ประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง สำหรับพยากรณ์ผลผลิตสับปะรดปัตตาเวีย
The Effectiveness of Machine Learning Forecasting of the Smooth Cayenne Pineapple.

ศิวาพร ศรีลาศักดิ์[1], ธัญญาธร ศรีวิเชียร[2] และอนุพงศ์ สุขประเสริฐ[2*]
คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม
Siwaphorn Srilasak[1], Tanyatorn Sriwichien[2] and Anupong Sukprasert[2*]
Mahasarakham Business School, Mahasarakham University

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องที่นำมาสร้างตัวแบบการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย โดยเก็บข้อมูล ตั้งแต่เดือนมกราคม พ.ศ. 2559 ถึง เดือนธันวาคม พ.ศ. 2563 ที่เก็บในรูปแบบรายเดือน รวมทั้งสิ้น 60 เดือน มาวิเคราะห์ตามกระบวนการวิเคราะห์ข้อมูล ด้วยเทคนิคการประมาณค่าข้อมูล ทั้งหมด 5 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น เทคนิคเพื่อนบ้านใกล้ที่สุด เทคนิคโครงข่ายประสาทเทียม เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคการเรียนรู้เชิงลึก ผลการวิจัยพบว่า เทคนิคที่มีความเหมาะสมมากที่สุดสำหรับนำมาสร้างตัวแบบพยากรณ์ผลผลิตสับปะรดปัตตาเวีย คือ วิธีเทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) ซึ่งให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เท่ากับ 23.982 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เท่ากับ 11.991 ค่าคลาดเคลื่อนสัมบูรณ์ (AE) เท่ากับ 14.575 และค่าความคลาดเคลื่อนยกกำลังสอง (SE) เท่ากับ 320.716 ผลการวิเคราะห์ข้อมูลนี้สามารถนำตัวแบบที่ได้ไปพัฒนาเป็นระบบสารสนเทศสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวียในอนาคต และนำไปต่อยอดในการพยากรณ์ผลผลิตทางการเกษตรชนิดอื่นๆ ต่อไป

คำสำคัญ: 1) การพยากรณ์ 2) สับปะรดปัตตาเวีย 3) เทคนิคการเรียนรู้ของเครื่อง

Abstract

The research aims to develop and compare the performance of machine learning techniques for building predictive models of the Smooth Cayenne Pineapple. Data were collected monthly from January 2016 to December 2020, totaling 60 months, and analyzed using the Regression Model technique along with four techniques. Including Linear Regression, k-Nearest Neighbors, Neural Network, Neural Network, Support Vector Machines and Deep Learning. The research findings indicate that the most suitable technique for constructing a predictive model of pineapple yield is k-Nearest Neighbors, which yields an average Mean Squared Error (MSE) of 23.982, Root Mean Square Error (RMSE) of 11.991, Absolute Error (AE) of 14.575, and Squared Error (SE) of 320.716. The findings of this data analysis can be utilized to develop a model for building an information system for forecasting the Smooth Cayenne Pineapple production in the future and can be extended to forecast the production of other agricultural commodities in the future.

Keywords: 1) Forecasting 2) Smooth Cayenne Pineapple 3) Machine learning

---

[1]นิสิตหลักสูตรบัญชีมหาบัณฑิต คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม
[2]อาจารย์ประจำคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม
[*]E-mail: anupong.s@acc.msu.ac.th

บทนำ

ปัจจุบันสับปะรดเป็นหนึ่งในพืชเศรษฐกิจของไทย เนื่องจากไทยเป็นผู้ส่งออกสับปะรดและผลิตภัณฑ์ที่ได้จากสับปะรดในลำดับต้นของโลก มูลค่าการส่งออกสูง ผลิตภัณฑ์ที่ส่งออกมาก คือ สับปะรดกระป๋องและน้ำสับปะรด พันธุ์สับปะรดที่นิยมปลูกและส่งเข้าโรงงานนิยมปลูกพันธุ์ ปัตตาเวีย ซึ่งมีชื่อเรียกหลายชื่อ เช่น สับปะรดศรีราชา พันธุ์ตาแดง พันธุ์ตาดำ กัลกัตตาหรือสับปะรดปราณบุรี เนื่องจากมีเนื้อแน่น รสหวานปานกลางจนถึงหวานจัด สามารถปลูกได้ทั่วไป การเก็บเกี่ยวควรเก็บเกี่ยวเมื่อตาสับปะรดเริ่มเปิด 2-3 ตา หรือผิวเปลือกเปลี่ยนเป็นสีเหลืองประมาณ 10 เปอร์เซ็นต์ ใช้มีดตัดก้านให้เหลือก้านยาวติดผลประมาณ 10 เซนติเมตร โดยไม่ต้องหักจุกของสับปะรดออก (ดนุสรณ์ ธนะปาละ และคณะ, 2559 : 111)

ประเทศไทยกำลังเผชิญปัญหาการส่งออกสับปะรด เนื่องจากเนื้อที่เก็บเกี่ยวผลผลิตสับปะรดทั้งประเทศคาดว่าลดลงจากปีที่ผ่านมา เป็นผลจากในปี 2565 เกษตรกรได้ลดเนื้อที่ปลูกสับปะรดใหม่จากต้นทุนการผลิต เช่น ปุ๋ยและสารเคมีที่ปรับตัวสูงขึ้น การขาดแคลนแรงงาน และการปรับตัวลดลงของราคาสับปะรดในปีที่ผ่านมา เกษตรกรจึงปรับเปลี่ยน ไปปลูกไม้ผลและไม้ยืนต้น เช่น ยางพารา ทุเรียน มะพร้าว และปรับเปลี่ยนไปปลูกมันสำปะหลัง โรงงานที่มีราคาอยู่ในเกณฑ์ดี ประกอบกับในปี 2565 บางพื้นที่ประสบภาวะแล้ง มีปริมาณฝนน้อยกว่าปี 2564 เกษตรกรบางรายจึงชะลอการบังคับออกดอกสับปะรดในปี 2565 ส่งผลให้เนื้อที่เก็บเกี่ยวลดลงจากปีที่ 2564 สำหรับผลผลิตต่อไร่คาดว่าลดลงจากปี 2564 เนื่องจากปุ๋ยเคมียังคงมีราคาสูง ประกอบกับ ราคาสับปะรดยังคงทรงตัว เกษตรกรขาดแรงจูงใจในการบำรุงรักษาจึงปรับลดปริมาณการใช้ปุ๋ยเคมีลง ทำให้ต้นสับปะรดไม่สมบูรณ์ ผลมีขนาดเล็ก จึงทำให้ภาพรวมผลผลิตต่อไร่ลดลง (สำนักงานเศรษฐกิจการเกษตร, 2566) เพื่อให้สับปะรดไทยยังคงเป็นที่ต้องการของตลาดโลก จึงจำเป็นต้องรักษาคุณภาพการส่งออกสับปะรด และมีการวางแผนการผลิตที่เหมาะสมกับฤดูกาลที่เปลี่ยนแปลงตลอดเวลา ความต้องการเหล่านี้สามารถบรรลุได้หากมีเทคโนโลยีสารสนเทศช่วยสนับสนุนข้อมูลในการตัดสินใจเพื่อการวางแผนปลูกสับปะรด ผลผลิตสามารถออกทันต่อความต้องการของตลาด ช่วยลดความเสี่ยงและความผิดพลาดของการปลูกสับปะรดได้ ทำให้สามารถเพิ่มปริมาณผลผลิตสับปะรดต่อพื้นที่เพาะปลูกสูงขึ้นได้ตามที่ต้องการ ซึ่งสอดคล้องกับงานวิจัยของ ดนุสรณ์ ธนะปาละ และคณะ (2559 : 111) ได้พยากรณ์ราคาสับปะรดที่ส่งเข้าโรงงานด้วยวิธีบอกซ์ – เจนกินส์ ซึ่งจากการศึกษาพบว่างานวิจัยดังกล่าวยังไม่ได้เปรียบเทียบเทคนิคการพยากรณ์หลายเทคนิคร่วมกันจากชุดข้อมูลเดียวกัน เพื่อให้การพัฒนาตัวแบบการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย มีประสิทธิภาพมากยิ่งขึ้น

ผู้วิจัยได้สนใจศึกษาการสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย โดยใช้เทคนิคการประมาณค่า 5 เทคนิคมาทำการสร้างตัวแบบพยากรณ์ที่ได้จากการศึกษางานวิจัยที่เกี่ยวข้องกับงานวิจัยในครั้งนี้ ได้แก่ ประกอบด้วย 1) เทคนิคการถดถอยเชิงเส้น (Linear Regression) 2) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) 3) เทคนิคโครงข่ายประสาทเทียม (Neural Network) 4) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และ 5) เทคนิคการเรียนรู้เชิงลึก (Deep Learning) เพื่อหาตัวแบบที่เหมาะสมสำหรับการนำไปใช้ในการสร้างตัวแบบพยากรณ์ผลผลิตสับปะรดปัตตาเวียที่อาจจะเกิดขึ้นในอนาคต และยังสามารถนำผลการวิจัยครั้งนี้ไปพัฒนาต่อยอดในการสร้างแบบจำลองสำหรับการพยากรณ์เกี่ยวกับผลผลิตทางการเกษตรชนิดอื่นๆ ต่อไป

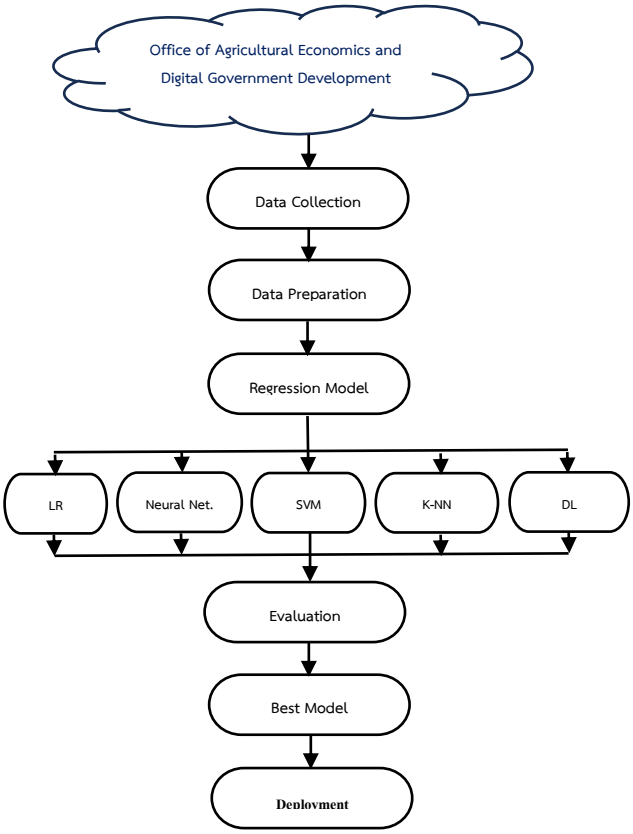
วัตถุประสงค์

1. เพื่อสร้างตัวแบบสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย
2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย

วิธีดำเนินการ

ผู้วิจัยได้ศึกษาเกี่ยวกับเทคนิคการเรียนรู้ของเครื่องสำหรับนำมาสร้างตัวแบบพยากรณ์ผลผลิตสับปะรดปัตตาเวีย โดยผ่านกระบวนการวิเคราะห์ข้อมูล (Data Analysis Process) โดยแบ่งข้อมูลออกเป็น 2 ส่วน เพื่อใช้

สำหรับการสร้างตัวแบบ (Training Set) และการทดสอบประสิทธิภาพของตัวแบบ (Testing Set) ด้วยเทคนิค Cross Validation โดยผู้วิจัยได้แบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน (10 – fold cross validation) (อนุพงศ์ สุขประเสริฐ, 2566 : 15) เทคนิคที่ใช้ในการสร้างตัวแบบพยากรณ์ ผู้วิจัยใช้เทคนิคในการประมาณค่าข้อมูล (Regression Model) จากนั้นทำการทดสอบประสิทธิภาพของตัวแบบ เพื่อหาตัวแบบที่มีความเหมาะสมสำหรับการนำไปใช้ในการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย ดังแสดงในภาพที่ 1



ภาพที่ 1 ขั้นตอนการดำเนินงาน

ในการสร้างตัวแบบสำหรับพยากรณ์ผลผลิตสับปะรดปัตตาเวีย และการเปรียบเทียบประสิทธิภาพสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย โดยใช้เทคนิคการเรียนรู้ของเครื่อง ผู้วิจัยได้ดำเนินการตามกระบวนการวิเคราะห์ข้อมูล (Data Analysis Process) การวิเคราะห์ข้อมูลแบ่งออกเป็น 6 ขั้นตอน (อนุพงศ์ สุขประเสริฐ, 2566) ดังนี้

1. การทำความเข้าใจปัญหา (Business Understanding)

ประเทศไทยกำลังเผชิญปัญหาการส่งออกสับปะรด เนื่องจากเนื้อที่เก็บเกี่ยวผลผลิตสับปะรดทั้งประเทศคาดว่าลดลงจากปีที่ผ่านมา เกษตรกรได้ลดเนื้อที่ปลูกสับปะรดใหม่จากต้นทุนการผลิต เช่น ปุ๋ยและสารเคมีที่ปรับตัวสูงขึ้น การขาดแคลนแรงงาน และการปรับตัวลดลงของราคาสับปะรดในปีที่ผ่านมา เกษตรกรจึงปรับเปลี่ยน ไปปลูกไม้ผลและไม้ยืนต้น เช่น ยางพารา ทุเรียน มะพร้าว และปรับเปลี่ยนไปปลูกมันสำปะหลัง โรงงานที่มีราคาอยู่ในเกณฑ์ดี ประกอบกับในปี 2565 บางพื้นที่ประสบภาวะแล้ง มีปริมาณฝนน้อยกว่าปี 2564 เกษตรกรบางรายจึงชะลอการบังคับออกดอกสับปะรดในปี 2565 ส่งผลให้เนื้อที่เก็บเกี่ยวลดลงจากปี 2564 สำหรับผลผลิตต่อไร่คาดว่าลดลงจากปี 2564 เนื่องจากปุ๋ยเคมียังคงมีราคาสูง ประกอบกับ ราคาสับปะรดยังคงทรงตัว เกษตรกรขาดแรงจูงใจในการบำรุงรักษา จึงปรับลดปริมาณการใช้ปุ๋ยเคมีลง ทำให้ต้นสับปะรดไม่สมบูรณ์ ผลมีขนาดเล็ก จึงทำให้ภาพรวมผลผลิตต่อไร่ลดลง (สำนักงานเศรษฐกิจการเกษตร, 2566 : ออนไลน์)

2. การทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลที่นำมาใช้ในการสร้างตัวแบบในการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย เป็นข้อมูลปริมาณการผลิตสับปะรด (ผลผลิต) จำนวน 5 ปีย้อนหลัง ตั้งแต่ปี พ.ศ.2559 ถึงปี พ.ศ.2563 ของส่วนปฏิบัติการสำนักงานเศรษฐกิจการเกษตร สำนักงานเศรษฐกิจการเกษตร (2566) และสำนักพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.) สำนักพัฒนารัฐบาลดิจิทัล (2566) ที่ถูกจัดเก็บไว้ในฐานข้อมูลขนาดใหญ่ รวมทั้งสิ้น 60 เดือน 27 แอตทริบิวต์ ที่ถูกจัดเก็บเป็นไฟล์ Excel สำหรับข้อมูลที่นำมาวิเคราะห์ในครั้งนี้ ประกอบด้วย ลำดับของข้อมูล (No.), ปีการเพาะปลูก (Crop_year), ปี (Year), ภูมิภาค/จังหวัด (Region/Province), สับปะรดปัตตาเวีย (Product), พื้นที่เพาะปลูกต่อไร (Sum_Plant_area), เนื้อที่เก็บเกี่ยวต่อไร (Sum_Harve_area), เนื้อที่ใส่ปุ๋ยต่อไร่ (Sum_Fertili_area), อัตราการใช้ปุ๋ยสำหรับเนื้อที่เพาะปลูกต่อกิโลกรัม (Rate_Fertili_Plantperkilo), อัตราการใช้ปุ๋ยสำหรับเนื้อที่ใส่ปุ๋ยต่อกิโกกรัม (Rate_ Fertiliarea_perkilo), ผลผลิตจากการเพาะปลูกต่อกิโลกรัม (Product_Plant_kilo), ผลผลิตจากการเก็บเกี่ยวต่อกิโลกรัม (Prod_Harve_kilo), ปริมาณปุ๋ยต่อตัน (Amount_Ferti_ton), ผลผลิตต่อตัน (Yield_per_ton), ผลผลิตต่อเดือนเขตภาคเหนือ (Yiels_per_Month_North), ผลผลิตต่อเดือนเขตภาคใต้ (Yiels_per_Month_South), ผลผลิตต่อเดือนเขตภาคตะวันออกเฉียงเหนือ (Yiels_per_Month_NORTHEAST), ผลผลิตต่อเดือนเขตภาคกลาง (Yiels_per_Month_CENTAL), อัตราส่วนของผลผลิตในเขตภาคเหนือ (Percent_Yield_NORTH), อัตราส่วนของผลผลิตในเขตภาคตะวันออกเฉียงเหนือ (Percent_Yield_NORTHEAST), อัตราส่วนของผลผลิตในเขตภาคกลาง (Percent_Yield_CENTRAL), อัตราส่วนผลผลิตในเขตภาคใต้ (Percent_Yield_SOUTH), ดัชนีราคาขายผลผลิต ณ ไร่นา (Agri_Price_Index), ดัชนีผลผลิต (Agri_Product_Index), ราคาขายต่อไร่เกษตรกร (Price_Former_rai), ดัชนีราคาขายผลผลิตปรับตามฤดูกาล (Agri_Price_Index_Season) และดัชนีผลผลิตปรับตามฤดูกาล (Agri_Product_Index_Season)

3. การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลเป็นขั้นตอนการทำความสะอาดข้อมูลให้อยู่ในรูปแบบที่เหมาะสมซึ่งต้องทำความสะอาดข้อมูลก่อน แสดงให้เห็นถึงความน่าเชื่อถือของข้อมูลก่อนการวิเคราะห์ โดยเริ่มการนำเข้าข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลขนาดใหญ่ ทำการคัดเลือกตัวแปรที่มีความสำคัญต่อการวิเคราะห์ข้อมูลโดยการศึกษาข้อมูลจากงานวิจัย และทำการแปลงข้อมูลให้เหมาะสมสำหรับเทคนิคที่ใช้ในการวิเคราะห์ โดยมีขั้นตอนการเตรียมข้อมูล (อนุพงศ์ สุขประเสริฐ, 2566) ดังนี้

3.1 การทำความสะอาดข้อมูล (Data Cleaning)

เนื่องจากข้อมูลดิบ (Raw data) ที่ได้รวบรวมมานั้น มีจำนวนแอทริบิวต์ ค่าสูญหาย (Missing Value) และสิ่งรบกวน (Noisy Data) เป็นจำนวนมาก หากนำไปวิเคราะห์อาจทำให้เกิดความผิดพลาด ไม่น่าเชื่อถือ ผู้วิจัยจึงได้ตัดตัวแปรที่มีค่าสูญหายออกไป 12 แอตทริบิวต์ ดังนั้นตัวแปรที่ใช้ในการศึกษาจะเหลือเพียง 15 แอตทริบิวต์ ดังแสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลที่ใช้ในการวิจัย

ลำดับ	แอทริบิวต์	คำอธิบาย
1	Crop_year (ID)	ปีการเพาะปลูก
2	Region	ภูมิภาค
3	Yield_per_Month_NORTH	ผลผลิตต่อเดือนเขตภาคเหนือ
4	Yield_per_Month_SOUTH	ผลผลิตต่อเดือนเขตภาคใต้
5	Yield_Per_Month_NORTHEAST	ผลผลิตต่อเดือนเขตภาคะวันออกเฉียงเหนือ
6	Yield_Per_Month_CENTAL	ผลผลิตต่อเดือนเขตภาคกลาง
7	Percent_Yield_NORTH	อัตราส่วนของผลผลิตในเขตภาคเหนือ
8	Percent_Yield_NORTHEAST	อัตราส่วนของผลผลิตในเขตภาคตะวันออกเฉียง เหนือ

ลำดับ	แอทริบิวต์	คำอธิบาย
9	Percent_Yield_CENTRAL	อัตราส่วนของผลผลิตในเขตภาคกลาง
10	Percent_Yield_SOUTH	อัตราส่วนของผลผลิตในเขตภาคใต้
11	Agri_Price_Index	ดัชนีราคาขายผลผลิต ณ ไร่นา
12	Agri_Product_Index	ดัชนีผลผลิต
13	Price_Former_rai	ราคาขายต่อไร่เกษตรกร
14	Agri_Price_Index_Season	ดัชนีราคาขายผลผลิตปรับตามฤดูกาล
15	Agri_Product_Index_Season	ดัชนีผลผลิตปรับตามฤดูกาล

3.2 การกำหนดหน้าที่ของตัวแปร (Set Role)

กำหนดให้หน้าที่ในแต่ละแอทริบิวต์เพื่อใช้ในการวิเคราะห์แต่ละแอทริบิวต์ โดยกำหนดให้ แอทริบิวต์ดัชนีผลผลิตปรับตามฤดูกาล ทำหหน้าที่เป็นตัวแปรตาม (Label) ที่ใช้สำหรับการพยากรณ์ และปีการเพาะปลูกมีหน้าที่เป็นรหัส (ID) สำหรับแอทริบิวต์ที่เหลือจะมีการกำหนดหน้าที่ให้เป็นตัวแปรอิสระ

4. การสร้างตัวแบบ (Modeling)

ขั้นตอนการสร้างตัวแบบเป็นการนำข้อมูลมาวิเคราะห์ตามอัลกอลิทึมด้วยเทคนิคการเรียนรู้ของเครื่องสำหรับการประมาณค่าข้อมูล (Regression Model) การวิจัยในครั้งนี้ ผู้วิจัยได้เลือกใช้เทคนิคการเรียนรู้ของเครื่องเพื่อใช้ในการประมาณค่าข้อมูล ทั้งหมด 5 เทคนิค ประกอบด้วย 1) **เทคนิคการถดถอยเชิงเส้น** (Linear Regression) 2) **เทคนิคเพื่อนบ้านใกล้ที่สุด** (k-Nearest Neighbors) 3) **เทคนิคโครงข่ายประสาทเทียม** (Neural Network) 4) **เทคนิคซัพพอร์ตเวกเตอร์** (Support Vector Machines) และ 5) **เทคนิคการเรียนรู้เชิงลึก** (Deep Learning) ด้วยโปรแกรม RapidMiner Studio version 10.1 (อนุพงศ์ สุขประเสริฐ, 2566 : 45) ในการสร้างตัวแบบพยากรณ์ในครั้งนี้

4.1 เทคนิคการถดถอยเชิงเส้น (Linear Regression) เป็นวิธีการทางสถิติที่ใช้ในการวิเคราะห์ข้อมูลที่มีตัวแปรต้น (Independent Variables) และตัวแปรตาม (Dependent Variable) เพื่อหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรสองตัวแปร หรือมากกว่าสองตัวแปร ซึ่งอาจมีความสัมพันธ์กันหรือแปรผกผันกันได้ โดยใช้สมการเส้นตรงเพื่อพยากรณ์ค่าของตัวแปรตามจากตัวแปรต้นที่กำหนดมา รูปแบบของสมการถดถอยเชิงเส้นเป็นรูปแบบพื้นฐานอย่างง่ายที่สุด โดยสร้างเป็นสมการเชิงเส้น มีรูปแบบสมการ (1) (อนุพงศ์ สุขประเสริฐ, 2566 : 212) ดังนี้

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \qquad (1)$$

โดยที่ Y คือ ตัวแปรตาม

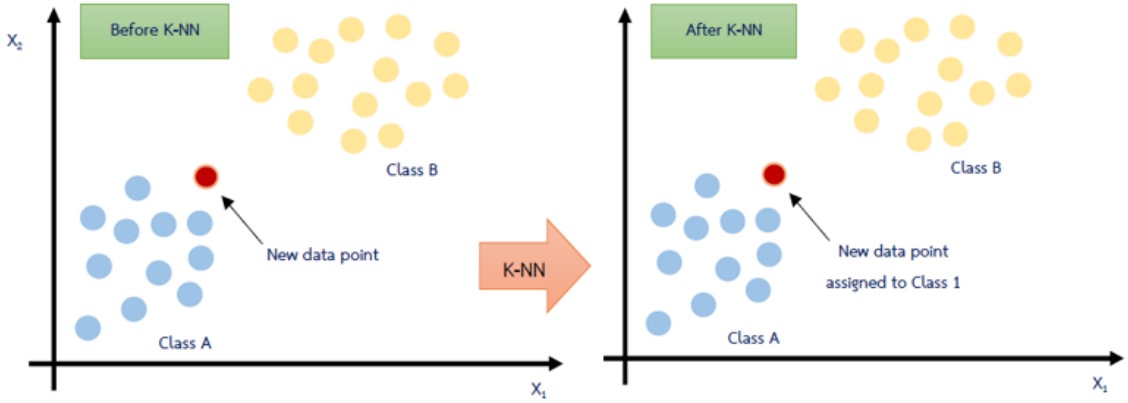
X คือ ตัวแปรอิสระ

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ คือ พารามิเตอร์

ε คือ ความคลาดเคลื่อน

4.2 เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) เป็นเทคนิคการประมาณค่าข้อมูลที่ใช้หลักการเปรียบเทียบความคล้ายคลึงกันระหว่างข้อมูลใหม่กับข้อมูลที่มีอยู่ โดยพิจารณาข้อมูลที่อยู่ใกล้ที่สุดจำนวน k ตัว ข้อมูลใหม่จะถูกจัดประเภทเดียวกับข้อมูลที่อยู่ใกล้มากที่สุด ขั้นตอนหลักของเทคนิคเพื่อนบ้านใกล้ที่สุด เริ่มต้นจากเลือกค่า k เป็นจำนวนของข้อมูลที่อยู่ใกล้ที่สุดที่จะใช้ในการตัดสินใจ คำนวณระยะห่างระหว่างข้อมูลใหม่กับข้อมูลที่มีอยู่ทั้งหมด เลือกข้อมูลที่อยู่ใกล้ที่สุดจำนวน k ตัว ข้อมูลใหม่จะถูกจัดประเภทเดียวกับข้อมูลที่อยู่ใกล้ที่สุดมากที่สุด เทคนิคเพื่อนบ้านใกล้ที่สุดเป็นเทคนิคการประมาณค่าข้อมูลแบบง่าย ๆ แต่มีประสิทธิภาพสูง

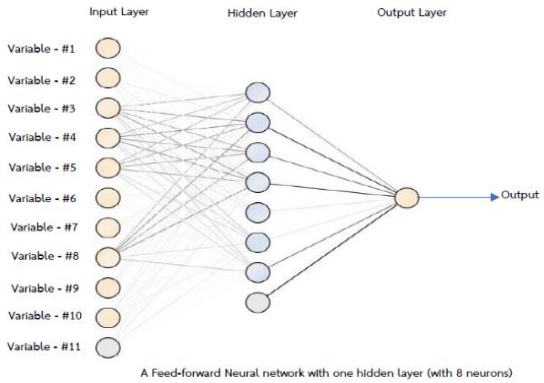
สามารถนำมาประยุกต์ใช้กับงานต่าง ๆ ได้อย่างหลากหลาย (อนุพงศ์ สุขประเสริฐ. 2566 : 234 - 242) ดังแสดงในภาพที่ 2



ภาพที่ 2 เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors)

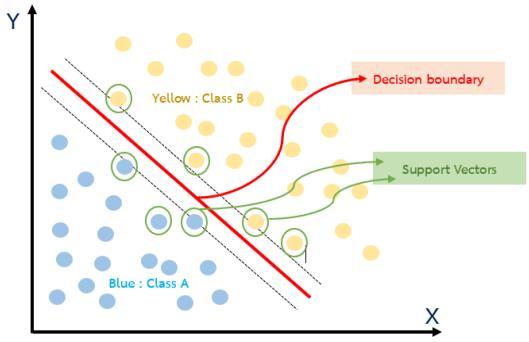
4.3 เทคนิคโครงข่ายประสาทเทียม (Neural Network) โมเดลคณิตศาสตร์ที่มีโครงสร้างและการทำงานคล้ายระบบประสาทเทียมในสมองของมนุษย์ สามารถใช้ในการเรียนรู้และการประมวลผลข้อมูลต่างๆ โดยมีการจำลองเซลล์ประสาทเทียม (Artificial Neurons) และการเชื่อมต่อระหว่างเซลล์ประสาทในรูปแบบของเครือข่าย ประกอบด้วยชั้นของเซลล์ประสาทที่เรียกว่าชั้น (Layer) ชั้นของเซลล์ประสาทแต่ละชั้นเชื่อมต่อกันด้วยน้ำหนัก (Weights) สามารถปรับเปลี่ยนในระหว่างการฝึก (Training) โดยทั่วไปการทำงานของ Artificial Neural Network (ANN) ถูกจำลองโดยเมื่อมีข้อมูลถูกนำเข้าจะคูณด้วยค่าน้ำหนัก แทนความสำคัญที่ให้กับข้อมูลนำเข้าในแต่ละตัว ผลรวมที่ได้ของค่าถ่วงน้ำหนักเกิดจากผลคูณของข้อมูลนำเข้า ค่าน้ำหนักจะถูกนำไปวิเคราะห์เพื่อตีความฟังก์ชันถ่ายโอน ทำหน้าที่เปรียบเสมือนกระบวนการทำงานของตัวเซลล์ประสาทได้เป็นผลลัพธ์ ฟังก์ชันถ่ายโอนที่สำคัญ คือ tan-sigmoid (tansig) log-sigmoid (logsig) และ pure-linear (purelin) นอกจากนี้โครงข่ายประสาทเทียมแบบหลายชั้น (Multi – layer Perceptron) ประกอบด้วย 3 ชั้น ได้แก่ ชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นข้อมูลออก แต่ละชั้นจะมีโหนดและเซลล์ประสาทหนึ่งเซลล์หรือมากกว่านั้น และข้อมูลจะส่งจากชั้นข้อมูลเข้าไปถึงชั้นข้อมูลออกโดยไม่ส่งย้อนกลับ ถึงแม้จะเป็นโหนดชั้นเดียวกันก็จะไม่มีการเชื่อมต่อกัน ซึ่งโครงข่ายประสาทเทียมหลายชั้นนี้จะเป็นโครงข่ายที่นิยมใช้ในการพยากรณ์

เนื่องจากจำแนกข้อมูลที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ (นรวัฒน์ เหลืองทอง และนันทชัย กานตานันทะ, 2559 : 374 - 375) ดังแสดงในภาพที่ 3



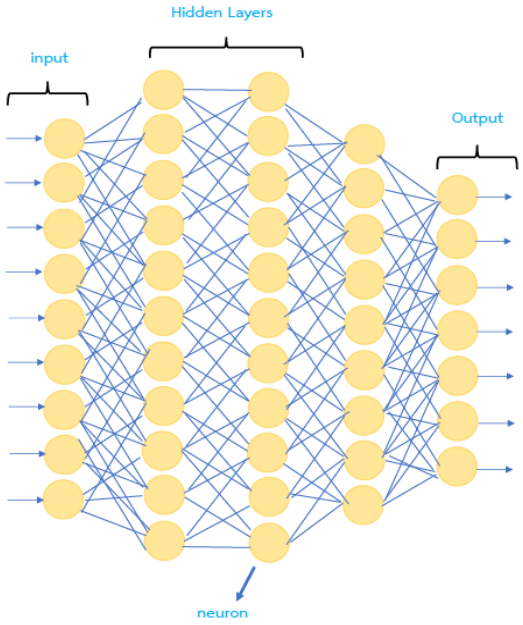
ภาพที่ 3 โครงข่ายประสาทเทียมที่มีการเชื่อมต่อกันผ่านโหนด

4.4 เทคนิควิธีซัพพอร์ตเวกเตอร์ (Support Vector Machines) เป็นเทคนิคการประมาณค่าข้อมูลที่มีความนิยมสูง สามารถนำมาประยุกต์ใช้กับงานต่าง ๆ ได้อย่างหลากหลายเนื้องจากสามารถจำแนกประเภทข้อมูล (classification) ที่ใช้หลักการหาเส้นแบ่ง (decision boundary) โดยจะแบ่งข้อมูลออกเป็นสองกลุ่มมีระยะห่างระหว่างข้อมูลทั้งสองกลุ่มมากที่สุด เทคนิคนี้หากข้อมูลที่อยู่ใกล้เส้นแบ่งมากที่สุดเรียกว่า เวกเตอร์สนับสนุน (support vector) ซึ่งเวกเตอร์สนับสนุนมีความสำคัญต่อประสิทธิภาพของโมเดล เนื่องจากโมเดลดังกล่าวจะพยายามปรับเส้นแบ่งให้ห่างจากเวกเตอร์สนับสนุนมากที่สุด โดยขั้นตอนหลักของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนเริ่มจากการกำหนดฟังก์ชั่นการสูญเสีย (loss function) กำหนดฟังก์ชั่นวัตถุประสงค์ (objective function) และหาค่าพารามิเตอร์ของโมเดล ซึ่งเทคนิคนี้สามารถจำแนกประเภทข้อมูลได้หลากหลายประเภท มีประสิทธิภาพสูง ทนต่อข้อมูลที่มี outliers แต่เทคนิคนี้ต้องใช้ข้อมูลตัวอย่างจำนวนมากในการสอน ใช้เวลาในการฝึกนาน อย่างเช่นการจำแนกประเภทภาพ ซึ่งต้องจำแนกประเภทวัตถุในภาพ จำแนกประเภทใบหน้า การจำแนกประเภทข้อความ เป็นการจำแนกประเภทอีเมลเป็นสแปมหรือไม่ใช่สแปม การจำแนกประเภทความคิดเห็นเป็นบวกหรือลบ และการประมาณค่าข้อมูลทางการเงิน โดยการจำแนกประเภทลูกค้าเป็นกลุ่มที่มีความเสี่ยงสูงหรือต่ำ (วิวัฒน์ กิตตินาราพร และคณะ, 2565 : 95 - 108) ดังแสดงในภาพที่ 4



ภาพที่ 4 เทคนิควิธีซัพพอร์ตเวกเตอร์ (Support Vector Machines)

4.5 เทคนิคการเรียนรู้เชิงลึก (Deep Learning) เทคนิคการประมาณค่าข้อมูลด้วยวิธีการเรียนรู้เชิงลึก (Deep Learning) เป็นกระบวนการทางปัญญาประดิษฐ์ที่ใช้โครงข่ายประสาทเทียมที่มีหลายชั้น (Deep Neural Networks) ในการเรียนรู้รูปแบบของข้อมูล มีการเรียนรู้จากข้อมูลขนาดใหญ่ และซับซ้อน โดยใช้หลายชั้นของโนด (nodes) ในแต่ละชั้น เช่น ชั้นนำเข้า (input layer) ชั้นซ่อน (hidden layers) และชั้นส่งออก (output layer) อย่างไรก็ตามโครงข่ายประสาทเทียมลึกยังสามารถจำแนก และพยากรณ์ข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ มีความสามารถในการเรียนรู้ลักษณะของข้อมูลโดยอัตโนมัติ โดยไม่ต้องกำหนดคุณลักษณะ (features) ข้อมูลล่วงหน้า (กฤติกา ทิพย์คำมี และคณะ, 2566 : 58 - 74 ; สุขสันต์ พรมบุญเรือง และคณะ, 2566 : 97 - 112) ดังแสดงในภาพที่ 5



ภาพที่ 5 เทคนิคการเรียนรู้เชิงลึก (Deep Learning)

5. การทดสอบประสิทธิภาพของตัวแบบ (Evaluation) หลังจากได้สร้างตัวแบบขึ้นมาแล้ว ขั้นตอนต่อไปต้องทำการวัดประสิทธิภาพของตัวแบบการประมาณค่าข้อมูลของตัวแบบที่สร้าง ก่อนที่จะนำตัวแบบไปใช้ในขั้นตอนของการพยากรณ์ สำหรับงานวิจัยนี้ ผู้วิจัยได้ใช้เกณฑ์ในการวัดค่าความแม่นยำของตัวแบบ มีทั้งหมด 4 วิธี ประกอบด้วย

5.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เป็นวิธีการวัดความคลาดเคลื่อนระหว่างค่าที่คำนวณได้และค่าจริงในการพยากรณ์หรือการประมาณค่าของตัวแปรต่าง ๆ โดยระยะห่างระหว่างค่าที่คำนวณได้และค่าจริงถูกยกกำลังสอง หลังจากนั้นนำมาหาค่าเฉลี่ย หากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) มีค่าน้อยแสดงถึงความแม่นยำของการพยากรณ์ที่ดีขึ้น และหากมีค่ามากขึ้นแสดงถึงการนำนายที่ไม่แม่นยำ สามารถหาได้จากสมการ (2) (วิรากานต์ กิตติบวรกุล และคณะ, 2565 ; กฤติกา ทิพย์คำมี และคณะ, 2566) ดังนี้

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2 \qquad (2)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใดๆ

5.2 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error : RMSE) เป็นการวัดความแตกต่างระหว่างค่าพยากรณ์ และค่าจริง หากค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) มีค่าน้อย แสดงว่าค่าพยากรณ์สามารถประมาณค่าได้ใกล้เคียงกับค่าจริง สามารถหาได้จากสมการ (3) (วิรากานต์ กิตติบวรกุล และคณะ, 2565 ; วรางคณา เรียนสุทธิ์, 2561 ; กฤติกา ทิพย์คำมี และคณะ, 2566) ดังนี้

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-\hat{Y}_t\right)^2} \quad (3)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใดๆ

$Y_t-\hat{Y}_t$ คือ ค่าของความแตกต่างระหว่างค่าจริงและค่าพยากรณ์

5.3 ค่าคลาดเคลื่อนสัมบูรณ์ (Absolute Error : AE) เป็นวิธีการวัดความคลาดเคลื่อนระหว่างค่าพยากรณ์และค่าจริงๆ โดยใช้ค่าสัมบูรณ์ของความแตกต่างระหว่างค่าทั้งสอง เพื่อหาค่าของความแตกต่างในทางบวกโดยไม่คำนึงถึงทิศทาง ซึ่งถ้าค่าคลาดเคลื่อนสัมบูรณ์มีค่าน้อย แสดงให้เห็นถึงการพยากรณ์ที่มีความแม่นยำในการประมาณค่าของข้อมูล ค่าน้อยแสดงถึงการประมาณค่าที่ใกล้เคียงกับค่าจริงมากยิ่งขึ้น แต่ถ้ามีค่าสูงแสดงว่าการพยากรณ์มีความคลาดเคลื่อนมาก โดยต้องระมัดระวังการใช้งานร่วมกับข้อมูลที่มีค่าสูง ค่าคลาดเคลื่อนสัมบูรณ์มักถูกใช้ในการประเมินประสิทธิภาพของการพยากรณ์ในสถานการณ์หลายรูปแบบที่ต่างกัน สามารถหาได้จากสมการ (4) (กฤติกา ทิพย์คำมีและคณะ, 2566 : 69) ดังนี้

$$AE=\left|Y_t-\hat{Y}_t\right| \quad (4)$$

โดยที่ Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใดๆ

$Y_t-\hat{Y}_t$ คือ ค่าของความแตกต่างระหว่างค่าจริงและค่าพยากรณ์

5.4 ค่าความคลาดเคลื่อนยกกำลังสอง (Square Error : SE) เป็นวิธีการวัดความคลาดเคลื่อนระหว่างค่าที่พยากรณ์และค่าจริงๆ โดยการยกกำลังสองของความแตกต่างระหว่างค่าทั้งสอง ซึ่งใช้ในการหาความคลาดเคลื่อนโดยให้ความสำคัญกับค่าที่มากขึ้น หากค่าความคลาดเคลื่อนยกกำลังสองมีค่าน้อยแสดงถึงความแม่นยำของการพยากรณ์ที่ดีขึ้น ค่าที่สูงขึ้นแสดงถึงความไม่แม่นยำของการพยากรณ์ในงานนั้นๆ ดังนั้น การประเมินค่าความคลาดเคลื่อนยกกำลังสอง (SE) เป็นอีกวิธีหนึ่งที่สำคัญในการประเมินประสิทธิภาพของการพยากรณ์ในหลายๆ สถานการณ์ที่ต่างกัน สามารถหาได้จากสมการ (5) (กฤติกา ทิพย์คำมีและคณะ. 2566 : 70) ดังนี้

$$SE=\left(\left|Y_t-\hat{Y}_t\right|\right)^2 \quad (5)$$

โดยที่ Y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใดๆ

$Y_t-\hat{Y}_t$ คือ ค่าของความแตกต่างระหว่างค่าจริงและค่าพยากรณ์

6. การนำไปใช้ (Deployment)

ผลลัพธ์จากการวิเคราะห์ข้อมูลครั้งนี้จะได้ตัวแบบที่มีความเหมาะสมสำหรับการสร้างตัวแบบพยากรณ์ผลผลิตสับปะรดปัตตาเวียในประเทศไทย เพื่อสนับสนุนให้หน่วยงานที่เกี่ยวข้องสามารถนำตัวแบบที่ได้ไปใช้ในการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย และวางแผนนโยบายต่างๆ เกี่ยวกับผลผลิตของสับปะรดปัตตาเวียในอนาคตให้กับเกษตรกร นอกจากนี้ยังสามารถนำผลการวิจัยในครั้งนี้ไปพัฒนาโดยการสร้างระบบสารสนเทศ

สำหรับการพยากรณ์เกี่ยวกับสับปะรดในแต่ละชนิดหรือการพยากรณ์ผลผลิตสินค้าเกษตรชนิดอื่นๆ โดยอาจเพิ่มปัจจัยในส่วนอื่น ที่ยังไม่ได้นำมาศึกษาในครั้งนี้ เพื่อให้มีความแม่นยำสำหรับการพยากรณ์มากยิ่งขึ้น

ผลการศึกษา

ผู้วิจัยทำการเปรียบเทียบค่าที่ทดสอบประสิทธิภาพของตัวแบบพยากรณ์ผลผลิตสับปะรด ด้วยเทคนิคการเรียนรู้ของเครื่อง 5 เทคนิค ได้แก่ (1) เทคนิคการถดถอยเชิงเส้น (Linear Regression) (2) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) (3) เทคนิคโครงข่ายประสาทเทียม (Neural Network) (4) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และ (5) เทคนิคการเรียนรู้เชิงลึก (Deep Learning) และทำการทดสอบประสิทธิภาพของประมาณค่าด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error) ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ค่าคลาดเคลื่อนสัมบูรณ์ (Absolute Error) และค่าความคลาดเคลื่อนยกกำลังสอง (Square Error) ผลการวิเคราะห์ข้อมูลแสดงดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของตัวแบบการประมาณค่าข้อมูล

Regression Model	Regression Performance			
	MSE	RMSE	AE	SE
Linear Regression	31.180	15.590	12.846	265.273
k-Nearest Neighbors*	23.982	11.991	14.575	320.716
Neural Network	29.752	14.876	11.432	231.725
Support Vector machines	31.902	15.951	12.988	301.608
Deep Learning	25.176	12.588	9.963	181.306

* คือ เทคนิคที่มีความเหมาะสมสำหรับนำมาสร้างตัวแบบการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย

จากตารางที่ 2 พบว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างตัวแบบพยากรณ์ผลผลิตสับปะรด คือ เทคนิคการประมาณค่าข้อมูลด้วยวิธีเทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) ซึ่งให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เท่ากับ 23.982 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เท่ากับ 11.991 ค่าคลาดเคลื่อนสัมบูรณ์ (AE) เท่ากับ 14.575 และค่าความคลาดเคลื่อนยกกำลังสอง (SE) เท่ากับ 320.716 ผู้วิจัยได้ทำการเปรียบเทียบข้อมูลผลผลิตระหว่างค่าที่ได้จากข้อมูลจริงของผลผลิตสับปะรด และค่าพยากรณ์ของผลผลิตสับปะรดด้วยเทคนิคการประมาณค่าข้อมูลด้วยวิธีเทคนิคเพื่อนบ้านใกล้ที่สุด มีค่าใกล้เคียงกันมาก ซึ่งแสดงให้เห็นว่าเทคนิคที่มีประสิทธิภาพในการนำไปใช้พยากรณ์ผลผลิตสับปะรดปัตตาเวีย คือ เทคนิคการประมาณค่าข้อมูลด้วยวิธีเทคนิคเพื่อนบ้านใกล้ที่สุดที่ให้ค่าประสิทธิภาพแม่นยำที่สุด

สรุปและอภิปรายผล

จากวัตถุประสงค์ของการวิจัยการสร้างตัวแบบและเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ผลผลิตสับปะรดปัตตาเวีย ด้วยเทคนิคการเรียนรู้ของเครื่อง จากผลการวิจัยสามารถ สรุปได้ว่าเทคนิคที่มีความเหมาะสมที่สุดสำหรับการสร้างแบบพยากรณ์ผลผลิตสับปะรดปัตตาเวีย คือ เทคนิคการประมาณค่าข้อมูลด้วยวิธีเทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) โดยให้ค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่นๆ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เท่ากับ เท่ากับ 23.982 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เท่ากับ 11.991 ค่าคลาดเคลื่อนสัมบูรณ์ (AE) เท่ากับ 14.575 และค่าความคลาดเคลื่อนยกกำลังสอง (SE) เท่ากับ 320.716 เป็นตัวแปรปัจจัยที่มีต่อผลผลิตสับปะรดปัตตาเวียที่ใช้ในการสร้างแบบพยากรณ์ แต่ไม่สอดคล้องกับดนุสรณ์ ธนะปาละ ธันวา เจริญศิริ และชนาธิป โสภณพิมล (2559 : 110 – 118) ที่ได้ศึกษาในเรื่อง

การพยากรณ์ราคาสับปะรดที่ส่งเข้าโรงงานด้วยวีธี บอกซ์ – เจนกินส์ ซึ่งเป็นวิจัยที่พยากรณ์สินค้าเกษตรชนิดเดียวกัน แต่วิธีการพยากรณ์ที่แตกต่างกัน

ข้อเสนอแนะ

จากการศึกษาในครั้งนี้ผู้วิจัยได้เลือกเทคนิคในการพยากรณ์ 5 เทคนิค มาเปรียบเทียบทำการพยากรณ์ ซึ่งอาจจะไม่ได้เทคนิคที่ดีที่สุด หากมีการทำวิจัยในครั้งถัดไปควรเลือกใช้เทคนิคอื่นๆ มาใช้ในการเปรียบเทียบประสิทธิภาพความแม่นยำ เพื่อให้ได้ตัวแบบพยากรณ์ที่ดีที่สุด และยังสามารถหาปัจจัยอื่นๆที่ส่งผลต่อผลผลิตสับปะรดปัตตาเวียและสับปะรดชนิดอื่น รวมถึงสามารถนำเทคนิคการพยากรณ์ต่างๆ มาใช้พยากรณ์สินค้าเกษตรชนิดอื่นๆ นอกเหนือจากที่ไม่ได้นำมาศึกษาในครั้งนี้มาทำการวิเคราะห์ในครั้งถัดไป

บรรณานุกรม

สำนักพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.). (7 พฤศจิกายน 2566). *ข้อมูลปริมาณผลผลิตการเกษตร.* https://www.data.go.th

สำนักงานเศรษฐกิจการเกษตร. (6 พฤศจิกายน 2566). *ข้อมูลการพยากรณ์ผลผลิตเกษตร.* https://www.oae.go.th/view/

กฤติกา ทิพย์คำมี, อนุพงศ์ สุขประเสริฐ, สุพัตรา กอผจญ และณัฐกานต์ ชุติมารังสรรค์. (2566). ประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ. *วารสารวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม มหาวิทยาลัยกาฬสินธุ์ ปีที่ 2 ฉบับที่ 1 (มกราคม – มิถุนายน 2566)*, 58 – 74.

ดนุสรณ์ ธนะปาละ, ธันวา เจริญศิริ และชนาธิป โสภณพิมล. (2559). พยากรณ์ราคาสับปะรดที่ส่งเข้าโรงงานด้วยวิธีบอกซ์-เจนกินส์. *วารสารวิทยาศาสตร์บูรพา*, 110 - 118.

สุขสันต์ พรมบุญเรือง, ธรรมรัตน์ บุญรอด, พีรญา ระดาสัย และศิรดา สุภาพันธ์. (2566). การคัดแยกพุทราจากภาพถ่ายด้วยเทคนิคการเรียนรู้เชิงลึก. *Journal of Science Ladkrabang 32(2)*, 97 - 112.

นรวัฒน์ เหลืองทอง และนันทชัย กานตานันทะ. 2559. การเลือกตัวแบบพยากรณ์ผลผลิตการเกษตรที่เหมาะสม. *Thai Science and Technology Journal***,** 370 - 381.

วรางคณา เรียนสุทธิ์. (2561). การพยากรณ์ราคาสับปะรดโรงงานด้วยวิธีบอกซ์ – เจนกินส์. *Thai Journal of Science and Technology 8(2)*, 105 - 115.

วิรากานต์ กิตติบวรกุล, ศรายุทธ นนท์ศิริ และพิชิตชัย คำอินทร์. (2565). การพยากรณ์ความเสียหายของชิ้นส่วนเครื่องยนต์อากาศยานด้วยเทคนิคการเรียนรู้ของเครื่อง. *APHEIT Journal (SCIENCE and Technology) 11(1)*, 1 - 14.

วิวัฒน์ กิตตินาราพร, เสรีย์ ตู้ประกาย วรานนท์, คงสง สิรวัลภ์, เรืองช่วย ตู้ประกาย, เสถียร เจริญเหรียญ และลาภยศ ประสิทธิโศภิน. (2565). การประยุกต์ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และโครงข่ายประสาทเทียมพยากรณ์ปริมาณงานก่อสร้างในประเทศไทย. *Engineering Journal of Research and Development 34(1)*, 95 - 108.

อนุพงศ์ สุขประเสริฐ. (2566). *การทำเหมืองข้อมูลด้วยโปรแกรม Rapidminer Studio* (พิมพ์ครั้งที่ 5). คณะบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม