# ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องสำหรับการพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทย

เยาวลักษณ์ เกิดปั้น[1*], สุวรรณี อัศวกุลชัย[2]

[1]คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏกาญจนบุรี กาญจนบุรี
[2]คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย กรุงเทพมหานคร
[*]Corresponding author email: yaowalak.koetpan@gmail.com

ได้รับบทความ: 12 ตุลาคม 2564
ได้รับบทความแก้ไข: 20 มิถุนายน 2565
ยอมรับตีพิมพ์: 22 มิถุนายน 2565

**บทคัดย่อ**

บทความนี้มีวัตถุประสงค์ เพื่อประยุกต์การเรียนรู้ของเครื่องในการพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทย ในการศึกษาครั้งนี้ใช้กระบวนการวิเคราะห์ข้อมูลมาตรฐานเรียกว่า Cross-industry standard process (CRISP) ประกอบด้วย 6 ขั้นตอน ได้แก่ เข้าใจปัญหาของธุรกิจ เข้าใจข้อมูล เตรียมข้อมูล พัฒนาแบบจำลอง การประเมิน และการนำไปใช้จริง ข้อมูลที่ใช้รวบรวมข้อมูลพื้นที่เพาะปลูก จำนวนครัวเรือน และผลผลิตกล้วยไม้จาก www.oae.go.th และ www.ditp.go.th ในปี 2559 - 2563 จำนวน 20 จังหวัด การเตรียมและคัดเลือกข้อมูลให้สมบูรณ์เหมาะสำหรับการทำเหมืองข้อมูลด้วยกระบวนการ Extract transform load (ETL) จากนั้นทำการโอนย้ายข้อมูล การลดขนาดของข้อมูล และการทำความสะอาดข้อมูล ใช้โปรแกรม Knime เป็นเครื่องมือ วิเคราะห์หาความสัมพันธ์ และใช้อัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้ด้วย 3 แบบ ได้แก่ Simple regression tree, Gradient boosted trees และ Random forest การทดสอบประสิทธิภาพอัลกอริทึมด้วยการหา $R^2$ เพื่อหาอัลกอริทึมที่เหมาะสมที่สุดและวัดผล ซึ่งผลการศึกษา พบว่า การพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทยจากอัลกอริทึม Gradient boosted tree มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 97.50 เมื่อเทียบกับอัลกอริทึม Simple regression tree และ Random forest ถูกต้องร้อยละ 96.40 และ 92.80 ตามลำดับ สามารถนำมาแสดงผลข้อมูลด้วยภาพโดยใช้ Power business intelligence (Power BI) ซึ่งเป็นประโยชน์ในการวางแผนการผลิต เพื่อเพิ่มยอดขายได้เป็นอย่างดี

**คำสำคัญ:** การเรียนรู้ของเครื่อง / กล้วยไม้ / การถดถอยอย่างง่าย / เกรเดียนท์บูตทรีส์ / แรนดอมฟอเรสท์

# The Results of Comparison of the Efficiency of Machine Learning Algorithms for Predicting Orchid Products in Thailand

Yaowalak Koetpan[1*], Suwannee Adsavakulchai[2]

[1]Faculty of Industrial Technology, Kanchanaburi Rajabhat University, Kanchanaburi
[2]School of Engineering, University of the Thai Chamber of Commerce, Bangkok
[*]Corresponding author email: yaowalak.koetpan@gmail.com

Received: 12 October 2021
Revised: 20 June 2022
Accepted: 22 June 2022

**Abstract**

This article aims to apply machine learning for predicting the orchid product in Thailand. The technique in this study was the cross-industry standard process (CRISP). There were 6 processes that were business understanding, data understanding, data preparation, model development, evaluation, and deployment. The data of planted were household and product's orchid collected from www.oae.go.th and www.ditp.go.th during 2016 till September 2020 that consisted of 20 provinces. To do the data preparation, feature selection for data mining i.e., data transfer, data reduction, data cleaning using Knime a tool. There were 3 algorithms for forecasting orchid product that were simple regression trees, gradient boosted tree and random forest. Algorithm performance testing by using the $R^2$ for most suitable. The results could demonstrate that the most accuracy algorithm was gradient boosted tree with accuracy 97.50% comparing with simple regression and random forest with accuracy 96.40% and 92.80%, respectively. To do the data visualization using power business intelligence (Power BI) that was very beneficial for planning and leads to higher sales.

**Keywords:** Machine learning / Orchid / Simple regression tree / Gradient boosted trees / Random forest

**บทนำ**

กล้วยไม้ไทยจัดเป็นกล้วยไม้เขตร้อน (Tropical orchid) ได้แก่ กล้วยไม้สกุลหวายมอคคารา และแวนดา ช่วงผลผลิตสูง คือ เดือนมิถุนายน - ตุลาคม แหล่งผลิต (5 อันดับแรก) นครปฐม สมุทรสาคร กรุงเทพมหานคร ราชบุรี นนทบุรี ผลผลิตกล้วยไม้ไทยเป็นผลผลิตที่ผลิตเพื่อส่งออกประมาณ ร้อยละ 49.00 ส่วนอีกร้อยละ 49.00 เป็นการผลิตเพื่อใช้ในประเทศ (เป็นกล้วยไม้ที่มีคุณภาพต่ำกว่ามาตรฐานการส่งออก) มีจำนวนผู้ส่งออกกล้วยไม้ไทยประมาณ 115 ราย และเกษตรกรผู้ปลูกกล้วยไม้ประมาณ 3,000 ราย [1]

จากสถานการณ์แพร่ระบาดของเชื้อไวรัสโควิด 19 ที่ผ่านมาตั้งแต่ช่วงเดือนมีนาคม 2563 ได้ส่งผลกระทบต่อการส่งออกสินค้าเกษตรอย่างต่อเนื่อง สำหรับการส่งออกกล้วยไม้ ประเทศไทยเป็นผู้ส่งออกกล้วยไม้ไทยเป็นอันดับ 1 ของโลก แต่สถานการณ์การแพร่ระบาดของเชื้อไวรัสโควิด 19 ที่เกิดขึ้นจนมาถึงการระบาดระลอกที่ 2 ได้ส่งผลกระทบต่อโลจิสติกส์ ในการขนส่งสินค้าเข้าไปยังประเทศต่าง ๆ ทำให้ผู้นำเข้าหลัก ได้แก่ จีน อิตาลี สหรัฐอเมริกา ออสเตรเลีย และญี่ปุ่นระงับการนำเข้าทำให้การส่งออกชะลอตัวและหยุดชะงักลง โดยช่วงเดือนมกราคม - มิถุนายน 2563 มีปริมาณการส่งออกดอกกล้วยไม้ 7,455 ตัน มูลค่า 670.71 ล้านบาท ซึ่งปริมาณการส่งออกและมูลค่าการส่งออก ลดลงจากปี 2562 ในช่วงเวลาเดียวกัน ร้อยละ 34.00 และ 40.00 ตามลำดับ ราคาที่เกษตรกรขายได้ปรับตัวลดลงเป็น 1.43 บาทต่อช่อ ราคาลดลงมากกว่าร้อยละ 50.00 [2]

การเรียนรู้ของเครื่อง (Machine learning) เป็นการวิเคราะห์ข้อมูลรูปแบบหนึ่ง มีความสำคัญอย่างมากในการนำมาใช้พัฒนา เรียนรู้ และมีปฏิสัมพันธ์กับชุดข้อมูลต่าง ๆ และปรับปรุงธุรกิจให้เหมาะสมผ่านการประมวลผลข้อมูลต่าง ๆ ให้มีประสิทธิภาพ อีกทั้งยังช่วยเพิ่มความรวดเร็วในการทำงาน ลดภาระค่าใช้จ่ายและช่วยให้สามารถปรับปรุงคุณภาพของชิ้นงานต่าง ๆ ให้ดียิ่งกว่าเดิม และนำไปสู่การตัดสินใจได้ [3]

Random forest เป็นการพัฒนาแบบจำลอง จาก Decision tree หลาย ๆ โมเดลย่อย (ตั้งแต่ 10 โมเดล ถึงมากกว่า 1,000 โมเดล) โดยแต่ละโมเดลจะได้รับชุดข้อมูลที่แตกต่างกัน เนื่องจากเป็นชุดข้อมูลย่อยของชุดข้อมูลทั้งหมด การพยากรณ์กรณี Classification ก็ให้แต่ละ Decision tree ทำการพยากรณ์แล้วคำนวณผลการพยากรณ์ด้วยการ Vote output ที่ถูกเลือกโดย Decision tree มากที่สุด หรือกรณี Regression เป็นการหาค่า Mean จาก Output ของแต่ละ Decision tree ทั้งนี้ Decision tree แต่ละโมเดลใน Random forest เมื่อนำเอาแต่ละ Decision tree มาทำการพยากรณ์ร่วมกันก็จะได้โมเดลรวมที่มีความแม่นยำมากกว่า Decision tree ที่ทำการพยากรณ์แบบเดี่ยว ๆ [4]

Gradient boosting เป็นวิธีการในการ Optimize ใช้นำ Classifier หลายตัวมาทำงาน เมื่อเรียนรู้เรียบร้อยแล้ว Classifier ทุกตัวจะพยากรณ์ร่วมกัน โดยเรียนรู้จากค่าความคลาดเคลื่อนสะสมที่เกิดจากการทำนายของ Instance ก่อนหน้า ส่วนการวิเคราะห์การถดถอย ศึกษาความสัมพันธ์ระหว่างตัวแปร ตั้งแต่ 2 ตัวแปรขึ้นไป มีตัวแปรอิสระ (Independent variable) และตัวแปรตาม (Dependent variable) นำมาสร้างเป็นแบบจำลองการถดถอยเชิงเส้นอย่างง่าย (Simple linear regression model) [5]

การใช้ Gradient boosting decision tree ทำนายระยะเวลาในการเดินทาง เปรียบเทียบประสิทธิภาพการทำงานกับ Backpropagation neural network และ Support vector machine (SVM) จากการวัดประสิทธิภาพแบบจำลองด้วย การหาค่าเฉลี่ยร้อยละความคลาดเคลื่อนสมบูรณ์ (MAPE) พบว่า แบบจำลองที่สร้างจาก Gradient boosting decision tree มีค่า MAPE น้อยกว่าแบบจำลองที่สร้างจาก BP Neural network และ SVM แสดงให้เห็นว่าแบบจำลองที่สร้างจาก Gradient boosting decision tree สามารถทำนายระยะเวลาการเดินทางได้ใกล้เคียงมากที่สุด [6]

ดังนั้นในการศึกษาครั้งนี้มีวัตถุประสงค์ในการประยุกต์การเรียนรู้ของเครื่องเพื่อการพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทยด้วยอัลกอริทึม Simple regression tree, Gradient boosted trees และ Random forest เนื่องจากอัลกอริทึมดังกล่าวมีความเหมาะสมกับลักษณะข้อมูลผลผลิตกล้วยไม้ในประเทศไทย

**วัสดุและวิธีการ**

การเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องสำหรับการพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทย เทคนิคการเรียนรู้ของเครื่อง มีกระบวนการวิเคราะห์ข้อมูล 6 ขั้นตอน ได้แก่

1) การทำความเข้าใจปัญหาของธุรกิจ (Business understanding)

ขั้นตอนนี้เป็นการทำความเข้าใจถึงพื้นที่ ปริมาณผลผลิต จำนวนครัวเรือน และจำนวนผู้ส่งออกกล้วยไม้ของประเทศไทย ตลอดจนจุดแข็ง จุดอ่อน โอกาส ภัยคุกคาม และปัญหาต่าง ๆ ในอุตสาหกรรมกล้วยไม้ของประเทศไทย โอกาสในการส่งเสริม และสนับสนุนการเพาะปลูก และปริมาณผลผลิตกล้วยไม้ของประเทศไทย ให้อยู่ในรูปแบบที่เหมาะสมต่อการนำมาวิเคราะห์ข้อมูล

2) การทำความเข้าใจข้อมูล (Data understanding)

ข้อมูลจาก www.oae.go.th และ www.ditp.go.th ศึกษาโครงสร้างของฐานข้อมูล ประกอบด้วย ขนาดพื้นที่ ปริมาณผลผลิต จำนวนครัวเรือน และจำนวนผู้ส่งออกกล้วยไม้

ของประเทศไทย หลังจากนั้นดำเนินการตรวจสอบข้อมูล ความถูกต้องของข้อมูล มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

3) การเตรียมข้อมูล (Data preparation)

จากข้อ 2 หลังจากที่ได้ศึกษาและรวบรวมข้อมูลในขั้นตอนนี้จะต้องมีการจัดเตรียมข้อมูลและคัดเลือกข้อมูลที่จะนำมาใช้งานต้องมีการจัดเตรียมข้อมูลให้เป็นไปตามมาตรฐานเดียวกัน และคัดเลือกข้อมูลที่มีความถูกต้อง สมบูรณ์มากที่สุด เพื่อให้ได้ผลการทดสอบที่มีประสิทธิภาพสูงสุด ขั้นตอนนี้ใช้เวลานานที่สุด เนื่องจากแบบจำลองที่ได้จากการทำเรียนรู้ของเครื่องจะให้ผลลัพธ์ที่ถูกต้อง ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ กล่าวคือ ถ้าข้อมูลที่ใช้นั้นไม่ถูกต้อง มีข้อผิดพลาด ย่อมส่งผลให้ผลลัพธ์คลาดเคลื่อนได้ สำหรับการเตรียมข้อมูลแบ่งเป็น 3 ขั้นตอนย่อยดังนี้

3.1) การคัดเลือกข้อมูล (Data selection) กำหนดเป้าหมายที่ทำการวิเคราะห์พื้นที่ จำนวนครัวเรือน ปริมาณผลผลิต และจำนวนผู้ส่งออกกล้วยไม้ของประเทศไทย เลือกใช้เฉพาะข้อมูลที่เกี่ยวข้องกับสิ่งที่จะวิเคราะห์ การกำหนดคุณลักษณะ (Attribute) และค่าที่เป็นไปได้ของคุณลักษณะ ศึกษาโครงสร้างของฐานข้อมูลและการจัดเตรียมข้อมูลคุณลักษณะที่คาดว่าจะนำมาใช้งาน เพื่อใช้ในการหาความรู้ และได้มีการกำหนดค่าที่เป็นไปได้ของคุณลักษณะ

3.2) การกลั่นกรองข้อมูล (Data cleaning) ในการศึกษาครั้งนี้ พบข้อมูลที่ไม่ถูกต้อง เนื่องมาจากปัญหาในระหว่างการจัดเก็บข้อมูล เช่น กรอกข้อมูลไม่ครบถ้วน กรอกข้อมูลซ้ำซ้อน การกลั่นกรองข้อมูล การทำความสะอาดข้อมูล (Data cleaning) สามารถลดขนาดข้อมูล (Data reduction)

3.3) การแปลงรูปข้อมูล (Data transformation) ขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ตามอัลกอริทึมที่เลือกใช้

4) แบบจำลอง (Modeling)

ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่อง ได้แก่ แบบจำลองที่ใช้ในการทำนายในการศึกษาครั้งนี้ เลือกใช้การพยากรณ์ปริมาณผลผลิตกล้วยไม้ของประเทศไทยที่ใช้ ได้แก่ Simple regression tree, Gradient boosted trees และ Random forest ทั้งนี้เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด เมื่อทำขั้นตอนนี้แล้ว อาจมีการย้อนกลับไปที่ขั้นตอนการเตรียมข้อมูล (Data preparation) เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย นอกจากนี้ยังมีการประเมินแบบจำลอง วิเคราะห์ข้อมูลที่ได้ในรูปแบบความถูกต้องของแบบจำลอง เพื่อเป็นตัวบ่งชี้ความน่าเชื่อถือของแบบจำลองที่ได้

5) การประเมินผล (Evaluation)

ขั้นตอนนี้เป็นการประเมินประสิทธิภาพของผลลัพธ์จากแบบจำลอง โดยนำผลลัพธ์ที่ได้มาวิเคราะห์ความถูกต้องจากข้อมูลที่นำมาทดสอบ ต้องอาศัยทักษะในการวิเคราะห์ข้อมูลเพื่อช่วยให้การวิเคราะห์ทำได้สะดวกและรวดเร็วขึ้น จึงมีการใช้เครื่องมือทางด้านกราฟฟิก เช่น การแสดงผลการวิเคราะห์ด้วยกราฟ รายงานรูปแบบต่าง ๆ เป็นต้น

6) การถ่ายทอดเพื่อนำไปปฏิบัติ (Deployment)

ดำเนินการเพื่อตอบสนองข้อกำหนดในเกณฑ์การถ่ายทอด เพื่อนำไปปฏิบัติเป็นแนวทางไปใช้กับการวิเคราะห์พื้นที่ จำนวนครัวเรือน ปริมาณผลผลิต และจำนวนผู้ส่งออกกล้วยไม้ของประเทศไทย และสามารถระบุโอกาสในการส่งเสริมและสนับสนุนการเพาะปลูกและปริมาณผลผลิตกล้วยไม้ของประเทศไทย

**ผลการศึกษา**

จากการประยุกต์การเรียนรู้ของเครื่องในการพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทย โดยใช้กระบวนการวิเคราะห์ข้อมูลมาตรฐาน Cross-industry standard process (CRISP) 6 ขั้นตอนนั้น ได้ผลการศึกษาโดยมีรายละเอียดดังต่อไปนี้

1) ผลการศึกษารวบรวมข้อมูลและทำความเข้าใจข้อมูล

รวบรวมข้อมูลพื้นที่เพาะปลูก จำนวนครัวเรือน และผลผลิตกล้วยไม้จาก www.oae.go.th และ www.ditp.go.th ในปี 2559 - 2563 จำนวน 20 จังหวัด ประกอบด้วย เชียงใหม่ สกลนคร นครพนม ขอนแก่น นครราชสีมา สระบุรี สุพรรณบุรี พระนครศรีอยุธยา นนทบุรี กรุงเทพมหานคร ปทุมธานี นครนายก ฉะเชิงเทรา ระยอง ชลบุรี สมุทรสาคร นครปฐม กาญจนบุรี ราชบุรี และเพชรบุรี เพื่อนำมาใช้ในกระบวนการต่าง ๆ ในการวิเคราะห์ได้อย่างเหมาะสม แสดงในตารางที่ 1

2) ผลการเตรียมข้อมูล (Data preparation)

จากการสกัดข้อมูลจากผลผลิตกล้วยไม้ในประเทศไทย เพื่อเข้าสู่กระบวนการจัดเตรียมพื้นที่ในการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมจะวิเคราะห์ต่อไป

3) การเตรียมข้อมูล (Data preparation)

3.1) ผลการคัดเลือกข้อมูล (Data selection) จากการคัดเลือกปัจจัย (Feature selection) เช่น จังหวัด ภาค เนื้อที่เพาะปลูก และจำนวนครัวเรือน แสดงในตารางที่ 2

**ตารางที่ 1** ตัวอย่างการคัดเลือกข้อมูล

| Column1 | Column2 | Column3 | Column4 | Column5 | Column6 | Column7 | Column8 | Column9 | Column10 | Column11 |
|---|---|---|---|---|---|---|---|---|---|---|
| กล้วยไม้ : ร้อ | ยละและปริมาณผลผลิต รายเดือน รวมทั้ง ประเทศ รายภาค และรายจังหวัด | null | null | null | null | null | null | ปี 2563 | null | nu |
| null | ร้อยละและปริมาณผลผลิตรายเดือน (ตน์ ) | null | null | null | null | null | null | null | null | nu |
| จังหวัด | ม.ค. | ก.พ. | มี.ค. | เม.ย. | พ.ค. | มิ.ย. | ก.ค. | null | ส.ค. | ก.ย. |
| นครนายก | 0.80 | 0.77 | 0.73 | 0.72 | 0.70 | 0.79 | 0.84 | null | 0.93 | 0.92 |
| สุพรรณบุรี | 7.46 | 7.36 | 6.82 | 6.72 | 13.56 | 8.40 | 4.55 | null | 9.08 | 8.36 |
| null | 30.88 | 30.47 | 28.23 | 27.82 | 56.14 | 34.78 | 18.84 | null | 37.59 | 34.61 |
| อยุธยา | 8.14 | 8.80 | 5.84 | 4.55 | 5.65 | 5.48 | 6.95 | null | 7.98 | 8.32 |
| null | 59 | 64 | 42 | 33 | 41 | 40 | 50 | null | 58 | 60 |
| นนทบุรี | 8.23 | 6.73 | 6.67 | 4.99 | 6.01 | 6.19 | 7.21 | null | 7.93 | 7.99 |
| null | 137 | 112 | 111 | 83 | 100 | 103 | 120 | null | 132 | 133 |
| กรุ เทพฯ | 8.12 | 6.59 | 5.64 | 4.49 | 6.87 | 7.15 | 7.68 | null | 8.38 | 8.18 |
| null | 291 | 236 | 202 | 161 | 246 | 256 | 275 | null | 300 | 293 |
| ปทุมธานี | 8.76 | 7.30 | 6.20 | 4.74 | 5.47 | 6.20 | 7.30 | null | 8.03 | 8.39 |
| null | 24 | 20 | 17 | 13 | 15 | 17 | 20 | null | 22 | 23 |
| ฉะเชิ เทรา | 9.63 | 7.11 | 6.39 | 5.04 | 6.12 | 5.94 | 6.84 | null | 7.65 | 8.10 |
| null | 1.07 | 0.79 | 0.71 | 0.56 | 0.68 | 0.66 | 0.76 | null | 0.85 | 0.90 |
| ระยอง | 9.30 | 7.64 | 6.42 | 5.76 | 7.20 | 8.19 | 9.86 | null | 9.63 | 9.75 |
| null | 0.84 | 0.69 | 0.58 | 0.52 | 0.65 | 0.74 | 0.89 | null | 0.87 | 0.88 |
| ชลบุรี | 8.96 | 7.46 | 6.47 | 5.47 | 6.97 | 7.96 | 9.95 | null | 10.45 | 10.45 |
| null | 18 | 15 | 13 | 11 | 14 | 16 | 20 | null | 21 | 21 |
| สมุทรสาคร | 9.11 | 7.05 | 6.27 | 4.78 | 5.68 | 6.10 | 7.25 | null | 7.49 | 7.74 |
| null | 933 | 722 | 642 | 490 | 582 | 625 | 743 | null | 767 | 793 |
| นครปฐม | 8.58 | 7.02 | 6.34 | 4.68 | 5.72 | 5.77 | 6.97 | null | 7.80 | 8.54 |
| null | 1,377 | 1,127 | 1,018 | 751 | 918 | 926 | 1,119 | null | 1,252 | 1,371 |

**ตารางที่ 2** ตัวอย่างการคัดเลือกข้อมูล

| ประเทศ/ภาค/จงัหวดั | Column3 | Column4 | Column5 | Column6 | Column7 | Column8 | Column9 | เนือทีเพาะปลูก | จฺ | นวนครวั | เรือน | เนือที/ครวั | เรือน_1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| null | null | null | null | null | null | null | null | (ไร่) | null | (ครวั | เรือน) | (ไร่) | null |
| รวมทงัประเทศ | null | null | null | null | null | null | null | 21,521 | null | null | 1,593 | null | 14 |
| ภาคเหนือ | null | null | null | null | null | null | null | 45 | null | null | 1 | null | 45 |
| ภาคตะวนั | null | null | null | null | null | null | ออกเฉียงเหนือ | 351 | null | null | 10 | null | 35 |
| ภาคกลาง | null | null | null | null | null | null | null | 21,125 | null | null | 1,582 | null | 13 |
| งใหม่ | null | null | null | null | null | null | null | 45 | null | null | 1 | null | 45 |
| สกลนคร | null | null | null | null | null | null | null | 34 | null | null | 2 | null | 17 |
| นครพนม | null | null | null | null | null | null | null | 13 | null | null | 1 | null | 13 |
| ขอนแก่น | null | null | null | null | null | null | null | 4 | null | null | 1 | null | 4 |
| นครราชสมี | null | null | null | null | null | null | า | 300 | null | null | 6 | null | 50 |
| สระบรุ | | null | null | null | null | null | null | 49 | null | null | 2 | null | 25 |
| รรณบรุ | null | null | null | null | | null | null | 232 | null | null | 13 | null | 18 |
| ยา | null | null | null | null | null | null | null | 450 | null | null | 14 | null | 32 |
| นนทบรุ | null | null | | null | null | null | null | 917 | null | null | 62 | null | 15 |
| กรงุเทพฯ | null | null | null | null | null | null | null | 2,066 | null | null | 180 | null | 11 |
| ปทมุ ธานี | null | null | null | null | null | null | null | 129 | null | null | 12 | null | 11 |

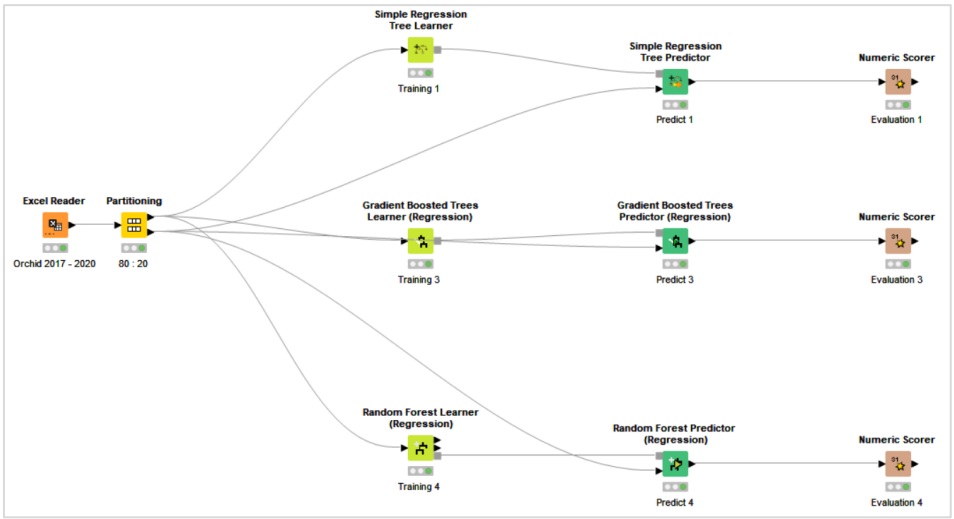
3.2) ผลการกลั่นกรองข้อมูล (Data cleaning) ดำเนินการลบข้อมูลซ้ำซ้อน แก้ไขข้อมูลผิดพลาด เช่น ข้อมูลที่ผิดรูปแบบ ข้อมูลที่หายไป ข้อมูล Outlier ที่แปลกแยกจากคนอื่น ผิดปกติค่อนข้างมาก เป็นต้น แสดงในตารางที่ 3

**ตารางที่ 3** ผลการกลั่นกรองข้อมูล

| Area | Planted area (Rai) | Product (Ton) | Household (Household) |
|---|---|---|---|
| **Thailand** | **21521.00** | **48794.00** | **1593.00** |
| **Northern** | **45.00** | **73.00** | **1.00** |
| **Northeastern** | **351.00** | **793.00** | **10.00** |
| **Central** | **21125.00** | **47928.00** | **1582.00** |
| Chiang Mai | 45.00 | 73.00 | 1.00 |
| Sakon Nakhon | 34.00 | 46.00 | 2.00 |
| Nakhon Phanom | 13.00 | 21.00 | 1.00 |
| Khon Kaen | 4.00 | 5.00 | 1.00 |
| Nakhon Ratchasima | 300.00 | 721.00 | 6.00 |
| Saraburi | 49.00 | 86.00 | 2.00 |
| Suphan Buri | 232.00 | 317.00 | 13.00 |
| Phra Nakhon Si Ayutthaya | 450.00 | 777.00 | 14.00 |
| Nonthaburi | 917.00 | 2112.00 | 62.00 |
| Bangkok | 2066.00 | 4651.00 | 180.00 |
| Pathum Thani | 129.00 | 267.00 | 12.00 |
| Nakhon Nayok | 7.00 | 11.00 | 1.00 |
| Chachoengsao | 3.00 | 5.00 | 1.00 |
| Rayong | 4.00 | 6.00 | 1.00 |
| Chonburi | 114.00 | 238.00 | 20.00 |
| Samut Sakhon | 5232.00 | 13033.00 | 505.00 |
| Nakhon Pathom | 8546.00 | 20177.00 | 653.00 |
| Kanchanaburi | 1372.00 | 2684.00 | 26.00 |
| Ratchaburi | 1986.00 | 3529.00 | 89.00 |
| Phetchaburi | 18.00 | 36.00 | 3.00 |

4) ผลการใช้อัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้

ผลวิเคราะห์หาความสัมพันธ์ และอัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้ ได้แก่ Simple regression tree, Gradient boosted tree และ Random forest แสดงในภาพที่ 1 สรุปเป็นตารางที่ 4




**ภาพที่ 1** ผลการใช้อัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้

**ตารางที่ 4** ความถูกต้องของแต่ละอัลกอริทึม

| | Regression tree | Gradient boosted tree | Random forest |
|---|---|---|---|
| $R^2$ | 0.964 | 0.975 | 0.928 |
| Mean absolute error | 299.725 | 288.061 | 590.459 |
| Mean squared error | 645,660.238 | 449,340.988 | 1,277,742.649 |
| Root mean squared error | 803.53 | 670.329 | 1,130.373 |
| Mean signed difference | -258.725 | -204.791 | 238.628 |
| Mean absolute percentage error | 0.147 | 0.19 | 5.726 |

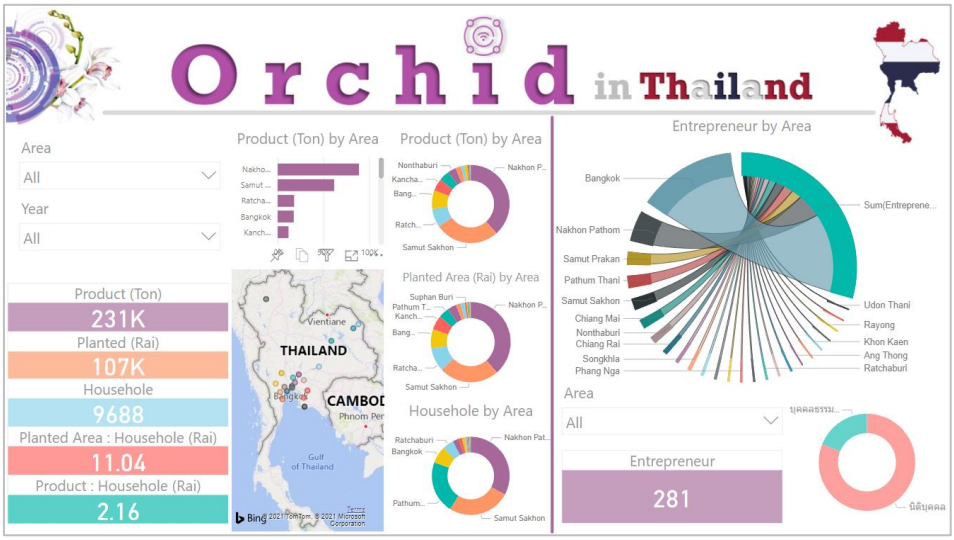
5) ผลการประเมินผล (Evaluation)

ผลการทดสอบประสิทธิภาพอัลกอริทึม ด้วยการหาค่า $R^2$ เพื่อหาอัลกอริทึมที่ดีที่สุด และวัดผล ผลการศึกษา พบว่า การพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทยจากอัลกอริทึม Gradient boosted tree มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 97.50 เมื่อเทียบกับอัลกอริทึม Simple regression tree และ Random forest ถูกต้องร้อยละ 96.40 และ 92.80 ตามลำดับ

6) ผลการถ่ายทอดเพื่อนำไปปฏิบัติ (Deployment)

จากอัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้ การถ่ายทอดเพื่อนำไปปฏิบัติ สามารถนำมาแสดงผลข้อมูลด้วยภาพ (Data visualization) โดยใช้ Power BI แสดงในภาพที่ 2 เพื่อนำข้อมูลการพยากรณ์ผลผลิตกล้วยไม้ใช้ในการวางแผนอุตสาหกรรมกล้วยไม้ เช่น การ

วางแผนการผลิต การวางแผนการส่งออก การจัดการความต้องการของประเทศคู่ค้า และการประชาสัมพันธ์




**ภาพที่ 2** สถิติผลผลิต พื้นที่เพาะปลูก ครัวเรือน และผู้ส่งออกกล้วยไม้ของประเทศไทย

## วิจารณ์

เทคนิคการเรียนรู้ของเครื่องเป็นเทคโนโลยีสมัยใหม่ที่สามารถประยุกต์ใช้ด้านการเกษตรเพื่อการเปลี่ยนถ่ายสู่เกษตรยุค 4.0 เป็นเกษตรแม่นยำ (Precision agriculture) ช่วยให้เกษตรกร และหน่วยงานที่เกี่ยวข้องสามารถพยากรณ์ผลผลิตได้แม่นยำขึ้น สนับสนุนการทำเกษตรกรรม ยกระดับชีวิตเกษตรกร และช่วยเพิ่มขีดความสามารถในการแข่งขันมากยิ่งขึ้น ตลอดจนสามารถประยุกต์ใช้ในอุตสาหกรรมของพืชเศรษฐกิจชนิดต่าง ๆ ได้

## สรุป

ในการวิเคราะห์หาความสัมพันธ์ และใช้อัลกอริทึมพยากรณ์ผลผลิตกล้วยไม้ด้วย 3 อัลกอริทึม ได้แก่ Simple regression, Gradient boosted tree และ Random forest การทดสอบประสิทธิภาพอัลกอริทึมด้วยการหาค่า $R^2$ เพื่อหาอัลกอริทึมที่เหมาะสมที่สุด และวัดผล ซึ่งผลการศึกษา พบว่า การพยากรณ์ผลผลิตกล้วยไม้ในประเทศไทยจากอัลกอริทึม Gradient boosted tree มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 97.50 เมื่อเทียบกับอัลกอริทึม Simple regression tree และ Random forest ถูกต้องร้อยละ 96.40 และ 92.80

ตามลำดับ สามารถนำมาแสดงผลข้อมูลด้วยภาพโดยใช้ Power business intelligence (Power BI) ซึ่งเป็นประโยชน์ในการวางแผนการผลิต เพื่อเพิ่มยอดขายได้เป็นอย่างดี

**กิตติกรรมประกาศ**

ขอขอบคุณหลักสูตรประกาศนียบัตร (Non-degree) การประยุกต์การเรียนรู้ของเครื่อง (Machine learning) เพื่อใช้สนับสนุนการตัดสินใจในธุรกิจ E-commerce คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย ที่ถ่ายทอดองค์ความรู้ที่เป็นประโยชน์ให้สามารถประยุกต์ในงานวิจัยนี้

**เอกสารอ้างอิง**

1. จิรนันท์ เอี่ยมศริยารักษ์. สถานการณ์และการพยากรณ์ปริมาณผลผลิตปริมาณส่งออกและราคาส่งออกของกล้วยไม้ตัดดอกของประเทศไทย. กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์; 2551.
2. Shearer C. The CRISP-DM model: the new blueprint for data mining. J Data Warehous 2000;5:13-22.
3. เอกสิทธิ์ พัชรวงศ์ศักดา. การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิงเบื้องต้น. พิมพ์ครั้งที่ 2. กรุงเทพฯ: เอเชีย ดิจิตอลการพิมพ์ จำกัด; 2557.
4. Smriti S. What is mean squared error, mean absolute error, root mean squared error and r squared? [Internet]. 2019 [cited 2021 August 4]. Available from: https://www.studytonight.com/post/what-is-mean-squared-error-mean-absolute-error-root-mean-quared-error-and-rsquared
5. นฤพนต์ ว่องประชานุกูล. วิธีที่เหมาะสมสำหรับการตัดกิ่งต้นไม้ตัดสินใจของการทำเหมืองข้อมูลทางด้านวิทยาศาสตร์ [วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต]. นครราชสีมา: มหาวิทยาลัยเทคโนโลยีสุรนารี; 2548.
6. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken: Wiley; 2013.