

การหาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องเพื่อประยุกต์ใช้จำแนกผลผลิตทางการเกษตร

An Efficiency of Algorithms Machine Learning for Agricultural Product Classification

ลัคนันท์ พลอยวัฒนาวงศ์¹, ปริญา นาโท², พยุง มีสัจ³

^{1,2}สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
1518 ถนนประชากรบุรี 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพมหานคร 10800

¹luxsanun.p@rmutsb.ac.th, ²p.natho@hotmail.com, ³pym@kmutnb.ac.th

บทนำ

การพัฒนาทางด้านการเกษตรในปัจจุบันมีการนำเทคโนโลยี นวัตกรรม และสิ่งประดิษฐ์เข้ามาช่วยในงานด้านต่างๆ ตั้งแต่การเพาะปลูก การเก็บเกี่ยว การคัดแยกผลผลิต จนถึงระบบการจัดจำหน่าย แต่ก็ยังคงพบปัญหาที่เกิดขึ้นในการคัดแยกผลผลิตที่จำเป็นต้องใช้แรงงานและระยะเวลาในการทำงานอย่างมาก งานวิจัยนี้ได้ศึกษาการนำเทคนิคการเรียนรู้ของเครื่องมาแก้ปัญหาในการคัดแยกผลผลิตทางการเกษตร โดยวิเคราะห์และเปรียบเทียบหาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องจากโมเดล 6 รูปแบบได้แก่ Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNb และ SVM ที่มีความเหมาะสมกับลักษณะข้อมูล โดยใช้ชุดข้อมูลเมล็ดพันธุ์ข้าวไทย เทคนิคการเรียนรู้ของเครื่องสามารถทำการคัดแยกสายพันธุ์ข้าวได้ 3 ประเภทจากจำนวนข้าว 130 สายพันธุ์ นอกจากนั้นพบว่ารูปแบบ KNeighbors Classifier ได้ค่าความถูกต้องสูงสุดเป็น 0.800000 ซึ่งเหมาะกับการเรียนรู้กับลักษณะข้อมูลการคัดแยกสายพันธุ์ข้าว แสดงให้เห็นว่าเทคนิคการเรียนรู้ของเครื่องสามารถลดขั้นตอนการทำงานในการคัดแยกสายพันธุ์ข้าวได้อย่างมีประสิทธิภาพ

คำสำคัญ: อัลกอริทึม, เทคนิคการเรียนรู้ของเครื่อง, การเรียนรู้ของเครื่อง, การคัดแยก

Abstract

The agriculture today was to use apply the innovative technologies and inventions to support the management of works such as: planting, harvesting classification product, the distribution systems. However, there were still problems in classification product the labour productivity and the length of work. This research to study and the use of machine learning techniques to solve the problems of classification product agricultural. Analyze and compare the efficiency of the machine learning algorithm from 6 models: Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNb and SVM. And using the Thai Rice Seed dataset as a test kit. The machine learning techniques can be identified into three types of rice varieties from 130 varieties. In addition, the KNeighbors Classifier has the highest accuracy value of 0.800000. It showed that the KNeighbors Classifier was suitable for learning the characteristics data of rice identification. This demonstrates that the machine learning technique was able to reduce the work process and improved the efficiency of rice classification varieties.

Keyword: algorithm, machine learning technique, machine learning, classification

1. บทนำ

ในปัจจุบันความก้าวหน้าทางด้านเทคโนโลยีมีวิวัฒนาการไปอย่างรวดเร็ว มีการใช้งานเทคโนโลยี อุปกรณ์อิเล็กทรอนิกส์ที่มีความฉลาดและสามารถเข้าใจในสิ่งที่มนุษย์ต้องการมากขึ้น การเรียนรู้และความฉลาดของเครื่องจักรกลมีการพัฒนามาจากวิทยาการทางด้านปัญญาประดิษฐ์ (Artificial Intelligence) [1] ซึ่งเป็นเทคโนโลยีที่เชื่อกันว่าเป็นวิทยาการที่จะช่วยให้มนุษย์สามารถนำไปใช้งานเพื่อแก้ไขปัญหาต่างสาขาต่างๆ ที่สำคัญ เช่น การให้คอมพิวเตอร์เข้าใจภาษามนุษย์ ตลอดจนการผลิตหุ่นยนต์ หรือแม้แต่การนำมาใช้งานในภาคเกษตรกรรมก็มีการนำความรู้ด้านปัญญาประดิษฐ์เข้ามาช่วยในการคัดแยก จัดกลุ่ม ควนคัม และการพยากรณ์สินค้าและการผลิตพืชผลด้านการเกษตร เพื่อช่วยลดปัญหาการใช้เวลาในการคัดแยกสายพันธุ์พืช การจัดกลุ่มเพื่อช่วยคัดเกรดของพืชผล [2] การควบคุมปริมาณของปัจจัยต่างๆ ที่อาจเกิดขึ้นในช่วงระหว่างการเพาะปลูก

การนำเทคโนโลยีปัญญาประดิษฐ์เข้ามาใช้สามารถประสานการโต้ตอบระหว่างเครื่องจักรกับเครื่องจักรได้โดยอัตโนมัติ ตัวอย่างเช่น ระบบควบคุมสำหรับศูนย์ข้อมูลสามารถใช้ AI เพื่อตรวจสอบกิจกรรมการทำงานของคอมพิวเตอร์ อุณหภูมิภายในและสภาพแวดล้อมได้ตลอดเวลา นอกจากนั้นสามารถปรับเปลี่ยนระบบระบายความร้อน เพื่อเพิ่มประสิทธิภาพการทำงานในขณะที่ลดค่าใช้จ่ายด้านพลังงานลง [3] ความสามารถนี้ยังช่วยให้ระบบ AI หลายระบบสามารถทำงานร่วมกันได้ เช่น กองยานอวกาศที่จัดการตัวเองเพื่อสร้างกองทหาร สามารถลดการใช้เชื้อเพลิงหรือหุ่นยนต์อวกาศที่สื่อสารกันเพื่อจัดเรียงสิ่งข้อมูล [4] และการพัฒนาเครื่องจักรให้มีความสามารถทำงานได้ดีเพียงอย่างหนึ่งหรืองานที่จำเพาะเจาะจง ซึ่งความสามารถระดับนี้คือปัญญาประดิษฐ์ในระดับที่เรียกว่า Narrow AI จะมีความสามารถเทียบเท่าหรือเหนือกว่ามนุษย์ในด้านใดด้านหนึ่งโดยเฉพาะ เช่น การแยกแยะภาพ การแปลภาษา การจดจำใบหน้า

การนำการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเข้ามาช่วยในการคัดแยกผลผลิตการเกษตรเป็นการพัฒนา

ระบบเทคโนโลยีเพื่อให้เครื่องคอมพิวเตอร์เข้ามาช่วยมนุษย์ในการทำงานให้มีความถูกต้องและแม่นยำเพิ่มมากขึ้นนั้น มีรูปแบบหรือโมเดลที่เข้ามาช่วยในการตรวจสอบความถูกต้องของข้อมูลอยู่ไม่น้อย แต่ปัญหาคือโมเดลหรืออัลกอริทึมแต่ละตัวต่างก็มีกระบวนการเรียนรู้ที่แตกต่างกัน ซึ่งถ้าหากเลือกใช้โมเดลเข้ามาช่วยในการเรียนรู้ที่ไม่เหมาะสมกับข้อมูลแล้วอาจจะทำให้ผลลัพธ์ที่ได้มีค่าไม่ถูกต้องหรือขาดความเชื่อถือเช่นกัน

ดังนั้นงานวิจัยฉบับนี้ได้ทำการสร้างโมเดลสำหรับทดสอบประสิทธิภาพการคัดแยกพืชผลทางการเกษตรด้วยการใช้ข้อมูลเมล็ดพันธุ์ข้าวเป็นชุดข้อมูล (Dataset) ในการหาประสิทธิภาพโมเดลการเรียนรู้ของเครื่องจักรและจัดกลุ่มเมล็ดพันธุ์ข้าว โดยใช้การเรียนรู้ของเครื่อง ซึ่งตัวชี้วัดที่จะนำมาวิเคราะห์ประสิทธิภาพของการทำงานจะประกอบด้วย การประเมินค่าความถูกต้องของโมเดลทั้งหมด 6 โมเดลคือ Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNb และ SVM ที่สามารถคัดแยกเมล็ดพันธุ์ข้าวไทยจำนวน 130 สายพันธุ์ได้ถูกต้องมากที่สุด

งานวิจัยนี้ได้นำเสนอหัวข้อตามลำดับ สำหรับส่วนที่สอง อธิบายถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่สาม วิเคราะห์การวิจัย ส่วนที่สี่ ผลการดำเนินการวิจัย และส่วนที่ห้า สรุปผลการวิจัย

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีโมเดล Logistic Regression

Logistic Regression เป็นเทคนิคการวิเคราะห์ตัวแปรเชิงพหุแบบหนึ่ง คือการวิเคราะห์ข้อมูลเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) และตัวแปรอิสระ (Independent Variable) แล้วนำสมการถดถอยที่ได้ไปประมาณหรือพยากรณ์ค่าตัวแปรตามเมื่อกำหนดค่าตัวแปรอิสระได้ แตกต่างกันที่การวิเคราะห์สมการถดถอยแบบปกติ มีเงื่อนไขคือ ต้องมีค่าความคลาดเคลื่อนมีความแจกแจงแบบปกติ และค่าแปรปรวนของค่าความคลาดเคลื่อนคงที่ แต่สมการถดถอยแบบ Logistic ไม่จำเป็นต้องมีเงื่อนไขดังกล่าว

เนื่องจากตัวแปรตามมีค่าได้เพียง 2 ค่า ทำให้ค่าคลาดเคลื่อนเป็นได้เพียง 2 ค่าด้วย ซึ่งเป็นไปไม่ได้ที่ค่าความคลาดเคลื่อนจะมีการแจกแจงแบบปกติจึงต้องใช้การวิเคราะห์ความถดถอย Logistic แทนการวิเคราะห์ความถดถอยทั่วไป [5]

2.2 ทฤษฎีโมเดล Linear Discriminant Analysis

การวิเคราะห์จำแนกกลุ่มด้วยเทคนิค Linear Discriminant Analysis [6] เป็นวิธีการทางสถิติ ที่ใช้วิเคราะห์จำแนกกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป ด้วยการวิเคราะห์จากตัวแปรตาม 1 ตัวและตัวแปรอิสระตั้งแต่ 1 ตัวขึ้นไป การวิเคราะห์ด้วยวิธีนี้ นอกจากจะสามารถจำแนกความแตกต่างระหว่างกลุ่มได้แล้ว ยังสามารถบอกธรรมชาติบางอย่างของการจำแนกกลุ่มได้ด้วย เช่น บอกได้ว่าตัวแปรใดจำแนกได้ดีมากน้อยไปกว่ากัน นั่นคือ สามารถบอกประสิทธิภาพ หรือน้ำหนักในการจำแนก ของตัวแปรเหล่านั้น การวิเคราะห์จำแนกกลุ่มเป็นการใช้ตัวแปรพยากรณ์หรือตัวแปรอิสระ ที่ร่วมกันพยากรณ์ตัวแปรตาม ซึ่งเป็นเทคนิคทางสถิติที่คล้ายคลึงกับการวิเคราะห์ถดถอยพหุคูณ (Multiple Regression Analysis)

2.3 ทฤษฎีโมเดล KNeighbors Classifier

KNeighbors Classifier เป็นการจัดกลุ่มข้อมูล โดยการจัดข้อมูลที่อยู่ใกล้กันให้เป็นกลุ่มเดียวกันซึ่งเทคนิคนี้จะทำให้ตัดสินใจได้ว่า คลาสไหนที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวน K ซึ่งถ้าหากเงื่อนไขของการตัดสินใจมีความซับซ้อน วิธีนี้สามารถสร้างโมเดลที่มีประสิทธิภาพได้ แต่ขั้นตอนวิธีการ KNeighbors Classifier จะใช้ระยะเวลาในการคำนวณนาน ถ้าตัวแปร (Attribute) มีจำนวนมาก จะเกิดปัญหาในการคำนวณค่าและค่อนข้างใช้ปริมาณงานในการคำนวณสูงมากบนคอมพิวเตอร์ เพราะเวลาที่ใช้สำหรับการคำนวณจะเพิ่มขึ้นแบบแฟกทอเรียลตามจำนวนจุดทั้งหมด ดังนั้นเพื่อจะเพิ่มความเร็วสำหรับเทคนิคขั้นตอนวิธีการนี้ให้มากขึ้น ข้อมูลทั้งหมดที่ใช้บ่อยจะต้องถูกเก็บไว้ในหน่วยความจำ (Memory) โดยวิธีการเข้าถึงหน่วยความจำพื้นฐานอย่างมีเหตุผล (Memory-Based Reasoning) ซึ่งจะเป็นวิธีที่นำมาอ้างอิงถึงเป็นประจำในการ

จัดเก็บกลุ่มคลาสของขั้นตอนวิธีการ KNeighbors Classifier ในหน่วยความจำ และ ถ้าหากข้อมูลที่ต้องการหาคำตอบมีตัวแปรอิสระเพียงไม่กี่ตัวแล้ว จะทำให้เราสามารถเข้าใจโมเดลขั้นตอนวิธีการ KNeighbors Classifier ได้ง่ายขึ้น ตัวแปรเหล่านี้ยังมีประโยชน์สำหรับนำมาสร้างโมเดลต่าง ๆ ที่เกี่ยวข้องกับชนิดของข้อมูลที่ไม่เป็นมาตรฐาน เช่น ข้อความ (Text) เพียงแต่อาจต้องมีมาตรฐานการวัดค่าสำหรับชนิดของข้อมูลดังกล่าวที่เหมาะสมด้วย นอกจากนี้ประสิทธิภาพของขั้นตอนวิธีการ KNeighbors Classifier จะขึ้นอยู่กับจำนวนระยะห่าง การอธิบายระหว่างข้อมูลทั้งคู่ ที่สามารถแบ่งแยกข้อมูลได้อย่างมีประสิทธิภาพระหว่างข้อมูลปกติและข้อมูลผิดปกติ การอธิบายจำนวนระยะห่างระหว่างข้อมูลเป็นความท้าทายอย่างมากเมื่อข้อมูลมีความซับซ้อน [7]

2.4 ทฤษฎีโมเดล Decision Tree Classifier

Decision Tree [8] เป็นกระบวนการเรียนรู้แบบจำแนกโดยการแบ่งแยกข้อมูลออกเป็นแขนงสาขา ซึ่งค่อนข้างสอดคล้องกับการตัดสินใจจริงของมนุษย์ สามารถแสดงเป็นภาพที่เข้าใจได้ง่าย และทำงานได้ดีกับข้อมูลที่มีความเกี่ยวพันของข้อมูลสูง และอาจมีหลายลำดับขั้น ในการเรียนรู้ของเครื่องรูปแบบ Decision Tree เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุ โดยพิจารณาจากลักษณะของวัตถุ โหนดภายใน (Inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนโหนดใบจะแสดงประเภทของวัตถุ

2.5 ทฤษฎีโมเดล GaussianNb

GaussianNb เป็นการแจกแจงความน่าจะเป็นของค่าของตัวแปรสุ่มที่เป็นค่าแบบต่อเนื่อง โดยที่ค่าของตัวแปรสุ่มมีแนวโน้มที่จะมีค่าอยู่ใกล้ๆ กับค่าๆ หนึ่ง (เรียกว่า ค่ามัชฌิม) กราฟแสดงค่าฟังก์ชันความหนาแน่น (Probability density function) จะเป็นรูปคล้ายระฆังคว่ำ หรือเรียกว่า Gaussian function โดยค่าฟังก์ชันความหนาแน่นของการแจกแจงปกติได้แก่

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

โดย X แทนตัวแปรสุ่ม พารามิเตอร์ μ แสดงค่ามัชฌิม และ σ^2 คือ ค่าความแปรปรวน (Variance) ซึ่งเป็นค่าที่ใช้บอกปริมาณการกระจายของการแจกแจง การแจกแจงปกติที่มีค่า $\mu = 0$ และ $\sigma^2 = 1$ จะถูกเรียกว่า การแจกแจงปกติมาตรฐาน [9]

2.6 ทฤษฎีโมเดล SVM

SVM (Support Vector Machine) เป็นโมเดลที่ใช้การคำนวณทางคณิตศาสตร์ ซึ่งจัดอยู่ในกลุ่มเดียวกับ Neural Network โดยวิธีการของ SVM [10] เป็นเทคนิคที่ใช้ในการแก้ปัญหาทางด้านการรู้จำรูปแบบข้อมูล โดยอาศัยหลักการของการหาสมมติฐานของสมการ เพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด สำหรับรากฐานเดิมของ SVM ถูกนำมาใช้กับข้อมูลที่เป็นชนิดเชิงเส้นแต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้ในกระบวนการสอนให้เครื่องเรียนรู้ โดยส่วนใหญ่แล้วมักจะเป็นข้อมูลชนิดแบบไม่เป็นเชิงเส้น

2.7 ทฤษฎีการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine Learning) [11] เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง การนำการเรียนรู้ของเครื่องเข้ามาใช้ในงานเกษตรได้รับสนใจเป็นอย่างมาก เช่น การพัฒนาระบบการทำฟาร์มในร่มแบบอิสระที่ใช้เซ็นเซอร์เครือข่ายและการเรียนรู้ของเครื่อง เพื่อเฝ้าติดตามการทำงานภายในอาคารเพาะปลูกตลอดเวลา ทั้งในด้านสภาพแวดล้อมของฟาร์มและการเจริญเติบโตของพืช การปรับแสง อุณหภูมิ ความชื้น น้ำและระดับธาตุอาหารในดิน เพื่อเพิ่มผลผลิตของฟาร์มอย่างมีประสิทธิภาพ [12] โดยใช้อัลกอริทึมการเรียนรู้

ของเครื่องในการทำการวิเคราะห์ข้อมูลและคาดการณ์ปัญหาที่อาจเกิดขึ้น เพื่อวางแผนดำเนินการป้องกันได้

สามารถนำมาพัฒนางานวิจัยในการเรียนรู้โรคของพืช โดยการพัฒนาแบบที่อาศัยการทำงานของเครือข่ายเซ็นเซอร์ของกล้องและอัลกอริทึมการเรียนรู้ของเครื่อง เพื่อตรวจสอบพืชและเตือนการเกิดโรค อัลกอริทึมสามารถระบุสัญญาณของเชื้อราแบคทีเรียหรือความเสียหายของการทำลายจากแมลง รวมทั้งการวิเคราะห์ข้อมูลโภชนาการพืช สภาพอากาศ การคาดการณ์และข้อมูลอื่นๆ เพื่อคาดการณ์ความเสี่ยงที่อาจเกิดขึ้นจากปัญหาข้างต้น จากข้อมูลในอดีตและวิเคราะห์การคาดการณ์สำหรับรองรับการแก้ปัญหา [13], [14]

นอกจากนี้ยังมีการพยากรณ์ผลผลิตการเกษตร เป็นงานวิจัยจาก Descartes Labs ที่ได้ทำการปรับใช้ซอฟต์แวร์วิเคราะห์ภาพด้วยเทคนิคการเรียนรู้เชิงลึก ในการวิเคราะห์ภาพถ่ายดาวเทียมของพื้นที่เพาะปลูก เพื่อคาดการณ์ผลผลิตการเกษตรได้รวดเร็วและถูกต้องมากขึ้น [15] ซอฟต์แวร์สามารถประมาณการเพาะปลูกและการผลิตเป็นรายสัปดาห์ โดยการเปรียบเทียบภาพถ่ายรายวัน จากฟาร์มข้าวโพดขนาด 3 ล้านตารางกิโลเมตร ซึ่งพบว่ามีความคลาดเคลื่อนน้อยกว่า 1% ส่งผลดีในการช่วยเหลือเกษตรกร บริษัทประกัน ธุรกิจประกันภัยและรัฐบาลในการตัดสินใจด้วยข้อมูลที่มีประสิทธิภาพมากขึ้น

อีกทั้งยังมีการนำเทคนิคการเรียนรู้ของเครื่องเพื่อเข้ามาใช้ในการคัดแยกประเภทของผลผลิตการเกษตร พบนงานวิจัยหลายชิ้นที่ได้พยายามนำเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้กับการคัดแยก จัดกลุ่มผลผลิตการเกษตร ซึ่งเป็นแนวความคิดที่เป็นไปได้และสามารถทำงานได้อย่างมีประสิทธิภาพ ตัวอย่างเช่น งานวิจัยที่เกิดขึ้นจากเกษตรกรสวนแตงกวาในประเทศญี่ปุ่น ผู้วิจัยได้สร้างระบบหุ่นยนต์ขับเคลื่อนโดยซอฟต์แวร์ TensorFlow เป็นซอฟต์แวร์การเรียนรู้ของเครื่อง และการเรียนรู้เชิงลึก ซึ่งเป็นซอฟต์แวร์ฟรีของบริษัท Google ทำให้สามารถจัดเรียง คัดแยกแตงกวาได้อย่างอัตโนมัติ โดยแยกแตงกวาตามความแตกต่างของภาพ เช่น ขนาด รูปร่าง [16] เกษตรกรได้ทำการฝึกขั้นตอนการมองเห็นของคอมพิวเตอร์บนระบบ เพื่อให้แยกความแตกต่าง

ของแตงกวาออกเป็น 9 ประเภท พบว่าระบบสามารถคัดแยกแตงกวาได้อย่างถูกต้องเป็นร้อยละ 70 และสามารถทำได้เร็วกว่าการคัดแยกด้วยมนุษย์ สำหรับฟาร์มขนาดเล็กที่ต้องใช้เวลาถึง 8 ชั่วโมงต่อวัน

3. วิธีดำเนินการวิจัย

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการคัดแยกเมล็ดพันธุ์ข้าวโดยใช้การเรียนรู้ของเครื่อง เพื่อทดสอบหาค่าความถูกต้องของโมเดลทั้งหมด 6 รูปแบบคือ Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNb และ SVM ว่าโมเดลรูปแบบใดมีประสิทธิภาพในการวิเคราะห์คัดแยกสายพันธุ์ข้าวได้ถูกต้องแม่นยำมากที่สุด สำหรับเครื่องมือที่ผู้วิจัยได้ทดสอบผู้วิจัยเลือกใช้ภาษา Python ในการพัฒนาโปรแกรม ทดสอบบนเครื่องคอมพิวเตอร์ Laptop ที่ติดตั้งระบบปฏิบัติการ Windows 7 Professional ความเร็ว CPU AMD E-450 APU with Radeon HD Graphics 1.65 GHz 32 bit ขนาดหน่วยความจำ 512MB โดยจะทำการทดสอบกับชุดข้อมูลพันธุ์ข้าวไทยทั้งหมด 130 สายพันธุ์

ผู้วิจัยได้กำหนดข้อมูลข้าวไทยจำนวน 130 สายพันธุ์ โดยกำหนดความกว้าง ความลึก และความหนา ของพันธุ์ข้าวต่างๆ ซึ่งพันธุ์ข้าวจะประกอบด้วย 3 ประเภทคือ ข้าวหอมแทนด้วยรหัส J ส่วนข้าวเจ้าแทนด้วยรหัส W และข้าวเหนียวแทนด้วยรหัส S โดยใช้สถานที่ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือและศูนย์วิจัยข้าวสุพรรณบุรีเป็นสถานที่ในการวิจัยครั้งนี้

ในการทดสอบประสิทธิภาพการคัดแยกสายพันธุ์ข้าวจำนวน 130 สายพันธุ์ ผู้วิจัยได้ใช้เทคนิคการเรียนรู้ของเครื่องในการจำแนกกลุ่มของพันธุ์ข้าวและหาค่าความถูกต้องของเทคนิคทั้ง 6 ประเภท

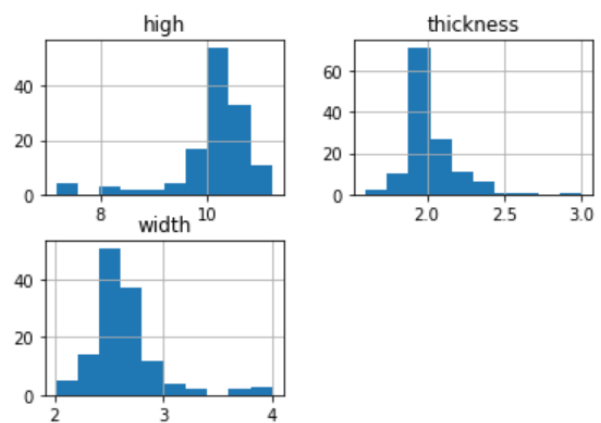
4. ผลการดำเนินการวิจัย

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการคัดแยกเมล็ดพันธุ์ข้าวโดยใช้การเรียนรู้ของเครื่อง เพื่อทดสอบหาค่าความถูกต้องของโมเดลทั้งหมด 6

รูปแบบคือ Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNb และ SVM ว่าโมเดลรูปแบบใดมีประสิทธิภาพในการวิเคราะห์คัดแยกสายพันธุ์ข้าวได้ถูกต้องแม่นยำมากที่สุด

4.1 ผลการทดสอบการจัดกลุ่มชุดพันธุ์ข้าว

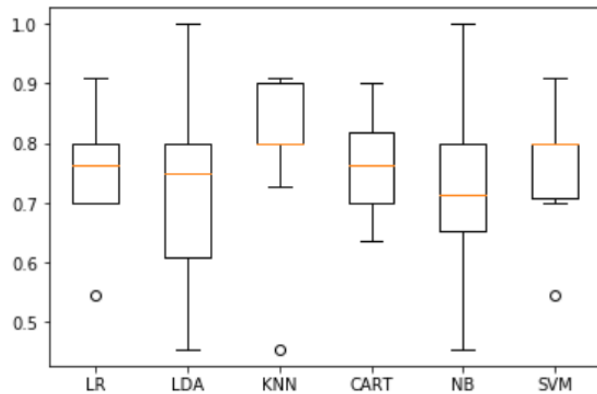
ในการจัดกลุ่มของชุดข้อมูลพันธุ์ข้าวสามารถจัดกลุ่มพันธุ์ข้าวได้ 3 ประเภทคือ ข้าวหอมมะลิมี 3 สายพันธุ์ ข้าวเหนียวจำนวน 29 สายพันธุ์ และข้าวเจ้าจำนวน 98 สายพันธุ์ แสดงดังภาพที่



ภาพที่ 1: กราฟรูปแบบ Histograms

4.2 ผลการทดสอบด้านความถูกต้อง

การประเมินค่าความถูกต้องของอัลกอริทึมทั้งหมด 6 รูปแบบคือ Logistic Regression ได้ผลลัพธ์ความถูกต้องคือ 0.750000 (0.093442), Linear Discriminant Analysis ได้ผลลัพธ์ค่าความถูกต้องคือ 0.720909 (0.146251), KNeighbors Classifier ได้ผลลัพธ์ค่าความถูกต้องคือ 0.800000 (0.129780), Decision Tree Classifier ได้ผลลัพธ์ค่าความถูกต้องคือ 0.770000 (0.085865), GaussianNB ได้ผลลัพธ์ค่าความถูกต้องคือ 0.721818 (0.137942) และ SVM ได้ผลลัพธ์ค่าความถูกต้องคือ 0.760000 (0.092905) แสดงดังภาพที่ 4



ภาพที่ 4: Algorithm Comparison

5. สรุปผลการดำเนินงาน

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง เพื่อทดสอบหาค่าความถูกต้องของโมเดลทั้งหมด 6 รูปแบบคือ Logistic Regression, Linear Discriminant Analysis, KNeighbors Classifier, Decision Tree Classifier, GaussianNB และ SVM ว่าโมเดลรูปแบบใดมีประสิทธิภาพในการวิเคราะห์คัดแยกข้อมูลสายพันธุ์ข้าวได้ถูกต้องแม่นยำมากที่สุด ผลจากการทดสอบพบว่าจากการใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อวิเคราะห์การจำแนกและการจัดกลุ่มชุดข้อมูลพันธุ์ข้าวได้ 3 ประเภทคือ ข้าวหอมมะลิ 3 สายพันธุ์ ข้าวเหนียว 29 สายพันธุ์ และข้าวเจ้า 98 สายพันธุ์ นอกจากนั้นการทดสอบหาค่าความถูกต้องของโมเดลทั้ง 6 รูปแบบพบว่าโมเดลที่ให้ค่าความถูกต้องมากที่สุดคือ โมเดล KNeighbors Classifier ได้ค่าความถูกต้องเป็น 0.800000 รองลงมาคือโมเดล Decision Tree Classifier ได้ค่าความถูกต้อง 0.770000 ถัดมาคือโมเดล SVM ได้ค่าความถูกต้อง 0.760000 สำหรับโมเดล Logistic Regression ได้ค่าความถูกต้อง 0.750000 ส่วนโมเดล GaussianNB ได้ค่าความถูกต้องคือ 0.721818 สุดท้ายคือโมเดล Linear Discriminant Analysis ได้ค่าความถูกต้องคือ 0.7200909

ซึ่งผลจากการทดสอบมีความสอดคล้องกับงานวิจัยที่ผ่านมา [15],[16] ที่ได้นำเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกมาประยุกต์ใช้ในการเกษตรได้อย่างมี

ประสิทธิภาพ สามารถช่วยลดขั้นตอนการทำงานในการคัดแยกพันธุ์ได้อย่างมีประสิทธิภาพ

สำหรับแนวทางในอนาคตสามารถนำอัลกอริทึมที่มีความเหมาะสมไปใช้ด้านการคัดแยกผลผลิตทางการเกษตรแบบอัตโนมัติ เป็นแนวคิดต้นแบบสำหรับการเลือกใช้อัลกอริทึมให้เหมาะสมกับลักษณะของข้อมูล ด้วยเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกในชุดข้อมูลที่มีความแตกต่างกันต่อไป เพื่อทำให้เกิดประโยชน์ในการสร้างผลผลิตทางการเกษตร ควบคุม จัดการและพยากรณ์ความต้องการของผู้บริโภคในอนาคต

เอกสารอ้างอิง

- [1] Stuart J. Russell, Peter Norvig. *Artificial Intelligence: A modern Approach (2nd Edition)*, Prentice Hall, New Jersey, 2003.
- [2] Castro, D., & New, J. (2016). The promise of artificial intelligence. Center for Data Innovation.
- [3] Joshua Hill, "Google's DeepMind AI reduces Data Center Cooling by 40%," Clean Technica, July 22, 2016, <https://cleantechnica.com/2016/07/22/google-deepmind-ai-reduces-data-center-cooling-bill-40/>.
- [4] Eric Limer, "A Fleet of Self-Driving Trucks Just Completed a 1,000-Mile Trip Across Europe," Popular Mechanics, April 7, 2016, <http://www.popularmechanics.com/cars/trucks/a20310/european-platooning-challenge-self-driving-trucks-1000-miles/>; Alexis Madrigal, "Autonomous Robots Invade Retail Warehouses," Wired, January 27, 2009, <https://www.wired.com/2009/01/retailrobots/>.
- [5] Follmann, D.A. and D. Lambert. 1989. Generalizing logistic regression by nonparametric mixing. J. Am. Stat. Assoc. 84:295–300.
- [6] Chelali, F.Z. 2009. Linear discriminant analysis for face recognition. *Multimedia Computing and Systems*, 2009.

- ICMCS '09. International Conference on*. 1 (Apr. 2009), 1–10.
- [7] Gregory Shakhnarovich, Trevor Darrell, Piotr Indyk. Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing series).
- [8] Mitchell, Tom. Machine Learning. McGraw-Hill, 1997, p. 52-80.
- [9] Casella, George; Berger, Roger L. (2001). Statistical inference (2nd ed.). Duxbury.
- [10] Chih-Chung Chang and Chih-Jen Lin. “LIBSVM.” A Library for Support Vector Machines. September, 2006.
- [11] Nelson Sizwe et al., “Integrating Artificial Intelligence into Data Warehousing and Data Mining,” Proceedings of the World Congress on Engineering and Computer Science 2015 2 (October 2015), pp. 819-823, http://www.iaeng.org/publication/WCECS2015/WCECS2015_pp819-823.pdf.
- [12] “How Motorleaf is Helping Automate Indoor Farming,” AgFunderNews, July 20, 2016, <https://agfundernews.com/motorleaf-helping-automate-indoor-farming.html>
- [13] Leanna Garfield, “This Robot Has ‘Eyes’ That Can See Dying Plants Before Farmers Can,” Tech Insider, July 8, 2016, <http://www.techinsider.io/prospera-robot-can-see-dying-plants-before-farmers-2016-7>.
- [14] Robinson Meyer, “The Start-Up That Watches Corn Grow, From Orbit,” The Atlantic, August 10, 2016, <http://www.theatlantic.com/technology/archive/2016/08/seeing-corn-with-satellites/495149/>.
- [15] Joshua New, “5 Q’s for Mark Johnson, Founder of Descartes Labs,” Center for Data Innovation, September 2, 2016, <https://www.datainnovation.org/2016/09/5-qs-for-mark-johnson-founder-of-descartes-labs/>.
- [16] Dave Gershgorn, “The Ultimate Promise of Artificial Intelligence Lies in Sorting Cucumbers,” Quartz, September 1, 2016, <http://qz.com/771921/the-ultimate-promise-of-artificial-intelligence-lies-in-sorting-cucumbers/>.