภาพ

การวัดประสิทธิภาพของตัวแบบที่ได้จากขั้นตอนวิธีถดถอยใช้ค่าความถูกต้องและความคลาดเคลื่อนโดยความถูกต้องนั้นใช้ค่าอาร์กำลังสอง (R^2) เป็นเครื่องมือวัด ส่วนความคลาดเคลื่อนใช้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสอง

เฉลี่ย (RMSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เพื่อเปรียบเทียบผลว่าขั้นตอนวิธีถดถอยแบบใดเหมาะสมกับชุดข้อมูลนี้

ค่า R^2 แสดงดังสมการที่ 11

$$R^2 = 1 - \frac{\sum_{i=1}^n (MaiPrice_{label,i} - MaiPrice_{predict,i})^2}{\sum_{i=1}^n (MaiPrice_{label,i} - MaiPrice_{mean,i})^2} \quad (11)$$

กรณีที่ R^2 มีค่าเข้าใกล้ 1 หมายความว่าราคาขายข้าวโพดเลี้ยงสัตว์ที่ขั้นตอนวิธีถดถอยพยากรณ์มีค่าใกล้เคียงกับค่าเฉลี่ย ส่วนค่า $RMSE$ แสดงดังสมการที่ 12 และค่า MAE แสดงดังสมการที่ 13

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (MaiPrice_{label,i} - MaiPrice_{predict,i})^2} \quad (12)$$

$$MAE = \frac{\sum_{i=1}^n |MaiPrice_{label,i} - MaiPrice_{predict,i}|}{n} \quad (13)$$

ผลการวิจัย

การทดลองที่ 1

ผู้วิจัยได้ทำการทดลองเพื่อค้นหาขั้นตอนวิธีถดถอยที่สามารถทำนายราคาข้าวโพดเลี้ยงสัตว์โดยใช้ไลบรารีจากโปรแกรม R รวมถึงการใช้ฟังก์ชัน stepAIC เพื่อคัดเลือกฟีเจอร์ที่สามารถสร้างโมเดลที่ดีที่สุด โดยขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณจะทำการทดลองเปรียบเทียบระหว่างการใช้ฟีเจอร์ทั้งหมด ($Model_{allfeature}$) และการใช้เฉพาะฟีเจอร์ที่ผ่านการคัดเลือก ($Model_{select}$) สำหรับขั้นตอนวิธีถดถอยแบบบริดจ์และลาโซจะเรียนรู้จากทุกฟีเจอร์เนื่องจากทั้งสองขั้นตอนวิธีดังกล่าวมีกลไกการคัดเลือกฟีเจอร์อยู่ภายในกระบวนการเรียนรู้ โดยมีผลการทดลองแสดงในตารางที่ 2

ตารางที่ 2 เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีถดถอย

ตัววัด	Linear Regression $Model_{allfeature}$	Linear Regression $Model_{select}$	Ridge Regression	Lasso Regression
R^2	0.86	0.94	0.86	0.86
MAE	0.43	0.31	0.43	0.43
RMSE	0.58	0.50	0.58	0.58

ผลการวัดประสิทธิภาพด้วยค่าอาร์กำลังสอง พบว่าโมเดลที่ได้จากขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณที่ใช้ฟีเจอร์ที่ผ่านขั้นตอนการคัดเลือกฟีเจอร์ ($Model_{select}$) สามารถทำนายราคาข้าวโพดได้ดีที่สุดโดยมีค่า R^2 เท่ากับ 0.94 และมีความคลาดเคลื่อนน้อยที่สุดซึ่งวัดจากค่า MAE และค่า RMSE มีค่าเป็น 0.31 และ 0.50 ตามลำดับ

จากกราฟในภาพที่ 7 ซึ่งแสดงผลการทำนายโดยใช้ขั้นตอนวิธีถดถอย 4 ขั้นตอนวิธี พบประเด็นที่น่าสนใจคือ ถึงแม้โมเดลที่ได้จากขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณที่ผ่านกลไกการคัดเลือกฟีเจอร์จะมีผลการทำนายที่ใกล้เคียงค่าเฉลี่ย

มากกว่าวิธีอื่น แต่เมื่อขั้นตอนวิธีเรียนรู้ไปเรื่อยๆ พบว่าในเดือนสุดท้ายโมเดล $Model_{allfeature}$ ขั้นตอนวิธีถดถอยแบบบริดจ์และลาซโซ สามารถทำนายค่าได้ใกล้เคียงกับค่าเฉลี่ยเช่นกัน โดยสามารถพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ที่แตกต่างจากค่าเฉลี่ยเพียง 0.02 บาทต่อกิโลกรัม ดังแสดงรายละเอียดในตารางที่ 3

ราคาข้าวโพด (บาท)



ภาพที่ 7 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยและผลการทำนายทั้ง 4 ขั้นตอนวิธี

ตารางที่ 3 ผลการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ในเดือนสุดท้ายด้วยขั้นตอนวิธีถดถอยแบบต่างๆ เทียบกับผลเฉลี่ยที่ราคา 8.06 บาทต่อกิโลกรัม

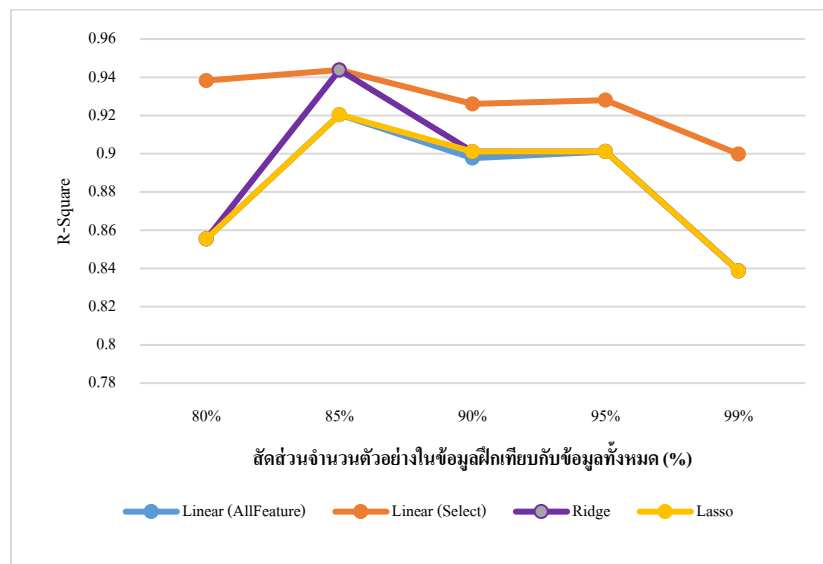
ราคาข้าวโพดเลี้ยงสัตว์	Linear Regression $Model_{allfeature}$	Linear Regression $Model_{select}$	Ridge Regression	Lasso Regression
ค่าพยากรณ์จากขั้นตอนวิธี	8.04	7.91	8.04	8.04
แตกต่างจากค่าเฉลี่ย (8.06 บาท/กก)	0.02	0.14	0.02	0.02

การทดลองที่ 2

ผู้วิจัยได้ทำการแบ่งข้อมูลเพื่อทดสอบความถูกต้องของการพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์จากขั้นตอนวิธีถดถอยด้วยค่าอาร์ก้าสอง โดยทำการแบ่งข้อมูลฝึกเป็น 85%, 90%, 95% และ 99% ตามลำดับ ผลการทดลองพบว่าโมเดล $Model_{select}$ เป็นโมเดลที่สามารถพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ได้ดีที่สุด ดังแสดงในตารางที่ 4 และ ภาพที่ 8

ตารางที่ 4 ค่าอาร์กำลังสองของขั้นตอนวิธีถดถอยที่เรียนรู้จากจำนวนตัวอย่างในข้อมูลฝึกที่แตกต่างกัน

ข้อมูลฝึก		Linear Regression	Linear Regression	Ridge Regression	Lasso Regression
%	จำนวนตัวอย่าง	Model _{allfeature}	Model _{select}		
80	167	0.86	0.94	0.86	0.855
85	177	0.92	0.94	0.94	0.92
90	188	0.90	0.93	0.90	0.901
95	198	0.90	0.93	0.90	0.901
99	206	0.84	0.90	0.84	0.84



ภาพที่ 8 กราฟเปรียบเทียบค่าอาร์กำลังสองของขั้นตอนวิธีถดถอยที่ใช้ขนาดของข้อมูลฝึกที่แตกต่างกัน

สรุปและอภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบจากขั้นตอนวิธีถดถอยเพื่อพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ เป็นการศึกษาข้อมูลจากข้อมูลที่มีการรวบรวมและสรุปทั้งประเทศเป็นรายเดือน ระหว่างเดือนมกราคม ปี พ.ศ. 2545 ถึง พฤษภาคม ปี พ.ศ. 2562 ประกอบด้วยข้อมูลราคาข้าวโพดเลี้ยงสัตว์และมันสำปะหลังโรงงาน ปริมาณที่ผลิตได้ เนื้อที่เพาะปลูก เนื้อที่เพาะปลูกมันสำปะหลังโรงงานรายปี ปริมาณและมูลค่าการนำเข้าส่งออกของข้าวโพดเลี้ยงสัตว์และมันสำปะหลังโรงงาน ข้อมูลปริมาณน้ำฝนในพื้นที่ปลูกเป็นต้น โดยมีตัวแปรทำนายคือราคาข้าวโพดเลี้ยงสัตว์ และมีฟีเจอร์จำนวน 46 ฟีเจอร์ ข้อมูลมีจำนวน 209 เดือนหรือตัวอย่าง โดยมีเป้าหมายของการวิจัยคือการวิเคราะห์ความสัมพันธ์ของฟีเจอร์และการเลือกขั้นตอนวิธีถดถอยหรือตัวแบบที่เหมาะสมกับข้อมูลชุดนี้ซึ่งมีคุณลักษณะเป็นข้อมูลเชิงปริมาณ ตัวแปรทำนายและฟีเจอร์ที่มีความสัมพันธ์กันแบบเชิงเส้นตรง จึงเลือกใช้ขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณ จำนวน 2 ตัวแบบคือตัวแบบที่สร้างจากทุกฟีเจอร์ (Model_{allfeature}) และตัวแบบที่สร้างจากการคัดเลือกฟีเจอร์ (Model_{select}) และใช้ขั้นตอน

วิธีการเรียนรู้ของเครื่องคือขั้นตอนวิธีถดถอยแบบบริดจ์ และขั้นตอนวิธีถดถอยลาซโซ เพื่อค้นหาขั้นตอนวิธีที่สามารถทำนายราคาข้าวโพดเลี้ยงสัตว์ได้อย่างมีประสิทธิภาพที่สุด

ผลการวิเคราะห์เพื่อหาความสัมพันธ์ระหว่างตัวแปรทั้งหมดพบว่า ข้อมูลมีความสัมพันธ์กันที่มีค่าระดับความน่าเชื่อถือทางสถิติระหว่าง 0.75 ถึง 1 มีดังนี้ ราคาข้าวโพดเลี้ยงสัตว์กับราคาขายข้าวโพดตลาดล่วงหน้า (ชิคาโก) (0.854) ราคาข้าวโพดเลี้ยงสัตว์กับราคามันสำปะหลังโรงงานมีค่าเป็น (0.813) มูลค่าการส่งออกแป้งมันสำปะหลังพื้นที่เพาะปลูกมันสำปะหลังมีค่าเป็น 0.812 ปริมาณการส่งออกแป้งมันสำปะหลังกับพื้นที่เพาะปลูกมันสำปะหลังมีค่าเป็น 0.78 และมูลค่าการนำเข้าแป้งมันสำปะหลังดัดแปรกับพื้นที่เพาะปลูกมันสำปะหลังมีค่าเป็น 0.775 อีกทั้งยังพบว่าข้อมูลที่มีความสัมพันธ์กันแบบผกผันที่มีค่าน้อยกว่า -5 คือ พื้นที่เพาะปลูกข้าวโพดกับปริมาณการส่งออกสาธุมมีค่าเป็น -0.652 และปริมาณน้ำฝนเฉลี่ยต่อพื้นที่ปลูกมันสำปะหลังกับปริมาณผลผลิตมันสำปะหลังที่เก็บเกี่ยวได้มีค่าเป็น -0.653

ผลการทดลองพบว่าการสร้างพีเจอรี่ใหม่ ได้แก่ *CasChg* และ *MaiChg_n* พบว่าเป็นพีเจอรี่ที่มีผลต่อการพยากรณ์ราคาขายข้าวโพดเลี้ยงสัตว์เนื่องจากถูกเลือกโดยขั้นตอน StepAIC เพื่อสร้างโมเดล ($Model_{select}$) ซึ่งเป็นโมเดลที่มีประสิทธิภาพสูงสุด จากการทดสอบค่า P-value ที่ได้พบว่า *MaiChg₄* มีค่าเป็น $< 2e-16$ ส่วน *CasChg* มีค่าเป็น 0.021 และ *MaiChg₁* มีค่าเป็น 0.087

ผู้วิจัยพบว่าการพยากรณ์ราคาข้าวโพดด้วยขั้นตอนวิธีถดถอยเชิงเส้นพหุคูณจากชุดข้อมูลที่มีพีเจอรี่ที่เกี่ยวข้องกับสินค้าเกษตรในรูปวัตถุดิบและสินค้าแปรรูปที่ทดแทนกันได้ จะสามารถสะท้อนอุปสงค์และอุปทานของข้าวโพดเลี้ยงสัตว์ จึงทำให้สามารถสร้างตัวแบบที่สามารถพยากรณ์ราคาข้าวโพดได้ใกล้เคียงกับค่าเฉลี่ยได้ ซึ่งเป็นไปตามข้อเสนอแนะของงานวิจัย [1] ที่ได้ให้ความเห็นว่าการพยากรณ์ราคาควรนำปัจจัยที่เกี่ยวข้องกับปริมาณความต้องการใช้วัตถุดิบมาสร้างเป็นพีเจอรี่ในชุดข้อมูล

อย่างไรก็ตามชุดข้อมูลที่ใช้ศึกษาวิจัยนี้ยังขาดข้อมูลเกี่ยวกับต้นทุนการผลิตเช่น ค่าแรงงาน ค่าเชื้อเพลิง รวมถึงข้อมูลด้านความต้องการข้าวโพดเลี้ยงสัตว์ ซึ่งจำเป็นต้องมีข้อมูลเกี่ยวกับปริมาณการเลี้ยงปศุสัตว์และการประมงที่เป็นสัตว์เศรษฐกิจ การส่งออกอาหารสัตว์เป็นต้น รวมถึงอัตราแลกเปลี่ยนเงินสกุลบาทและสกุลดอลลาร์ที่อาจส่งผลกระทบต่อราคาและปริมาณการนำเข้าส่งออกสินค้าและผลิตภัณฑ์ข้าวโพดเลี้ยงสัตว์ได้ ดังนั้นจึงควรนำข้อมูลต่างๆ เหล่านี้มาพิจารณาร่วมเพื่อให้ข้อมูลสำหรับการศึกษาครั้งต่อไปมีความสมบูรณ์มากยิ่งขึ้น

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ สำนักงานเศรษฐกิจการเกษตรที่ได้อนุเคราะห์ชุดข้อมูลสำหรับดำเนินการวิจัย และขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์สำหรับทรัพยากรที่ใช้ในการทำวิจัย

เอกสารอ้างอิง

1. Ge Y, Wu H. Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Comput & Applic.* 2019; 32:16843–16855.
2. Ayush J, Marvaniya S, Godbole S, Munigala V. A framework for crop price forecasting in emerging economies by analyzing the quality of time-series data. *arXiv.* 2020:10.
3. Silva RF, Barreira BL, Cugnasca CE. Prediction of corn and sugar prices using machine Learning, econometrics, and ensemble models. *MDPI.* 2021; 9(31):1-4.



4. Ouyang H, Wei X, Wu Q. Agricultural commodity futures prices prediction via long- and short-term time series network. *J Appl Econ.* 2019; 22(1): 468 - 483.
5. Shahhosseini M, Hu G, Archontoulis S. Forecasting corn yield with machine learning ensembles. *Front. Plant Sci.* 2020;31.
6. Rencher AC, Christensen WF. Multivariate regression, *Methods of Multivariate Analysis*: 3 ed: John Wiley & Sons; 2012.
7. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Royal Stat Soc.* 1996; 58(1): 267-288.
8. Hoerl AE, Kennard RW, Hoerl R. Practical use of ridge regression: A Challenge Met. *Journal of the Royal Statistical Society.* 1985; 34(2): 114-120.
9. Soonthornphisaj N. Machine learning. Bangkok: Kasetsart University; 2020.