



แบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้  
ของเครื่องเพื่อตรวจจับรูปแอปเปิ้ล

MACHINE LEARNING-BASED ONE-CLASS FRUIT CLASSIFICATION MODELS  
FOR APPLE IMAGES

ศรัทธา ไซติชาติมาลา

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2563

แบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้  
ของเครื่องเพื่อตรวจจํารูปแอปเปิ้ล



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล  
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ  
ปีการศึกษา 2563  
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

MACHINE LEARNING-BASED ONE-CLASS FRUIT CLASSIFICATION MODELS  
FOR APPLE IMAGES



A Master's Project Submitted in Partial Fulfillment of the Requirements  
for the Degree of MASTER OF SCIENCE  
(Data Science)

Faculty of Science, Srinakharinwirot University

2020

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

แบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้  
ของเครื่องเพื่อตรวจจับรูปแบบเปิด

ของ

ศิริชัย โชติชาติมาลา

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล  
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก ..... ประธาน  
(ผู้ช่วยศาสตราจารย์ ดร.นุ้ย วิวัฒน์วัฒนา) (อาจารย์ ดร.ดวงดาว วิชาดากุล)

..... กรรมการ  
(อาจารย์ ดร.วีระ สอึ้ง)

|                  |                                                                                    |
|------------------|------------------------------------------------------------------------------------|
| ชื่อเรื่อง       | แบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้ของเครื่องเพื่อตรวจจับรูปแอปเปิ้ล |
| ผู้วิจัย         | ศิริชัย โชติชาติมาลา                                                               |
| ปริญญา           | วิทยาศาสตร์มหาบัณฑิต                                                               |
| ปีการศึกษา       | 2563                                                                               |
| อาจารย์ที่ปรึกษา | ผู้ช่วยศาสตราจารย์ ดร. นุรีย์ วิวัฒน์วัฒนา                                         |

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้ของเครื่องเพื่อตรวจจับรูปแอปเปิ้ล โดยใช้เทคนิคการตรวจสอบสิ่งใหม่ (Novelty Detection) คือเทคนิคที่มีชุดข้อมูลฝึกฝนซึ่งคือรูปแอปเปิ้ลเพียงชนิดเดียวทั้งหมด และต้องการตรวจจับข้อมูลใหม่หรือข้อมูลผิดปกติในชุดข้อมูลทดสอบ ที่อยู่นอกเหนือจากชุดข้อมูลปกติซึ่งก็คือการทำนายว่าข้อมูลใหม่นั้นเป็นหรือไม่เป็นรูปแอปเปิ้ล โดยในงานวิจัยนี้ใช้แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และแบบจำลองป่าแยก (Isolation Forest) ซึ่งทั้ง 2 แบบจำลองเป็นเทคนิคสำหรับการสร้างแบบจำลองแบบกึ่งมีผู้สอนหรือไม่มีผู้สอน ซึ่งสามารถนำแบบจำลองที่ได้ไปประยุกต์ใช้กับการคัดแยกผลไม้แอปเปิ้ลได้ต่อไป ผู้วิจัยกำหนดให้มีการวัดประสิทธิภาพของแบบจำลอง 4 วิธี ประกอบด้วย Confusion Matrix, ค่าความถูกต้อง (Accuracy), F1-score macro average และพื้นที่ใต้โค้ง (ROC AUC) หลังจากทำการหาค่าพารามิเตอร์ที่ดีที่สุดแล้ว แม้ว่าผลการทดลองชี้ว่าประสิทธิภาพการทำนายมีความใกล้เคียงกัน แต่แบบจำลองป่าแยกดีกว่าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว ในแง่ของความแม่นยำการทำนายรูปภาพแอปเปิ้ล

คำสำคัญ : Support Vector Machine, Isolation Forest, Novelty Detection, แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว, แบบจำลองป่าแยก, เทคนิคการตรวจจับสิ่งใหม่

|                |                                                                                     |
|----------------|-------------------------------------------------------------------------------------|
| Title          | MACHINE LEARNING-BASED ONE-CLASS FRUIT<br>CLASSIFICATION MODELS<br>FOR APPLE IMAGES |
| Author         | SIRICHA CHOTICHATMALA                                                               |
| Degree         | MASTER OF SCIENCE                                                                   |
| Academic Year  | 2020                                                                                |
| Thesis Advisor | Assistant Professor Nuwee Wiwatwattana , Ph.D.                                      |

The objective of this research is to create models for one-class apple classification using two novelty detection algorithms. The novelty detection learns from a training data set comprising only apple images; while able to detect whether the test data are apple or not. The One-Class Support Vector Machine (OCSVM) and the Isolation Forest algorithms, both of which are unsupervised/semi-supervised learning methods, were employed in this research. The research compared model performance with four metrics, i.e., Confusion Matrix, Accuracy, F1-score macro average, and ROC AUC. After tuning for the best hyperparameter values, both methods were comparatively similar. However, the Isolation forest model performed better than the One-Class Support Vector Machine model in the aspect of predicting the apple images correctly.

Keyword : Support Vector Machine, Isolation Forest, Novelty Detection

## กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จลงได้ด้วยความช่วยเหลือจาก ผศ.ดร.นุรีย์ วิวัฒน์วัฒนา อาจารย์ที่ปรึกษาที่ให้คำปรึกษา คำแนะนำในการทำสารนิพนธ์ ตลอดจนสนับสนุนข้อมูลทางวิชาการและข้อมูลสำหรับทำสารนิพนธ์นี้

ขอกราบขอบพระคุณคณะกรรมการสอบสารนิพนธ์ที่ได้ให้คำแนะนำและข้อเสนอแนะสำหรับการปรับปรุงสารนิพนธ์

ศิริชัย โชติชาติมาลา



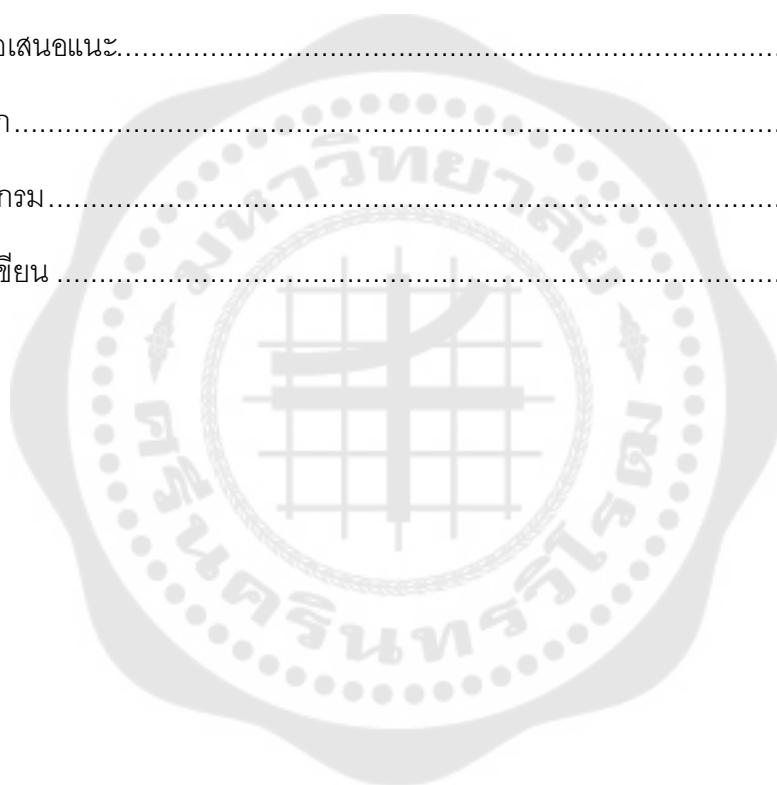
## สารบัญ

### หน้า

|                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------|----|
| บทคัดย่อภาษาไทย.....                                                                                | ง  |
| บทคัดย่อภาษาอังกฤษ .....                                                                            | จ  |
| กิตติกรรมประกาศ .....                                                                               | ฉ  |
| สารบัญ .....                                                                                        | ช  |
| สารบัญตาราง.....                                                                                    | ญ  |
| สารบัญรูปภาพ .....                                                                                  | ฎ  |
| บทที่ 1 บทนำ.....                                                                                   | 1  |
| 1.ความสำคัญและความเป็นมาของงานวิจัย .....                                                           | 1  |
| 2 วัตถุประสงค์ .....                                                                                | 3  |
| 3 ขอบเขตของงานวิจัย.....                                                                            | 4  |
| 4.ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย .....                                                        | 5  |
| บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....                                                       | 6  |
| 1.ทฤษฎีเกี่ยวกับอัลกอริทึม.....                                                                     | 6  |
| 1.1 แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine)..... | 6  |
| 1.2 แบบจำลองป่าแยก (Isolation Forest).....                                                          | 8  |
| 1.3 โครงข่ายประสาทเทียมแบบลึก 50 ชั้น (ResNet50).....                                               | 9  |
| 1.4 การวิเคราะห์องค์ประกอบหลัก (Principles Component Analysis: PCA).....                            | 9  |
| 1.5 การวัดประสิทธิภาพของแต่ละแบบจำลอง.....                                                          | 10 |
| 2.งานวิจัยที่เกี่ยวข้อง.....                                                                        | 13 |
| 2.1 บทความวิจัยเรื่อง Pure-CNN: A Framework for Fruit Images Classification.....                    | 13 |

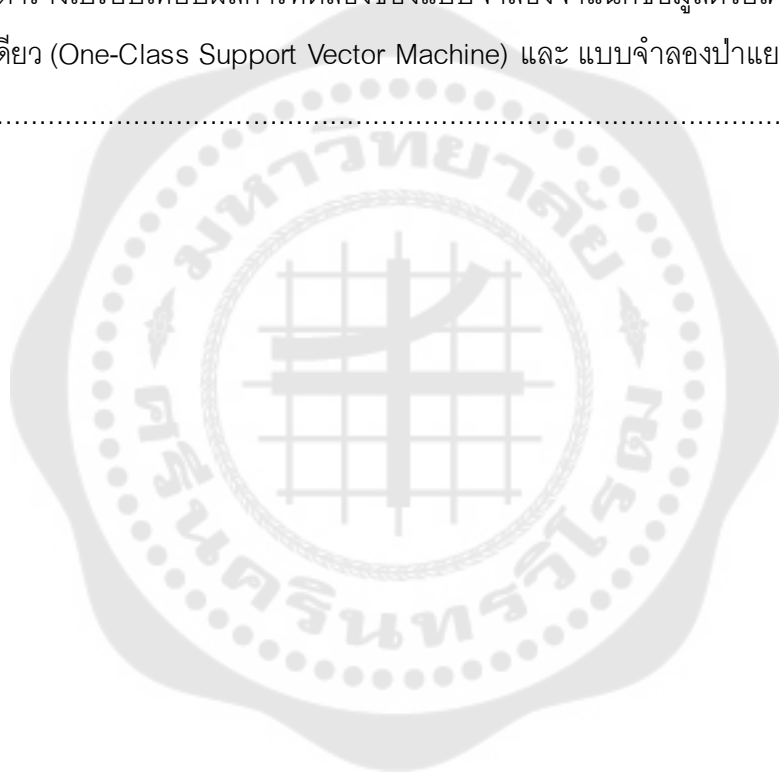
|                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------|----|
| 2.2 บทความวิจัยเรื่อง A Comparative Analysis on Fruit Freshness Classification...                                              | 13 |
| 2.3 บทความวิจัยเรื่อง Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks .....  | 14 |
| 2.4 บทความวิจัย Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool.....          | 14 |
| 2.5 บทความวิจัยเรื่อง Fruit Classification Based on Six Layer Convolutional Neural Network.....                                | 14 |
| 2.6 บทความวิจัย Image-based Processing for Ripeness Classification of Oil Palm Fruit .....                                     | 15 |
| 2.7 บทความวิจัย Research on Fruit Category Classification Based on Convolution Neural Network and Data Augmentation .....      | 15 |
| 2.8 บทความวิจัย Classification on Passion Fruit's Ripeness using K-Means Clustering and Artificial Neural Network .....        | 15 |
| 2.9 บทความวิจัย Fruit Classification using Statistical Features in SVM Classifier ..                                           | 16 |
| 2.10 บทความวิจัย Snake Fruit Classification by Using Histogram of Oriented Gradient Feature and Extreme Learning Machine ..... | 16 |
| 2.11 บทความวิจัย Deep Transfer Learning for Multiple Class Novelty Detection ..                                                | 16 |
| 2.12 บทความวิจัย Open-Set Bottle Classifying using a Convolution Neural Network .....                                          | 17 |
| 2.13 บทความวิจัย Fault Detection of Hydroelectric Generators using Isolation Forest.....                                       | 17 |
| 2.14 บทความวิจัย Weed Classification Using One Class Support Vector Machine .....                                              | 18 |
| บทที่ 3 วิธีดำเนินการวิจัย.....                                                                                                | 19 |
| 3.1 กระบวนการสร้างแบบจำลอง .....                                                                                               | 19 |

|                                                           |    |
|-----------------------------------------------------------|----|
| 3.2 การสำรวจข้อมูล (Exploratory Data Analysis: EDA) ..... | 20 |
| 3.3 การเตรียมข้อมูล (Data Preprocessing) .....            | 38 |
| 3.4 การสร้างแบบจำลอง (Model Creation) .....               | 40 |
| บทที่ 4 การทดลอง.....                                     | 41 |
| บทที่ 5 สรุปผลการทดลอง.....                               | 48 |
| 5.1 สรุปผลการวิจัย.....                                   | 48 |
| 5.2 ข้อเสนอแนะ.....                                       | 50 |
| ภาคผนวก.....                                              | 51 |
| บรรณานุกรม.....                                           | 55 |
| ประวัติผู้เขียน .....                                     | 59 |



## สารบัญตาราง

|                                                                                                                                                                                   | หน้า |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| ตาราง 1 แสดงจำนวนข้อมูลผลไม้ทั้งหมดใน Dataset Fruit-360 ของศฯ .....                                                                                                               | 20   |
| ตาราง 2 ชุดข้อมูลที่นำมาใช้ในงานวิจัย.....                                                                                                                                        | 35   |
| ตาราง 3 กำหนดไฮเปอร์พารามิเตอร์ของ ResNet50.....                                                                                                                                  | 38   |
| ตาราง 4 ตารางเปรียบเทียบผลการทดลองของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยัน<br>ประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation<br>Forest)..... | 43   |



## สารบัญรูปภาพ

### หน้า

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ภาพประกอบ 1 ผลไม้ที่อยู่บนชั้นวางประกอบด้วยแอปเปิ้ลหลากหลายชนิด(Lehnert, 2015).....                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 2  |
| ภาพประกอบ 2 ผลไม้ที่ประกอบด้วยแอปเปิ้ลและไม่ใช่แอปเปิ้ล(Disease, 2016).....                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 2  |
| ภาพประกอบ 3 เส้นขอบแสดงการทำงานของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยัน<br>เส้นขอบแสดงการทำงานของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว<br>(One-Class Support Vector Machine) จัดเป็นการเรียนรู้แบบไม่ได้รับคำตอบ (Unsupervised<br>Learning) คือการทำระบุในชุดข้อมูลทดสอบ (Test data) ว่าค่าใดที่เหมือนกับข้อมูลชุดฝึกฝน<br>(Training data) ค่าจะระบุในวงรี ส่วนค่าใดที่ไม่เหมือนกับข้อมูลชุดฝึกฝน (Training data) จะอยู่<br>นอกวงรีเรียกว่าข้อมูลใหม่ เรียกว่ากระบวนการตรวจจับข้อมูลใหม่ (Novelty Detection)..... | 7  |
| ภาพประกอบ 4 แสดงถึงการทำงานของแบบจำลองป่าแยก (Isolation Forest) ประกอบด้วย ราก<br>(Root Node) คือโหนดตั้งต้นที่อยู่บนสุด , โหนดภายใน (Internal Node) คือการแยกย่อยของ<br>ต้นไม้ออกไป และ โหนดที่เป็นใบ (Leaf Node) คือโหนดที่มีดีกรีเป็นศูนย์ (Zhang, Kang, Li,<br>2019) .....                                                                                                                                                                                                                                                               | 8  |
| ภาพประกอบ 5 ตาราง Confusion matrix เพื่อวัดประสิทธิภาพของแบบจำลองจำแนกข้อมูลด้วย<br>เครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine : SVM) และ<br>แบบจำลองป่าแยก (Isolation Forest) ในงานวิจัยนี้ผลลัพธ์เป็นจำนวนตัวเลขทั้งหมดเพื่อ<br>การวัดผลอย่างชัดเจน .....                                                                                                                                                                                                                                                          | 11 |
| ภาพประกอบ 6 เส้นกราฟ ROC จะมีพื้นที่ใต้กราฟเรียกว่า AUC (Area Under Curve) เป็นวิธีวัด<br>ประสิทธิภาพของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class<br>Support Vector Machine : SVM) และแบบจำลองป่าแยก (Isolation Forest) ยิ่งค่าพื้นที่ใต้กราฟ<br>มาก แปลว่าแบบจำลองมีประสิทธิภาพมาก ในทางกลับกันถ้าเส้นกราฟเป็นเส้นตรงทแยงมุม<br>หมายความว่าแบบจำลองไม่มีประสิทธิภาพ (Prati, Monard, 2008) .....                                                                                                                    | 12 |
| ภาพประกอบ 7 แผนผังกระบวนการสร้างแบบจำลอง.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 19 |
| ภาพประกอบ 8 ตัวอย่างข้อมูลรูปภาพชุดข้อมูล Fruit-360 ที่เป็นแอปเปิ้ล .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 37 |
| ภาพประกอบ 9 ตัวอย่างข้อมูลรูปภาพชุดข้อมูล Fruit-360 ที่ไม่ใช่แอปเปิ้ล .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 37 |

|                                                                                                                                                                                                                                  |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ภาพประกอบ 10 Confusion matrix ของ One-Class support vector machine โดยกำหนดค่า 0 คือ ไม่ใช่รูปผลไม้แอปเปิ้ล และค่า 1 คือรูปผลไม้แอปเปิ้ล.....                                                                                    | 41 |
| ภาพประกอบ 11 Confusion matrix ของ Isolation Forest โดยกำหนดค่า 0 คือ ไม่ใช่รูปผลไม้แอปเปิ้ล และค่า 1 คือรูปผลไม้แอปเปิ้ล.....                                                                                                    | 42 |
| ภาพประกอบ 12 True Positive คือเป็นภาพผลไม้แอปเปิ้ล และทำนายว่าเป็นภาพแอปเปิ้ลจริง                                                                                                                                                | 43 |
| ภาพประกอบ 13 True Negative คือไม่ใช่ภาพผลไม้แอปเปิ้ล และทำนายว่าไม่เป็นภาพแอปเปิ้ลจริง.....                                                                                                                                      | 44 |
| ภาพประกอบ 14 False Positive คือไม่ใช่ภาพผลไม้แอปเปิ้ล แต่ทำนายว่าเป็นภาพแอปเปิ้ลจริง.....                                                                                                                                        | 44 |
| ภาพประกอบ 15 False Negative คือเป็นภาพผลไม้แอปเปิ้ล แต่ทำนายว่าไม่ใช่ภาพแอปเปิ้ลจริง.....                                                                                                                                        | 45 |
| ภาพประกอบ 16 True Positive คือเป็นภาพผลไม้แอปเปิ้ล และทำนายว่าเป็นภาพแอปเปิ้ลจริง                                                                                                                                                | 45 |
| ภาพประกอบ 17 True Negative คือไม่ใช่ภาพผลไม้แอปเปิ้ล และทำนายว่าไม่เป็นภาพแอปเปิ้ลจริง.....                                                                                                                                      | 46 |
| ภาพประกอบ 18 False Positive คือไม่ใช่ภาพผลไม้แอปเปิ้ล แต่ทำนายว่าเป็นภาพแอปเปิ้ลจริง.....                                                                                                                                        | 46 |
| ภาพประกอบ 19 False Negative คือเป็นภาพผลไม้แอปเปิ้ล แต่ทำนายว่าไม่ใช่ภาพแอปเปิ้ลจริง.....                                                                                                                                        | 47 |
| ภาพประกอบ 20 เปรียบเทียบผลการทดลองของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ซึ่งประสิทธิภาพของทั้ง 2 แบบจำลองใกล้เคียงกันอย่างมาก..... | 49 |

## บทที่ 1

### บทนำ

#### 1. ความสำคัญและความเป็นมาของงานวิจัย

ปัจจุบันการรับรู้และแยกแยะวัตถุต่างๆเกิดขึ้นจากการรับรู้ภาพต่างๆของสมองของคน ซึ่งในวงการอาหารและผลไม้ก็เช่นเดียวกัน ในอุตสาหกรรมผลไม้เป็นอุตสาหกรรมหลักของโลกและของประเทศไทย โดยข้อมูลการส่งออกปี 2561 และ 2562 ที่ 2,959.37 ล้านเหรียญสหรัฐฯ และ 3,953.62 ล้านเหรียญสหรัฐฯ ตามลำดับ ผลไม้มีหลายชนิดและหลากหลายสายพันธุ์ให้เลือกซื้อในซูเปอร์มาร์เก็ตตามภาพประกอบ 1 ซึ่งปัญหาของการคัดแยกผลไม้ชนิดต่างๆเกิดขึ้นเนื่องจาก ขนาดที่ต่างกัน, สีที่ต่างกัน, เนื้อผลไม้ที่ต่างกัน หรือความเหมือนกันทั้งขนาด, สีและเนื้อผลไม้ และการเกิดปัญหาอย่างน้อย 2 ครั้งคือเมื่อมีการคัดแยกผลไม้คัดแยกสายพันธุ์เพื่อนำขึ้นชั้นวาง และเมื่อลูกค้าหยิบแล้วคืนในชั้นวางที่ไม่ใช่ชนิดของผลไม้ นั่นทำให้ในชั้นวางมีทั้งผลไม้แอปเปิ้ลและไม่ใช่ผลไม้แอปเปิ้ล ตามภาพประกอบ 2 ซึ่งปัญหาตรงนี้ต้องใช้ทรัพยากรในแต่ละวันค่อนข้างมาก ทั้งนี้การใช้ทรัพยากรคนอาจจะก่อให้เกิดความผิดพลาดในการแยกประเภทของผลไม้เนื่องจาก บางชนิดนั้นมีรูปร่างที่คล้ายกันแต่ไม่ใช่ชนิดเดียวกัน ยกตัวอย่างเช่น ผลไม้สาลี่กับแอปเปิ้ลเขียว เป็นต้น จากปัญหาข้างต้นปัจจุบันหลายๆที่ยังไม่ได้รับการแก้ไข สวนทางกับการบริโภคผลไม้ของคนที่มีมากขึ้นทุกวัน ทำให้ความสำคัญของการสร้างวิธีแยกชนิดของผลไม้โดยลดทรัพยากรคน ด้วยวิธีการเรียนรู้ของเครื่องจักร (Machine Learning) ซึ่งข้อดีคือเป็นระบบที่สามารถเรียนรู้ได้ด้วยตนเองโดยเรียนรู้จากการใช้ข้อมูลเพียงอย่างเดียวเพื่อให้ค่าความถูกต้องของผลลัพธ์ที่แม่นยำได้ ซึ่งวิธีการเรียนรู้ของเครื่องจักร (Machine Learning) นี้ประกอบด้วยข้อมูลและเครื่องมือทางสถิติ ซึ่งตัวอย่างที่เราเห็นในปัจจุบันเช่นการจัดประเภท (Classification) สำหรับสมาชิกของ Netflix ว่ามีความสนใจหนังประเภทไหน, การตรวจจับการคอร์รัปชันในองค์กร, การหากลุ่มธุรกรรมทางการเงิน (Transaction) ที่ผิดปกติอันนำไปสู่การทุจริตฟอกเงิน รวมถึงการคัดแยกผลไม้ที่ปะปนกัน ซึ่งในการคัดแยกผลไม้แอปเปิ้ลในงานวิจัยนี้จะใช้เทคนิคที่เรียกว่าการตรวจจับข้อมูลใหม่ (Novelty Detection) โดยอนุมานว่ามีหุ่นยนต์คอยคัดแยกผลไม้ โดยที่หุ่นยนต์จะตรวจสอบว่ารูปภาพผลไม้ชนิดใดมีแนวโน้มที่จะเป็นแอปเปิ้ล หรือรูปภาพใดมีแนวโน้มที่จะไม่ใช่แอปเปิ้ล

การตรวจจับข้อมูลใหม่ (Novelty Detection) คือการนำรูปภาพเพียงประเภทเดียว มาใช้ในการฝึกฝนข้อมูลในที่นี้คือผลไม้แอปเปิ้ลเพียงชนิดเดียวเพื่อการจัดประเภท (Classification) หลังจาก

แบบจำลองเรียนรู้ว่ารูปภาพแอปเปิ้ลเป็นอย่างไร หลังจากนั้นทำการทดสอบโดยนำรูปภาพทั้งแอปเปิ้ลและไม่ใช่แอปเปิ้ลทดสอบกับแบบจำลองที่สร้างขึ้น เพื่อตรวจสอบค่าความถูกต้องของการตรวจจับ



ภาพประกอบ 1 ผลไม้ที่อยู่บนชั้นวางประกอบด้วยแอปเปิ้ลหลากหลายชนิด (Lehnert, 2015)



ภาพประกอบ 2 ผลไม้ที่ประกอบด้วยแอปเปิ้ลและไม่ใช่แอปเปิ้ล (Disease, 2016)

## 2 วัตถุประสงค์

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งความมุ่งหมายไว้ดังนี้

1. เพื่อสร้างแบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้ของเครื่อง เพื่อตรวจจับรูปแอปเปิ้ล โดยใช้เทคนิคการตรวจจับข้อมูลใหม่ (Novelty Detection) คือเทคนิคที่มีชุดข้อมูลฝึกฝน (Training Data) ซึ่งคือรูปแอปเปิ้ลเพียงชนิดเดียวในทุกสายพันธุ์ของข้อมูล ทำการตรวจจับข้อมูลใหม่หรือข้อมูลผิดปกติในชุดข้อมูลทดสอบ (Test data) ที่ประกอบด้วยรูปภาพแอปเปิ้ลและรูปที่ไม่ใช่แอปเปิ้ล โดยในงานวิจัยนี้ใช้แบบจำลองการคัดแยกผลไม้แอปเปิ้ลแบบหนึ่งชนิดด้วย แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest)

2. เพื่อเปรียบเทียบวัดประสิทธิภาพของแบบจำลองทั้ง 2 แบบที่สร้างจากการใช้ข้อมูลรูปภาพแอปเปิ้ลเพียงอย่างเดียว กับรูปภาพผลไม้ที่ไม่ใช่แอปเปิ้ล ด้วย Confusion Matrix, F1-Score, Accuracy และ ROC AUC

### 3 ขอบเขตของงานวิจัย

ภาพที่นำมาใช้ในงานวิจัยคือเป็นภาพที่ใช้การถ่ายทำผลไม้ในขณะที่กำลังหมุน ภาพแล้วทำการแยกเฟรม โดยฉากหลังของผลไม้เป็นสีขาว ขนาดของภาพอยู่ในขนาด 100 x 100 pixels โดยเป็นชุดข้อมูลที่นำมาบางส่วนจากชุดข้อมูลภาพผลไม้ 360 องศา (Fruit-360) โดยได้รับลิขสิทธิ์จาก (c) 2017-2020 Mihai Oltean, Horea Muresan ประกอบด้วยภาพผักและผลไม้จำนวน 90,380 ภาพ ผ่านการจัดประเภท 131 ชนิด โดยทางผู้วิจัยได้จัดแบ่งเป็นแอปเปิ้ลและไม่ใช่อแอปเปิ้ลเท่านั้น จากตารางที่ 1 แสดงให้เห็นถึงจำนวนข้อมูลผลไม้ที่ประกอบด้วยแอปเปิ้ลและอื่นๆที่ไม่ใช่อแอปเปิ้ล เช่น ส้ม, มังคุด, องุ่น เป็นต้น ซึ่งมีขนาดทั้งหมด 127 MB จะเห็นได้ว่าผู้วิจัยได้ทำการดึงรูปผลไม้ในข้อมูล Training ที่ไม่ใช่แอปเปิ้ลออก เพื่อสร้างสถานการณ์ของการสร้างแบบจำลองที่ตรวจสอบสิ่งที่ไม่ใช่แอปเปิ้ล โดยฝึกฝนเฉพาะภาพแอปเปิ้ลจึงนำมาซึ่งหัวข้อการแก้ปัญหาแบบ โดยใช้เทคนิคการตรวจจับข้อมูลใหม่ (Novelty Detection) อัลกอริทึมที่เลือกใช้ในการวิจัยนี้ประกอบด้วยโครงข่ายประสาทเทียมแบบลึกที่เรียกว่า ResNet50 ใช้สำหรับดึงคุณลักษณะเด่นของรูปภาพออกมา ถัดมาคือ Standard Scaler เมื่อการทำ ResNet50 ส่งผลให้มิติที่ออกมาสูงขึ้น หน้าทีของ Standard Scaler คือทำให้ภาพที่ออกมาอยู่ในมาตรฐานเดียวกัน สุดท้ายคือการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis : PCA) เพื่อช่วยในการลดมิติของข้อมูลลง และเตรียมนำเข้าสู่แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ต่อไปการวัดประสิทธิภาพและประสิทธิผลของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ประกอบด้วย 4 ประเภท ประกอบด้วย Confusion Matrix, F1-Score, Accuracy และ ROC AUC

#### 4. ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. แบบจำลองสำหรับเพื่อคัดแยกผลไม้แอปเปิ้ลและไม่ใช่แอปเปิ้ลที่ปนกันโดยใช้แบบจำลอง ที่พัฒนาขึ้นโดยเทคนิคที่มีประสิทธิภาพ
2. เพิ่มประสิทธิภาพและประสิทธิผลการคัดแยกผลไม้แอปเปิ้ลได้ในทันที
3. แบบจำลองอาจนำมาซึ่งการลดต้นทุนค่าใช้จ่ายทรัพยากรคนเพื่อการแยกประเภทผลไม้ระหว่างแอปเปิ้ลและไม่ใช่แอปเปิ้ล
4. สามารถนำแบบจำลองไปประยุกต์ใช้กับปัญหาอื่นๆในการแยกวัตถุที่มีรูปร่าง, สี หรือขนาดใกล้เคียงกันได้



## บทที่ 2

### วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยนี้เลือกใช้แบบจำลอง 2 แบบคือแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine : SVM) และแบบจำลองป่าแยก (Isolation Forest) เนื่องจาก งานวิจัยนี้เป็นงานวิจัยสำหรับตรวจจับข้อมูลใหม่ (Novelty Detection) ซึ่งทั้ง 2 แบบจำลองนี้เป็นแบบจำลองที่สอดคล้องกับหลักการของการตรวจจับข้อมูลใหม่ (Novelty Detection) และได้นำเสนอหัวข้อดังนี้

#### 1. ทฤษฎีเกี่ยวกับอัลกอริทึม

การเรียนรู้ของเครื่อง (Machine Learning) คืออัลกอริทึมที่มีพัฒนาการเรียนรู้ข้อมูลและทำนายข้อมูลได้ ซึ่งเป็นส่วนหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งอัลกอริทึมจะทำงานโดยอาศัยสร้างแบบจำลองทางคณิตศาสตร์จากข้อมูลฝึกฝน (Training data) เพื่อนำแบบจำลองมาทดสอบกับข้อมูลทดสอบ (Test data) และใช้ในการคาดการณ์หรือตัดสินใจ การเรียนรู้ของเครื่องสามารถแบ่งได้หลายประเภทดังนี้ การเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning), การเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning), การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) และ การเรียนรู้แบบกึ่งมีผู้สอน (Semi Supervised Learning) เทคนิคการตรวจจับข้อมูลใหม่ (Novelty Detection) คือการที่ชุดข้อมูลฝึกฝน (Training data) จะมีข้อมูลเพียงชนิดเดียว เช่นในงานวิจัยนี้มีเพียงรูปภาพผลไม้แอปเปิ้ลเพียงชนิดเดียว และสนใจที่จะตรวจจับข้อมูลใหม่ที่ไม่ได้อยู่ในชุดฝึกฝน ในที่นี้คือรูปภาพผลไม้ที่ไม่ใช่แอปเปิ้ล ซึ่งแตกต่างจากการตรวจจับค่าผิดปกติ (Outlier Detection) คือชุดข้อมูลฝึกฝน (Training data) จะมีข้อมูลรวมข้อมูลที่แตกต่างเข้าไปด้วย ตัวอย่างเช่นมีรูปภาพผลไม้แอปเปิ้ลและภาพผลไม้อื่นรวมอยู่ด้วย ซึ่งในการทำจะสนใจเพียงข้อมูลที่เป็นปกติ และจะไม่สนใจข้อมูลที่อยู่นอกเหนือจากข้อมูลปกติ ที่เรียกว่าค่าผิดปกติ (Outlier)

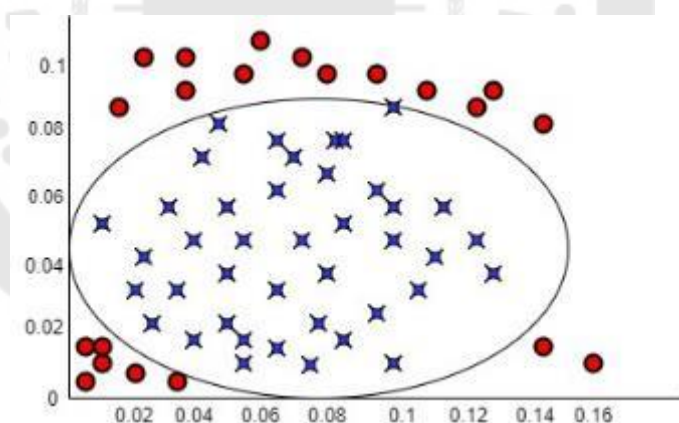
##### 1.1 แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine)

แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine : SVM) เป็นอัลกอริทึมการเรียนรู้ที่แบบ Unsupervised Learning ซึ่งได้รับการฝึกฝนเฉพาะกับข้อมูลผลไม้แอปเปิ้ล คือมี Class Label เพียง 1 Class แบบจำลองจะ

เรียนรู้ขอบเขตของภาพแอปเปิ้ลดังนั้นจึงสามารถจำแนกภาพผลไม้ใด ๆ ที่อยู่นอกขอบเขตตามที่คาดเดาได้ว่าไม่เป็นภาพแอปเปิ้ล จากสมการที่ 1 แสดงหลักการทำงานของ One-Class Support Vector Machine กำหนดให้  $x$  คือ ค่าของรูปภาพที่ส่งเข้าแบบจำลอง และ  $S$  คือค่าความน่าจะเป็นที่จะเป็นรูปภาพที่เป็นผลไม้แอปเปิ้ล กรณีที่ถ้าค่า  $x$  น้อยกว่า  $S$  จะให้ค่า 1 ซึ่งคือเป็นภาพผลไม้แอปเปิ้ล แต่ถ้า  $x$  มากกว่า  $S$  จะให้ค่า 0 คือไม่ใช่ภาพผลไม้แอปเปิ้ล และจากภาพประกอบ 3 เส้นขอบแสดงการทำงานของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) แสดงถึงภาพที่ความน่าจะเป็นของรูปภาพผลไม้แอปเปิ้ลจะแสดงค่าอยู่ในวงรี ส่วนรูปภาพผลไม้ใดที่ไม่ใช่รูปภาพแอปเปิ้ลจะแสดงว่าอยู่นอกวงรี

$$f(x) = \begin{cases} 1, & \text{if } x < S \\ 0, & \text{if } x \geq S \end{cases} \quad (1)$$

จากสมการที่ 1 หลักการทำงานของ One-Class Support Vector Machine

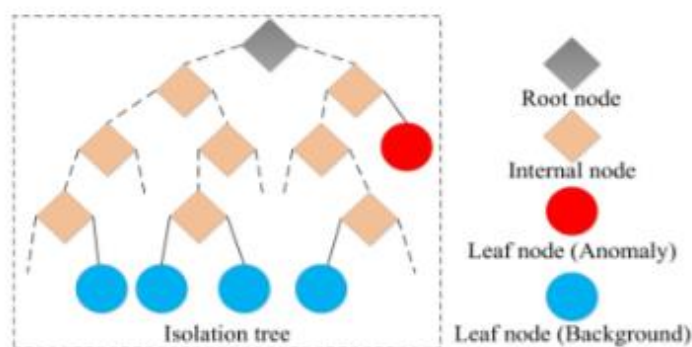


ภาพประกอบ 3 เส้นขอบแสดงการทำงานของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันเส้นขอบแสดงการทำงานของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) จัดเป็นการเรียนรู้แบบไม่ได้รับคำตอบ (Unsupervised Learning) คือการทำระบุในชุดข้อมูลทดสอบ (Test data) ว่าค่าใดที่เหมือนกับข้อมูลชุดฝึกฝน (Training data) ค่าจะระบุในวงรี ส่วนค่าใดที่ไม่เหมือนกับข้อมูลชุดฝึกฝน (Training data) จะอยู่นอกวงรีเรียกว่าข้อมูลใหม่ เรียกว่ากระบวนการตรวจจับข้อมูลใหม่ (Novelty Detection)

## 1.2 แบบจำลองป่าแยก (Isolation Forest)

เป็นอัลกอริทึมการเรียนรู้ที่เป็นแบบ Unsupervised Learning เป็นแบบจำลองที่พัฒนามาจาก Decision Tree แต่ต่างกันเนื่องจาก แบบจำลองป่าแยก (Isolation Forest) คือการเพิ่มจำนวนต้นไม้เป็นหลายๆต้น ซึ่งวิธีการทำงานปกติคือจะแยกค่าที่ผิดปกติโดยการสุ่มเลือกคุณลักษณะ (Feature) ออกมาหลังจากนั้นทำการแยกค่าสูงสุดและค่าต่ำสุดของของคุณลักษณะที่เลือกมา ซึ่งการแบ่งคุณลักษณะแบบสุ่มนี้จะเป็นการสร้างกิ่งของต้นไม้ หรือก็คือข้อมูลค่าที่ผิดปกติออกจากข้อมูลที่ปกติ โดยจะสังเกตได้ว่าข้อมูลที่ผิดปกติจะอยู่ในชั้นต้นๆของต้นไม้ หรือว่า Leaf Node (Anomaly) ตามภาพประกอบ 4 ส่งผลให้ทำให้ประสิทธิภาพการทำงานของแบบจำลองป่าแยกสูงขึ้น ซึ่งเป็นเทคนิคที่เหมาะสมกับการตรวจจับข้อมูลใหม่โดยยึดจากชุดข้อมูลฝึกฝน (Training data) เป็นหลัก ข้อดีของเป้าหมายการทำงานของแบบจำลองป่าแยกคือ การลดค่าความเกี่ยวข้องกัน (Correlation) ระหว่างกันและยิ่งแต่ละต้นไม้ (Single tree) ยิ่งมากแบบจำลองยิ่งมีประสิทธิภาพ นอกจากนี้การใช้ขนาดของหน่วยความจำน้อยกว่าอัลกอริทึมแบบอื่นในการตรวจจับข้อมูลใหม่ ดังภาพประกอบ 4

ซึ่งพารามิเตอร์พื้นฐานต่างๆในงานวิจัยกำหนดดังนี้ Contamination เป็นตัวกำหนดสัดส่วนการทำนายของสิ่งที่ไม่ใช่แอปเปิ้ลบนชุดข้อมูล, Max\_features จำนวนสูงสุดที่จะดึงคุณสมบัติมาฝึกฝน ต้นไม้แต่ละต้น หากค่ามากจะไปเพิ่ม ค่าความเกี่ยวข้องกัน (Correlation) , Max\_samples คือการระบุจำนวนของชุดข้อมูลที่จะสุ่มขึ้นมาเพื่อใช้ในการฝึกฝนแบบจำลอง และ N\_estimators คือการระบุจำนวนต้นไม้ที่จะสร้าง



ภาพประกอบ 4 แสดงถึงการทำงานของแบบจำลองป่าแยก (Isolation Forest) ประกอบด้วย ราก (Root Node) คือโหนดตั้งต้นที่อยู่บนสุด , โหนดภายใน (Internal Node) คือ

การแยกย่อยของต้นไม้ออกไป และ โหนดที่เป็นใบ (Leaf Node) คือโหนดที่มีดีกรีเป็นศูนย์ (Zhang, Kang, Li, 2019)

### 1.3 โครงข่ายประสาทเทียมแบบลึก 50 ชั้น (ResNet50)

สถาปัตยกรรมโครงข่ายประสาทเทียมแบบลึก 50 ชั้น (ResNet50) เป็นอัลกอริทึมที่ผ่านการฝึกฝนจากชุดข้อมูลมากกว่าล้านภาพในชุดข้อมูล ImageNet โดยจะให้ค่าน้ำหนัก (Weight) และค่าอคติ (Bias) พร้อมใช้งานแล้ว ซึ่งในงานวิจัยนี้ใช้สำหรับการดึงคุณลักษณะจากภาพออกมาซึ่งจะทำให้มิติมีขนาดสูงขึ้น ซึ่งทำให้ต้องทำ Standard Scaler เพื่อให้ภาพที่ออกมาจาก ResNet50 อยู่ในมาตรฐานเดียวกัน โดยองค์ประกอบของสถาปัตยกรรมโครงข่ายประสาทเทียมแบบลึก 50 ชั้น (ResNet50) ประกอบด้วย 50 เลเยอร์ ดังนี้

ชั้นที่ติดกับนำข้อมูลเข้ามีขนาด Kernel เท่ากับ  $7 \times 7$  และ Kernel ที่ต่างกัน 64 แบบ ได้ทั้งหมด 1 เลเยอร์

ชั้นถัดมาประกอบด้วย Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 64 แบบ, Kernel เท่ากับ  $3 \times 3$  และ Kernel ที่ต่างกัน 64 แบบ และ Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 256 แบบ ซึ่งในเลเยอร์นี้กำหนดให้ทำซ้ำ 3 ครั้ง จะทำให้ได้ทั้งหมด 9 เลเยอร์

ชั้นถัดมาประกอบด้วย Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 128 แบบ, Kernel เท่ากับ  $3 \times 3$  และ Kernel ที่ต่างกัน 128 แบบ และ Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 512 แบบ ซึ่งในเลเยอร์นี้กำหนดให้ทำซ้ำ 4 ครั้ง จะทำให้ได้ทั้งหมด 12 เลเยอร์

ชั้นถัดมาประกอบด้วย Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 256 แบบ, Kernel เท่ากับ  $3 \times 3$  และ Kernel ที่ต่างกัน 256 แบบ และ Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 1,024 แบบ ซึ่งในเลเยอร์นี้กำหนดให้ทำซ้ำ 6 ครั้ง จะทำให้ได้ทั้งหมด 18 เลเยอร์

ชั้นถัดมาประกอบด้วย Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 512 แบบ, Kernel เท่ากับ  $3 \times 3$  และ Kernel ที่ต่างกัน 512 แบบ และ Kernel เท่ากับ  $1 \times 1$  และ Kernel ที่ต่างกัน 2,048 แบบ ซึ่งในเลเยอร์นี้กำหนดให้ทำซ้ำ 3 ครั้ง จะทำให้ได้ทั้งหมด 9 เลเยอร์

ชั้นสุดท้ายคือการทำนายผล (Fully Connected Layer) ประกอบด้วย 1,000 Node และ Activation Function กำหนดเป็น Softmax นับได้ 1 เลเยอร์

### 1.4 การวิเคราะห์องค์ประกอบหลัก (Principles Component Analysis: PCA)

งานวิจัยนี้ได้ใช้การวิเคราะห์องค์ประกอบหลัก (PCA) เข้ามาช่วยในงานวิจัยซึ่งนำมัลติมิติของภาพที่เกิดจากการทำการดึงคุณลักษณะ (Feature Extraction) ของภาพออกมาโดยหลักการทำงานคือ เป็นวิธีที่ใช้วิเคราะห์ข้อมูลหลายตัวแปรเพื่อหาความสัมพันธ์ของตัวแปรทำ

ให้เกิดการลดขนาดลง ลดความซับซ้อนและนำมาประยุกต์ใช้การลดขนาดคุณลักษณะ (Feature) ซึ่งจะไม่ส่งผลกระทบต่อข้อมูลแต่วิธีนี้ทำให้ข้อมูลง่ายต่อการนำไปประยุกต์ใช้และนำเข้าแบบจำลองต่อไป โดยการทำงานของวิธีการวิเคราะห์องค์ประกอบหลัก (PCA) จะทำการสร้างตัวแปรที่เรียกว่าองค์ประกอบ (Component) ซึ่งแต่ละองค์ประกอบจะไม่สัมพันธ์กันและจะแสดงค่าความผันแปร (Variance) โดยองค์ประกอบที่จะนำมาใช้จะเลือกใช้ที่ครอบคลุมกับความผันแปรมากที่สุด

### 1.5 การวัดประสิทธิภาพของแต่ละแบบจำลอง

งานวิจัยนี้ได้วัดประสิทธิภาพของแบบจำลอง 4 แบบประกอบด้วย Confusion Matrix, F1-Score, Accuracy และ ROC AUC

Confusion matrix คือตารางสำคัญในการวัดความสามารถของแบบจำลองในการแก้ปัญหา Classification จากภาพประกอบ 5 ประกอบด้วย

True Positive (TP) คือภาพที่ทำนาย (Prediction) ตรงกับสิ่งที่เกิดขึ้นจริง (Actual) ก็คือทำนายว่าจริงและสิ่งที่เกิดขึ้นก็คือจริง เช่น เป็นรูปภาพแอปเปิ้ลจริง และแบบจำลองว่าทำนายว่าเป็นแอปเปิ้ลจริง

True Negative (TN) คือภาพที่ทำนาย (Prediction) ตรงกับสิ่งที่เกิดขึ้นจริง (Actual) ก็คือทำนายว่าไม่จริงและสิ่งที่เกิดขึ้นก็คือไม่จริง เช่น ไม่ใช่ภาพผลไม้แอปเปิ้ล และแบบจำลองทำนายว่าไม่เป็นภาพผลไม้แอปเปิ้ล

False Positive (FP) คือการทำนาย (Prediction) ไม่ตรงกับสิ่งที่เกิดขึ้นจริง (Actual) ก็คือทำนายว่าจริง แต่สิ่งที่เกิดขึ้นก็คือไม่จริง เช่น ไม่ใช่ภาพผลไม้แอปเปิ้ล แต่แบบจำลองทำนายว่าเป็นภาพผลไม้แอปเปิ้ล จัดอยู่ในกลุ่มของ True Negative.

False Negative (FN) คือการทำนาย (Prediction) แล้วไม่ตรงกับสิ่งที่เกิดขึ้นจริง (Actual) ก็คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้นก็คือจริง เช่น เป็นภาพแอปเปิ้ลจริง แต่แบบจำลองทำนายว่าไม่เป็นภาพผลไม้แอปเปิ้ล จัดอยู่ในกลุ่มของ True Positive.

|                  |              | Actual Values       |                     |
|------------------|--------------|---------------------|---------------------|
|                  |              | Positive (1)        | Negative(0)         |
| Predicted Values | Positive (1) | True Positive (TP)  | False Positive (FP) |
|                  | Negative(0)  | False Negative (FN) | True Negative (TN)  |

ภาพประกอบ 5 ตาราง Confusion matrix เพื่อวัดประสิทธิภาพของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine : SVM) และแบบจำลองป่าแยก (Isolation Forest) ในงานวิจัยนี้ผลลัพธ์เป็นจำนวนตัวเลขทั้งเลขทั้งหมดเพื่อการวัดผลอย่างชัดเจน

การประเมินผลด้วยค่าความถูกต้อง (Accuracy) เป็นค่าความถูกต้องที่ได้จากนำชุดข้อมูลทดสอบเข้าแบบจำลองทั้ง 2 แบบ ดังสมการนี้

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + RN)} \quad (2)$$

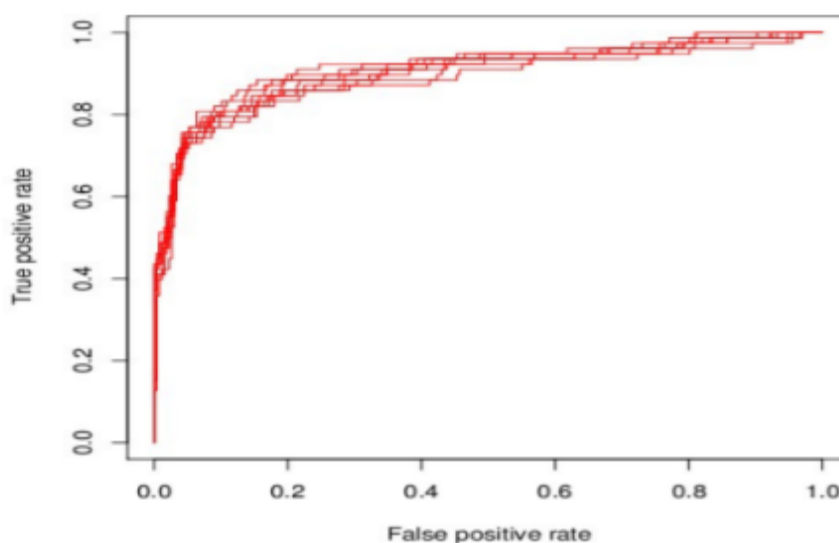
การประเมินผลด้วยค่า F1-score แบบ Macro Average เกิดจากการนำค่าของ Precision คือ จำนวนที่แบบจำลองสามารถทำนายได้ภาพ(ไม่ใช่)รูปแอปเปิ้ลถูกต้องหารด้วยจำนวนที่แบบจำลองสามารถทำนายว่าได้ภาพ(ไม่ใช่)รูปแอปเปิ้ลทั้งหมด และ Recall คือจำนวนที่แบบจำลองสามารถทำนายได้ภาพ(ไม่ใช่)รูปแอปเปิ้ลถูกต้องหารด้วยจำนวนที่เป็นรูป(ไม่ใช่)แอปเปิ้ลทั้งหมด โดยนำมาเฉลี่ยทั้ง 2 Class (ใช่และไม่ใช่แอปเปิ้ล) ซึ่งค่า F1-score แบบ Macro Average สามารถบอกถึงความสามารถแบบจำลองในแง่ที่แอปเปิ้ลเป็นคลาสที่สนใจ กับแ่งที่แอปเปิ้ลเป็นคลาสที่ไม่สนใจ

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Accuracy = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (5)$$

วัดประสิทธิภาพค่า ROC AUC คือ พื้นที่ใต้กราฟ ระหว่าง False Positive Rate คือแบบจำลองทำนายว่าเป็นผลไม่แอปเปิ้ลแต่ความจริงไม่ใช่ผลไม่แอปเปิ้ล และ True Positive Rate คือแบบจำลองทำนายว่าเป็นรูปภาพแอปเปิ้ลได้ตรงกับรูปภาพแอปเปิ้ลจริงซึ่งค่าโดยรวมของ ROC AUC จะต้องมีความสูงซึ่งจะเป็นตัววัดผลว่าแบบจำลองมีประสิทธิภาพ ดังภาพประกอบ 6



ภาพประกอบ 6 เส้นกราฟ ROC จะมีพื้นที่ใต้กราฟเรียกว่า AUC (Area Under Curve) เป็นวิธีวัดประสิทธิภาพของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-

Class Support Vector Machine : SVM) และแบบจำลองป่าแยก (Isolation Forest) ยิ่งค่าพื้นที่ใต้กราฟมาก แปลว่าแบบจำลองมีประสิทธิภาพมาก ในทางกลับกันถ้าเส้นกราฟเป็นเส้นตรงแยงมุม หมายความว่าแบบจำลองไม่มีประสิทธิภาพ (Prati, Monard, 2008)

## 2.งานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยนี้ได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังนี้

### 2.1 บทความวิจัยเรื่อง Pure-CNN: A Framework for Fruit Images Classification

บทความกล่าวถึงเรื่องการจัดหมวดหมู่ภาพผลไม้โดยใช้ Pure Convolutional Neural Network (PCNN) ด้วย Global Average Pooling (GAP) พบว่าการใช้เลเยอร์ GAP สามารถแก้ปัญหาเรื่องการ overfitting และได้ประเมินความถูกต้องของการจำแนกในชุดข้อมูล Fruit-360 ซึ่งมีรูปภาพผลไม้ 81 หมวดและผลไม้ชนิดเดียวกันได้ค่าความถูกต้องร้อยละ 98.88 เมื่อใช้ PCNN กับชั้น GAP ซึ่งทำให้สามารถจำแนกวัตถุและการจำแนกภาพได้หลายระดับซึ่งดีกว่าการใช้เพียง PCNN เท่านั้น โดยทำนายภาพผลไม้ผิดเพียง 46 รูปภาพเท่านั้นและใช้เวลาคำนวณน้อยที่สุดที่ 20.02 วินาที ในส่วนของ CNN และ Fully Connected Layer ให้ค่าความถูกต้องที่ร้อยละ 97.41 โดยทำนายภาพผิดไป 107 รูปภาพและใช้เวลาคำนวณที่ 20.89 วินาที (Kausar, Sharif, Park, & Shin, 2018)

### 2.2 บทความวิจัยเรื่อง A Comparative Analysis on Fruit Freshness Classification

บทความกล่าวถึงเรื่องความพยายามคัดแยกผลไม้สดและผลไม้เน่าเสียโดยใช้ Support Vector Machine (SVM) มีทั้งหมด 6 Classes ด้วยกันประกอบด้วย ผลไม้สด 3 ประเภท แอปเปิ้ล,กล้วย,ส้ม และผลไม้เสีย 3 ประเภท แอปเปิ้ล,กล้วย,ส้ม โดยใช้การดึงคุณลักษณะ (Feature Extraction) 4 แบบ กับ Histogram Features (Hist), Gray Level Co-Occurrence Matrix (GLCM), Bag of Features(BoF), Convolutional Neural Networks (CNNs) ซึ่งสามารถแยกผลไม้สดกับผลไม้เน่าเสียด้วยอัตราความถูกต้องเฉลี่ยสูงสุดคือ CNNsF ที่ร้อยละ 99.19 ส่วนกรณีที่เป็นกรเปรียบเทียบโดยแยกผลไม้สดและผลไม้เน่าเสียแต่ละประเภทคือแอปเปิ้ล , กล้วย และส้ม ได้อัตราความถูกต้องเฉลี่ยสูงสุดคือ CNNsF ที่ร้อยละ 97.61 (Prati, Monard, 2008)

### 2.3 บทความวิจัยเรื่อง Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks

บทความกล่าวถึงการคัดแยกรูปภาพมะกอกจำนวน 2,800 รูปภาพโดยแบ่งเป็น 7 สายพันธุ์ หรือว่า 7 Class โดยใช้ Architecture คือ Convolutional Neural Network (CNN) ทั้ง 6 แบบ คือ Alexnet, Inception-ResnetV2, InceptionV1, InceptionV3, Resnet-50 และ Resnet-101 โดยในแบบจำลองทั้ง 6 แบบนั้นค่าความถูกต้องเกินกว่าร้อยละ 90 ทั้งหมด โดยแบบจำลองที่ดีที่สุดคือ CNN แบบ Inception-ResnetV2 ค่าความถูกต้องอยู่ที่ร้อยละ 95.91 (Ponce, Aquino, & Andújar, 2019)

### 2.4 บทความวิจัย Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool

บทความกล่าวถึงการคัดแยกประเภทของผลไม้จำนวน 49 ภาพชุดข้อมูลจาก Google Image โดยใช้ Algorithm 6 แบบประกอบด้วย Logistic Regression, Random Forest, Naïve Bayes, K-Nearest neighbor, Decision Tree และ Neural Network จากการเทียบค่า Accuracy และ Precision โดยแบ่งข้อมูลออกเป็น 2 กลุ่มได้ผลดังนี้ ในกลุ่มที่ 1 Logistic Regression ได้ผลดีที่สุด ค่า Accuracy 85% และ Precision 86% กลุ่มที่ 2 Logistic Regression ได้ผลดีที่สุด ค่า Accuracy 91% และ Precision 92% (Vaishnav & Rao, 2018)

### 2.5 บทความวิจัยเรื่อง Fruit Classification Based on Six Layer Convolutional Neural Network

บทความกล่าวถึงการคัดแยกผลไม้ที่มีรูปร่างลักษณะใกล้เคียงกัน เนื่องจากการจ่ายเงินที่แคชเชียร์ซูเปอร์มาร์เก็ตราคาจะต่างกันผลไม้ที่ต่างชนิดกัน ซึ่งกรณีนี้ที่คิดราคาผิดพลาดกระทบต่อกำไรขาดทุนของธุรกิจ โดยใช้ข้อมูลในการวิจัยคือ ภาพผลไม้จำนวน 1,800 ภาพ ซึ่งแบ่งเป็นผลไม้ 9 ชนิด วิธีที่ใช้แยกประเภทผลไม้ใช้คือ Six-Layer Convolutional Neural Network ประกอบด้วย Convolutional Layer, Pooling Layer และ Fully connected Layer โดยเปรียบเทียบการทำ Six-Layer CNN เทียบกับ WE (wavelet entropy), GA (genetic algorithm) และ VB-SVM (Computer Vision and a Multiclass Support Vector Machine) โดยค่าความถูกต้องที่ดีที่สุดคือ Six-Layer CNN ที่ 91.44% (Lu, Lu, Aok, & Graham, 2018)

## 2.6 บทความวิจัย Image-based Processing for Ripeness Classification of Oil Palm Fruit

บทความกล่าวถึงการคัดแยกผลปาล์มที่ดิบ, ผลปาล์มที่ไม่สุกและผลปาล์มที่สุก วัตถุประสงค์คือต้องการที่จะเพิ่มพัฒนาประสิทธิภาพผลผลิตของผลปาล์ม โดยในบทความใช้วิธีการ Segmentation , Feature Extraction และ Classification ในขั้นตอนของการแยกประเภทของผลไม้แบ่งเป็น 3 Class ดิบ, ไม่สุก, สุก ผสมผสานกับ RGB และ grey color features และ SVM โดยรูปที่นำมาทดสอบทั้งหมด 160 รูปภาพ ประกอบด้วยรูปผลปาล์มดิบ 60 รูป , รูปผลปาล์มไม่สุก 50 รูป , รูปผลปาล์มสุก 50 รูป โดยแบ่งเป็น Training และ Testing โดยทั้ง 2 Stage จะแบ่งเป็น 4 กระบวนการคือ นำรูปผลปาล์มเข้า, Segmentation, Feature Extraction และ Classification (ใช้ SVM) ซึ่งผลการทดสอบพบว่าการใช้ SVM ที่ K-fold = 20 ให้ค่าความถูกต้องที่สูงที่สุดคือ 92.50% (Septiarini, Hamdani, Hatta, & Kasim, 2019)

## 2.7 บทความวิจัย Research on Fruit Category Classification Based on Convolution Neural Network and Data Augmentation

บทความนี้กล่าวถึงความสำคัญของการคัดแยกผลไม้ที่มีส่วนสำคัญต่อการผลผลิตของผลไม้ โดยใช้ข้อมูลการทดสอบคือ Fruit-360 Dataset โดยในการคัดแยกประเภทของผลไม้เน้นใช้วิธี Convolutional Neural Network (CNN) เป็นหลัก โดยเปรียบเทียบค่าความถูกต้องระหว่าง IANET (5 Convolutional Layers, 4 Fully Connection Layers and 1 output layer), HoreaCNN, ALexNet และ 13-LayerCNN ประกอบกับการใช้เทคนิค Data Augmentation พบว่าวิธี IANET พร้อมกับการใช้ Data Augmentation ให้ค่าความถูกต้อง 98.60% ซึ่งเมื่อเทียบกับวิธี IANET และไม่ใช้ Data Augmentation จะได้ค่าความถูกต้องเพียง 98.06% (Zhu, Wang, Zou, Shen, & Luo, 2019)

## 2.8 บทความวิจัย Classification on Passion Fruit's Ripeness using K-Means Clustering and Artificial Neural Network

บทความนี้กล่าวถึงการระบุคัดแยกระดับความสุกของผลไม้เสาวรส โดยแบ่งเป็น 3 ประเภทคือ เสาวรสที่สุก, เสาวรสที่ไม่สุก และเสาวรสที่ดิบ เนื่องจากการคัดแยกผลไม้ในอุตสาหกรรมใช้ต้นทุนแรงงานในการคัดแยกเยอะมาก ซึ่งส่งผลต่อการแข่งขันในตลาด วิธีที่ใช้ในงานวิจัย K-Means Clustering และ ANN (Artificial Neural Network) ในการแบ่งประเภท

ฐานข้อมูลที่ใช้ในการทดสอบคือใช้คลิปวิดีโอยาว 5 วินาที ทั้งหมด 95 วิดีโอ แบ่งเป็นข้อมูลฝึกฝน (Training Data) 75 วิดีโอ และข้อมูลทดสอบ (Test Data) 20 วิดีโอ ซึ่งจากผลการวิจัยพบว่า ค่าความถูกต้องอยู่ที่ 90% ซึ่งความผิดพลาดที่เกิดขึ้น เกิดบนภาพที่เสาวรสีไม่สูงหรือมืดเนื่องจากมีสีที่คล้ายกัน (Sidehabi, Suyuti, Areni, & Nurtanio, 2018)

## 2.9 บทความวิจัย Fruit Classification using Statistical Features in SVM Classifier

บทความนี้กล่าวถึงความสำคัญของการคัดแยกผลไม้โดยอัตโนมัติในอุตสาหกรรมผลไม้และในซูเปอร์มาร์เก็ต การคัดแยกผลไม้ชนิดต่างๆ ประกอบด้วย 10 ชนิด ทั้งหมด 941 รูปภาพ ซึ่งใช้วิธีการ 3 ขั้นตอน คือ Data Preprocessing (ใช้ Image Preprocessing), Feature extraction และ Classification ซึ่งในกรณีนี้ใช้ SVM ซึ่งผลการทดสอบในภาพรวมค่าความถูกต้องอยู่ที่ 95.3% (Kumari & Gomathy, 2018)

## 2.10 บทความวิจัย Snake Fruit Classification by Using Histogram of Oriented Gradient Feature and Extreme Learning Machine

บทความกล่าวถึงการคัดเลือกเพื่อทำการส่งออกผลไม้สด โดยแบ่งเป็น 2 Class คือ สลัดที่ส่งออกได้กับสลัดที่ส่งออกไม่ได้ โดยกระบวนการแบ่งเป็น Preprocess, Feature Extraction และทำการ Classification ซึ่งในงานวิจัยนี้ใช้ Extreme Learning Machine (ELM) และ Support Vector Machine (SVM) โดยทั้ง 2 วิธีสามารถแยกผลไม้สดที่สามารถส่งออกได้และส่งออกไม่ได้ ซึ่งผลการทดสอบปรากฏว่าค่าความถูกต้องในการ Classification สำหรับ SVM คือ 97.2% ส่วน ELM อยู่ที่ 95.3% ซึ่ง SVM สามารถแยกความแตกต่างของคุณภาพสลัด (Rismiyati & Wibawa, 2019)

## 2.11 บทความวิจัย Deep Transfer Learning for Multiple Class Novelty Detection

งานวิจัยนี้ผู้วิจัยพยายามที่จะสร้างตัว Deep Novelty ซึ่งเป็นการตรวจจับข้อมูลใหม่ (Novelty Detection) ภาพที่มันผิดปกติออกไป โดยใช้ Membership Loss คือกรณีที่ Known Class คือ Only Single Label เพื่อบรรเทา Cross-entropy Loss นักวิจัยจะเรียกว่า Membership Loss เพิ่มเติม และ Globally Negative filter คือ การตั้งค่าจัดหมวดหมู่ Negative Filter ในคลาสหนึ่งสามารถเป็น Positive Filter ในอีกคลาสหนึ่ง ซึ่งในสูตรนักวิจัยใช้สำหรับลดค่าความเป็นไปได้ไม่ให้สูงเกินไปข้อสรุปจากงานวิจัยนี้ในการเรียนรู้เชิงลึกสำหรับการตรวจจับข้อมูลใหม่ (Novelty Detection) ของภาพโดยผู้วิจัยได้สร้างเครือข่าย Conventional Classification และแนะนำ 2 เรื่อง

ได้แก่ Membership Loss และ Training procedure ซึ่งเป็นตัวสร้าง Globally Negative Filter โดยในงานวิจัยนี้ทดสอบประสิทธิภาพของวิธีที่นักวิจัยเตรียมมาทดสอบเทียบกับวิธีอื่น เช่น Finetune, One-Class SVM, KNFST pre, KNFST, Local KNFST pre, Local KNFST, K-extremes, OpenMax, Finetune (c+C), Deep Novelty โดยทั้งหมดที่กล่าวมาข้างต้นจะใช้ Feature เดียวกันคือ VGG16 และ AlexNet (Perera & Patel, 2019)

## 2.12 บทความวิจัย Open-Set Bottle Classifying using a Convolution Neural Network

งานวิจัยนี้มุ่งเน้นการจัดประเภทขวดเครื่องดื่ม ภายใต้สภาพแวดล้อมแบบเปิดซึ่งภาพที่นำเข้ามาอาจจะไม่ได้เข้ามาอยู่ในช่วงเวลาของการเทรนข้อมูล โดยแบบจำลองที่จะใช้มี 3 แบบจำลองคือ N-Binary (คือวิธีที่พยายามระบุ Class ที่ไม่รู้โดยระบุเป็น Positive หรือ Negative), N+Unknown (คือวิธีการสร้าง 3 Class ที่เป็นเป้าหมาย ส่วน Class ที่ไม่รู้จักจะถูกเทรนด้วยภาพที่ไม่ได้เป็น Class เป้าหมาย) และ N+Combination (คือวิธีการสร้าง 3 Class ที่เป็นเป้าหมาย ส่วน Class ที่ไม่รู้จักจะถูกเทรนด้วยภาพที่เป็น Class เป้าหมาย) ซึ่งชุดข้อมูลที่นำมาประเมินประกอบด้วย 11,309 รูป ประกอบด้วย Product1 จำนวน 1,024 รูป, Product2 จำนวน 1,027 รูป และ Product3 จำนวน 1,032 รูป และ Unknown product ประกอบด้วย 8,226 รูป ผลการทดสอบทั้ง 3 วิธีพบว่า ค่าความถูกต้องที่สูงที่สุดคือ N+Unknown ที่ 92% (Jintawatsakoon & Charoenruangkit, 2019)

## 2.13 บทความวิจัย Fault Detection of Hydroelectric Generators using Isolation Forest

บทความงานวิจัยนี้ กล่าวถึงการตรวจจับข้อมูลที่มีความผิดปกติจากเครื่องกำเนิดไฟฟ้าพลังน้ำ ซึ่งเปรียบเทียบประสิทธิภาพระหว่างวิธีป่าแยก (Isolation Forest) กับวิธีควบคุมกระบวนการผลิตด้วยหลักสถิติหลายตัวแปร (Multivariate Statistical Process Control : MSPC) โดยกำหนดไฮเปอร์พารามิเตอร์ของวิธีป่าแยก (Isolation Forest) กำหนดค่า Contamination ที่ 0.333 และวิธีควบคุมกระบวนการผลิตด้วยหลักสถิติหลายตัวแปร (Multivariate statistical Process Control: MSPC) ใช้ ค่า Q-Value ที่ 0.997 ซึ่งตัววัดประสิทธิภาพคือพื้นที่ใต้เส้นโค้ง (AUC) โดยแบ่งเป็น 4 แบบคือ ค่าเฉลี่ย (Average), ค่าสูงสุด (Max), ค่าต่ำสุด (Min) และ ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ซึ่งผลปรากฏว่าแบบจำลองป่าแยก (Isolation forest) ให้ค่าความถูกต้องที่มากกว่าในทุกวิธี (Hara & Fukuyama & Murakami & Iizaka & Matsui, 2020)

## 2.14 บทความวิจัย Weed Classification Using One Class Support Vector Machine

งานวิจัยกล่าวถึงการคัดแยกพันธุ์ของหญ้าเพื่อประโยชน์ในการดำรงไว้ซึ่งประสิทธิภาพและคุณภาพของหญ้า โดยเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทสองชนิด (Binary Support Vector Machine) ซึ่งกำหนดขนาดภาพอยู่ที่  $100 \times 100$  พิกเซล และกำหนดพารามิเตอร์ดังนี้ Kernel คือ RBF , กำหนดค่าความผิดพลาดในการทำนายพลาด ( $\nu$ ) แตกต่างกันว่า 0.7, 0.9 และ 0.2 พบว่าค่าความถูกต้องของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) มีความแม่นยำกว่า แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทสองชนิด (Binary Support Vector Machine) ทุกรูปแบบ (Shahbudin & Zamri & Kassim & Afzal & Abdullah & Suliman, 2021)

### บทที่ 3

#### วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

##### 3.1 กระบวนการสร้างแบบจำลอง



ภาพประกอบ 7 แผนผังกระบวนการสร้างแบบจำลอง

กระบวนการวิจัยในการสร้างแบบจำลองตรวจจับข้อมูลใหม่ (Novelty Detection) เริ่มต้นงานวิจัยด้วยนำข้อมูลภาพผลไม้ 360 องศามาทำการสำรวจข้อมูล (Exploratory data analysis : EDA) ถึงจำนวนรูปภาพผลไม้ทั้งหมดทั้งที่เป็นรูปภาพแอปเปิ้ลและไม่ใช่รูปภาพแอปเปิ้ล จากนั้นทำการเลือกข้อมูลชุดฝึกฝนที่จะใช้ทำการนำเข้าแบบจำลอง ทั้ง 2 แบบได้แก่ ใช้ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) หลังจากนั้นทำการประเมินผลลัพธ์ของแบบจำลองด้วยการวัดประสิทธิภาพ 4 แบบด้วย Confusion Matrix, F1-Score, Accuracy และ ROC AUC

### 3.2 การสำรวจข้อมูล (Exploratory Data Analysis: EDA)

1. ภาพที่นำมาใช้ในงานวิจัยคือเป็นภาพที่ใช้การถ่ายทำผลไม้ในขณะที่กำลังหมุนภาพแล้วทำการแยกเฟรม โดยด้านหลังของผลไม้วางสีขาว ขนาดของภาพอยู่ในขนาด 100 x 100 pixels โดยในชุดข้อมูลประกอบด้วยภาพผักและผลไม้จำนวน 90,380 ภาพ ผ่านการจัดประเภท 131 Class (Label) ซึ่งแต่ละรูปประกอบด้วยรูปเดี่ยวของผักและผลไม้เท่านั้น ประกอบด้วยภาพ

ตาราง 1 แสดงจำนวนข้อมูลผลไม้ทั้งหมดใน Dataset Fruit-360 องศา

| Label              | Number of Training Images | Number of Test Images |
|--------------------|---------------------------|-----------------------|
| Apple Braeburn     | 492                       | 164                   |
| Apple Crimson Snow | 444                       | 148                   |
| Apple Golden 1     | 480                       | 160                   |

ตาราง 1 (ต่อ)

| Label                    | Number of Training Images | Number of Test Images |
|--------------------------|---------------------------|-----------------------|
| Apple Golden 2           | 492                       | 164                   |
| Apple Golden 3           | 481                       | 161                   |
| Apple Granny Smith       | 492                       | 164                   |
| Apple Pink Lady          | 456                       | 152                   |
| Apple Red 1              | 492                       | 164                   |
| Apple Red 2              | 492                       | 164                   |
| Apple Red 3              | 429                       | 144                   |
| Apple Red Delicious<br>1 | 490                       | 166                   |
| Apple Red Yellow 2       | 492                       | 164                   |

ตาราง 1 (ต่อ)

| Label              | Number of Training Images | Number of Test Images |
|--------------------|---------------------------|-----------------------|
| Apple Red Yellow   | 672                       | 219                   |
| Apricot            | 492                       | 164                   |
| Avocado            | 427                       | 143                   |
| Avocado ripe       | 491                       | 166                   |
| Banana             | 490                       | 166                   |
| Banana Lady Finger | 450                       | 152                   |
| Banana Red         | 490                       | 166                   |
| Beetroot           | 450                       | 150                   |
| Blueberry          | 462                       | 154                   |

ตาราง 1 (ต่อ)

| Label            | Number of Training Images | Number of Test Images |
|------------------|---------------------------|-----------------------|
| Cactus fruit     | 490                       | 166                   |
| Cantaloupe 1     | 492                       | 164                   |
| Cantaloupe 2     | 492                       | 164                   |
| Carambula        | 490                       | 166                   |
| Cauliflower      | 702                       | 234                   |
| Cherry 1         | 492                       | 164                   |
| Cherry 2         | 738                       | 246                   |
| Cherry Rainier   | 738                       | 246                   |
| Cherry Wax Black | 492                       | 164                   |

ตาราง 1 (ต่อ)

| Label             | Number of Training Images | Number of Test Images |
|-------------------|---------------------------|-----------------------|
| Cherry Wax Red    | 492                       | 164                   |
| Cherry Wax Yellow | 492                       | 164                   |
| Chestnut          | 450                       | 153                   |
| Clementine        | 490                       | 166                   |
| Cocos             | 490                       | 166                   |
| Dates             | 490                       | 166                   |
| Eggplant          | 468                       | 156                   |
| Ginger Root       | 297                       | 99                    |
| Granadilla        | 490                       | 166                   |

ตาราง 1 (ต่อ)

| Label            | Number of Training Images | Number of Test Images |
|------------------|---------------------------|-----------------------|
| Grape Blue       | 984                       | 328                   |
| Grape Pink       | 492                       | 164                   |
| Grape White      | 490                       | 166                   |
| Grape White 2    | 490                       | 166                   |
| Grape White 3    | 492                       | 164                   |
| Grape White 4    | 471                       | 158                   |
| Grapefruit Pink  | 490                       | 166                   |
| Grapefruit White | 492                       | 164                   |
| Guava            | 490                       | 166                   |

ตาราง 1 (ต่อ)

| Label           | Number of Training Images | Number of Test Images |
|-----------------|---------------------------|-----------------------|
| Hazelnut        | 464                       | 157                   |
| Cucumber Ripe 2 | 468                       | 156                   |
| Cucumber Ripe   | 392                       | 130                   |
| Corn            | 450                       | 150                   |
| Corn Husk       | 462                       | 154                   |
| Huckleberry     | 490                       | 166                   |
| Kaki            | 490                       | 166                   |
| Kiwi            | 466                       | 156                   |
| Kohlrabi        | 471                       | 157                   |

ตาราง 1 (ต่อ)

| Label       | Number of Training Images | Number of Test Images |
|-------------|---------------------------|-----------------------|
| Kumquats    | 490                       | 166                   |
| Lemon       | 492                       | 164                   |
| Lemon Meyer | 490                       | 166                   |
| Limes       | 490                       | 166                   |
| Lychee      | 490                       | 166                   |
| Mandarine   | 490                       | 166                   |
| Mango       | 490                       | 166                   |
| Fig         | 702                       | 234                   |
| Mango Red   | 426                       | 142                   |

ตาราง 1 (ต่อ)

| Label              | Number of Training Images | Number of Test Images |
|--------------------|---------------------------|-----------------------|
| Mangostan          | 300                       | 102                   |
| Maracuja           | 490                       | 166                   |
| Melon Piel de Sapo | 738                       | 246                   |
| Mulberry           | 492                       | 164                   |
| Nectarine          | 492                       | 164                   |
| Nectarine Flat     | 480                       | 160                   |
| Nut Forest         | 654                       | 218                   |
| Nut Pecan          | 534                       | 178                   |
| Onion Red          | 450                       | 150                   |

ตาราง 1 (ต่อ)

| Label            | Number of Training Images | Number of Test Images |
|------------------|---------------------------|-----------------------|
| Onion Red Peeled | 445                       | 155                   |
| Onion White      | 438                       | 146                   |
| Orange           | 479                       | 160                   |
| Papaya           | 492                       | 164                   |
| Passion Fruit    | 490                       | 166                   |
| Peach            | 492                       | 164                   |
| Peach 2          | 738                       | 246                   |
| Peach Flat       | 492                       | 164                   |
| Pear             | 492                       | 164                   |

ตาราง 1 (ต่อ)

| Label         | Number of Training Images | Number of Test Images |
|---------------|---------------------------|-----------------------|
| Pear Abate    | 490                       | 166                   |
| Pear Forelle  | 702                       | 234                   |
| Pear Kaiser   | 300                       | 102                   |
| Pear Monster  | 490                       | 166                   |
| Pear Red      | 666                       | 222                   |
| Pear 2        | 696                       | 232                   |
| Pear Stone    | 711                       | 237                   |
| Pear Williams | 490                       | 166                   |
| Pepino        | 490                       | 166                   |

ตาราง 1 (ต่อ)

| Label              | Number of Training Images | Number of Test Images |
|--------------------|---------------------------|-----------------------|
| Pepper Green       | 444                       | 148                   |
| Pepper Red         | 666                       | 222                   |
| Pepper Yellow      | 666                       | 222                   |
| Pepper Orange      | 702                       | 234                   |
| Physalis           | 492                       | 164                   |
| Physalis with Husk | 492                       | 164                   |
| Pineapple          | 490                       | 166                   |
| Pineapple Mini     | 493                       | 163                   |
| Pitahaya Red       | 490                       | 166                   |

ตาราง 1 (ต่อ)

| Label            | Number of Training Images | Number of Test Images |
|------------------|---------------------------|-----------------------|
| Plum             | 447                       | 151                   |
| Plum 2           | 420                       | 142                   |
| Plum 3           | 900                       | 304                   |
| Pomegranate      | 492                       | 164                   |
| Pomelo Sweetie   | 450                       | 153                   |
| Potato Red       | 450                       | 150                   |
| Potato RedWashed | 453                       | 151                   |
| Potato Sweet     | 450                       | 150                   |
| Potato White     | 450                       | 150                   |

ตาราง 1 (ต่อ)

| Label           | Number of Training Images | Number of Test Images |
|-----------------|---------------------------|-----------------------|
| Quince          | 490                       | 166                   |
| Rambutan        | 492                       | 164                   |
| Raspberry       | 490                       | 166                   |
| Redcurrant      | 492                       | 164                   |
| Salak           | 490                       | 162                   |
| Strawberry      | 492                       | 164                   |
| StrawberryWedge | 738                       | 246                   |
| Tamarillo       | 490                       | 166                   |
| Tangelo         | 490                       | 166                   |

ตาราง 1 (ต่อ)

| Label              | Number of Training Images | Number of Test Images |
|--------------------|---------------------------|-----------------------|
| Tomato 1           | 738                       | 246                   |
| Tomato Heart       | 684                       | 228                   |
| Tomato not Ripened | 474                       | 158                   |
| Watermelon         | 475                       | 157                   |
| Tomato 2           | 672                       | 225                   |
| Tomato 3           | 738                       | 246                   |
| Tomato 4           | 479                       | 160                   |
| Tomato Cherry Red  | 492                       | 164                   |
| Tomato Maroon      | 367                       | 127                   |

ตาราง 1 (ต่อ)

| Label         | Number of Training Images | Number of Test Images |
|---------------|---------------------------|-----------------------|
| Tomato Yellow | 459                       | 153                   |
| Walnut        | 735                       | 249                   |
| Total         | 67,692                    | 22,688                |

ในงานวิจัยนี้เลือกใช้แอปเปิ้ลใช้สำหรับชุดข้อมูลการฝึกฝน (Training data) แอปเปิ้ล 13 ชนิดจำนวน 6,404 ภาพ และข้อมูลที่จะนำมาทดสอบทั้งแอปเปิ้ลและไม่ใช้แอปเปิ้ลจำนวน 2,134 รูปและ 20,554 รูปตามลำดับ ดังตาราง 2

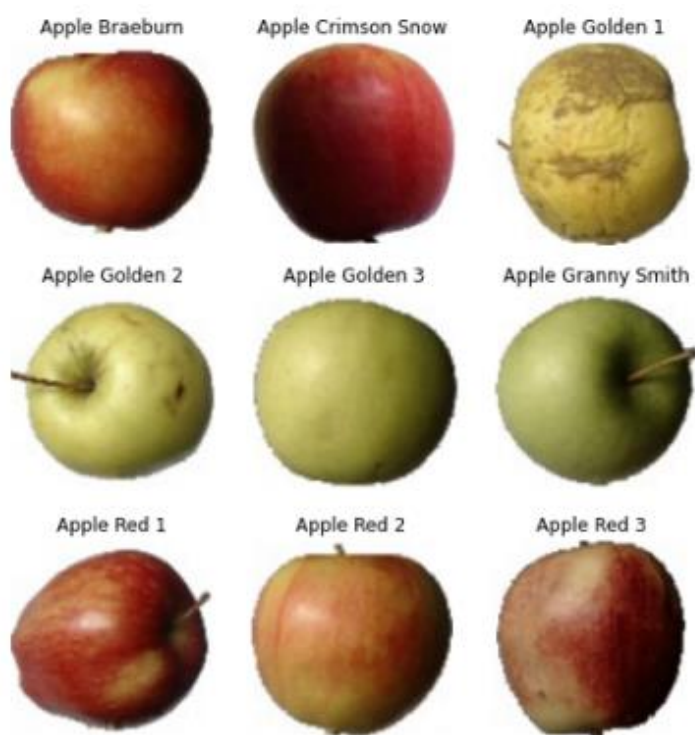
ตาราง 2 ชุดข้อมูลที่นำมาใช้ในงานวิจัย

| Label | Number of Training Images | Number of Test Images |
|-------|---------------------------|-----------------------|
| Apple | 6,404                     | 2,134                 |

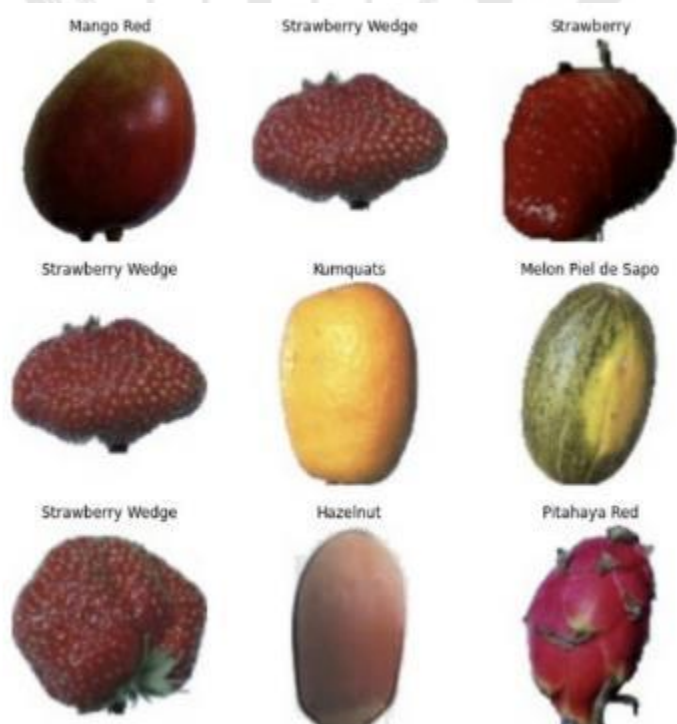
ตาราง 2 (ต่อ)

| Label     | Number of Training Images | Number of Test Images |
|-----------|---------------------------|-----------------------|
| Not Apple | -                         | 20,554                |
| Total     | 6,404                     | 22,688                |

2.จากการนำข้อมูลแอปเปิ้ลและผลไม้ชนิดอื่นๆเดี่ยวมาทำการสำรวจข้อมูล (Exploratory data analysis: EDA) โดยเลือกภาพผลไม้แอปเปิ้ล 13 ประเภท ที่ใช้รูปภาพในการฝึกฝนข้อมูล (Training data) และผลไม้ชนิดอื่นๆที่ไม่ใช่ภาพแอปเปิ้ลใช้ในการทดสอบแบบจำลอง (Testing data) ดังภาพประกอบ 7 และ 8



ภาพประกอบ 8 ตัวอย่างข้อมูลรูปภาพชุดข้อมูล Fruit-360 ที่เป็นแอปเปิ้ล



ภาพประกอบ 9 ตัวอย่างข้อมูลรูปภาพชุดข้อมูล Fruit-360 ที่ไม่ใช่แอปเปิ้ล

### 3.3 การเตรียมข้อมูล (Data Preprocessing)

ในงานวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบลึกที่เรียกว่า ResNet50 ใช้สำหรับการดึงคุณลักษณะจากภาพออกมาซึ่งจะทำให้มิติหรือว่า Dimension สูงขึ้น โดยเป็นแบบจำลองที่มีชั้นของ Layer จำนวน 50 เลเยอร์ ซึ่งสถาปัตยกรรมประกอบด้วย 4 Block ในแต่ละ Block จะมี Convolutional Layer ประกอบด้วย 3, 4, 6 และ 3 ตามลำดับซึ่ง รวม Convolutional Layer ชั้นที่ติดกับข้อมูลที่นำเข้า และชั้นที่ติดกับข้อมูลที่ทำนายออกจากสถาปัตยกรรม ถัดมาคือการทำ Standard Scaler เพื่อให้ภาพที่ออกมาจาก ResNet50 ที่อยู่ในมิติที่สูงขึ้นนั้นอยู่ในมาตรฐานเดียวกัน หลังจากนั้นทำการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis : PCA) เป็นการทำให้องค์ประกอบที่สัมพันธ์กันมาอยู่รวมกันทำให้ข้อมูลเล็กลงหรือลดสิ่งรบกวนลง หลักการทำงานคือหาเส้นตรงหรือ Hyperplane เพื่อฉายภาพจากมิติที่สูงกว่าลงมายัง Hyperplane ในมิติที่ต่ำกว่า ในงานวิจัยนี้ใช้สำหรับลดมิติหรือ Dimension ของข้อมูลลงหลังจากผ่านขั้นตอนการเตรียมข้อมูลแล้วขั้นตอนถัดไปคือการนำข้อมูลเข้าแบบจำลองเพื่อฝึกฝน

วิธีการทดลองดำเนินงานวิจัยในการสร้างแบบจำลองการคัดแยกผลไม้แอปเปิ้ลออกจากผลไม้อื่นๆ การดำเนินการโดยทำบนแพลตฟอร์มของ Kaggle เนื่องจากข้อมูลภาพมีขนาดใหญ่จึงต้องใช้ทรัพยากรที่ค่อนข้างมาก โดยกระทำการดำเนินงานวิจัยมีขั้นตอนดังนี้

ขั้นตอนการเตรียมข้อมูลเพื่อนำเข้าแบบจำลองประกอบด้วย 2 ส่วนหลักๆคือ ResNet50 ใช้สำหรับการดึงคุณลักษณะเด่นของภาพออกมาโดยกำหนดพารามิเตอร์ดังตารางที่ 3

ตาราง 3 กำหนดไฮเปอร์พารามิเตอร์ของ ResNet50

| Hyperparameter          | Value                           |
|-------------------------|---------------------------------|
| ขนาดของภาพ (Image_size) | 100 x 100 pixel                 |
| จำนวนน้ำหนัก (Weight)   | Imagenet เป็น Transfer Learning |

ตาราง 3 (ต่อ)

| Hyperparameter                               | Value                                                            |
|----------------------------------------------|------------------------------------------------------------------|
| กำหนด Fully-Connected layer<br>(Include_top) | False คือกำหนดให้ไม่รวม Fully-Connected layer                    |
| รูปแบบการทำ Feature Extraction<br>(Pooling)  | Avg เป็น ค่าเฉลี่ยในการดึงรูปภาพ<br>ประยุกต์ใช้กับผลลัพธ์ 2 มิติ |

ซึ่งขั้นตอนถัดมาคือ Standard Scaler เพื่อให้ภาพที่ผ่านออกมาจาก ResNet50 อยู่ในมาตรฐานเดียวกัน หลังจากนั้นทำการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis : PCA) เพื่อลดมิติหรือ Dimension ของข้อมูลลง โดยในงานวิจัยนี้ผู้วิจัยได้กำหนดพารามิเตอร์ที่ผ่านการทดลองเปลี่ยนค่าแล้วให้ผลลัพธ์ที่ดีที่สุดดังนี้ n\_components เท่ากับ 512 และ whiten เท่ากับ True โดย n\_components คือ จำนวนองค์ประกอบหลักที่ใช้ ส่วนWhiten จะทำการสเกลข้อมูล หลังจากผ่านขั้นตอนการเตรียมข้อมูลแล้วขั้นตอนถัดไปคือการนำข้อมูลเข้าแบบจำลองที่เตรียมไว้

### 3.4 การสร้างแบบจำลอง (Model Creation)

แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) กำหนดไฮเปอร์พารามิเตอร์ต่างๆของดังนี้ Kernel คือ RBF (ฟังก์ชันฐานแนวรัศมี) เป็น Kernel ที่ใช้งานทั่วไป ใช้เมื่อไม่เคยเห็นข้อมูลและทำงานในมิติที่ไม่จำกัด , กำหนดค่าความผิดพลาดในการทำนายพลาด ( $\nu$ ) เท่ากับ 0.08 คือจำนวนที่เราคิดว่าน่าจะทำนายไม่ถูกต้อง และพารามิเตอร์  $\gamma$  เท่ากับ 0.001 เป็นตัวกำหนดความเรียบของเส้นชั้นความสูงแบบจำลองป่าแยก (Isolation Forest) กำหนดไฮเปอร์พารามิเตอร์ดังนี้ Contamination คือ 0.08, Max\_features คือ 1.0, Max\_samples คือ 1.0 และ n\_estimators คือ 4

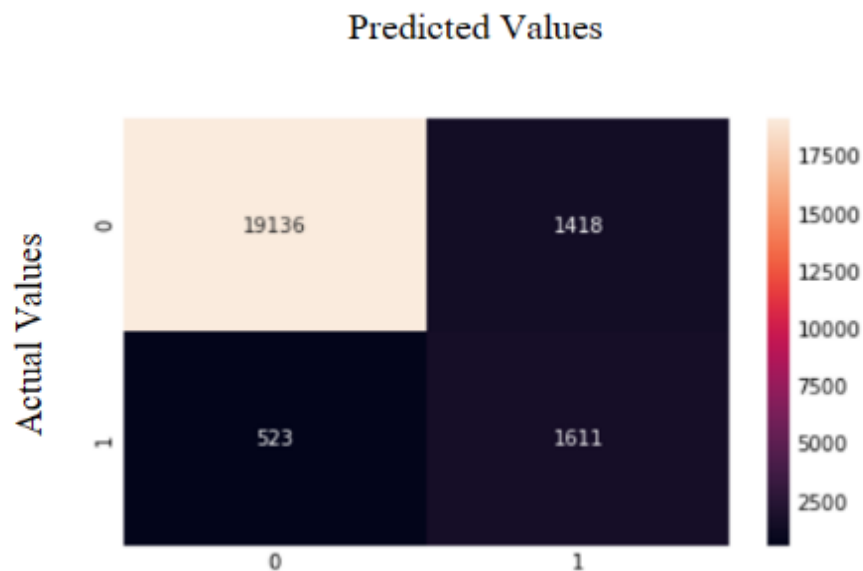
ในงานวิจัยนี้ผู้วิจัยได้ทดลองปรับค่าไฮเปอร์พารามิเตอร์ต่างๆของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และแบบจำลองป่าแยก (Isolation Forest) พบว่าค่าไฮเปอร์พารามิเตอร์ที่แสดงก่อนหน้าให้ผลที่ดีที่สุด

## บทที่ 4

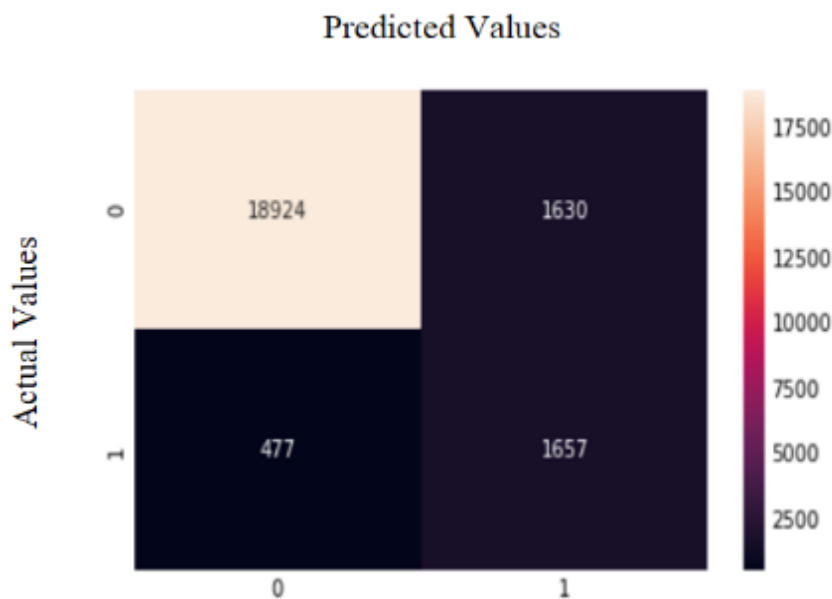
### การทดลอง

จากผลการทดลองพบว่าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) จากภาพประกอบ 10 การทำนายภาพแอปเปิ้ลถูกต้องจำนวน 1,611 รูป , ทำนายรูปที่ไม่ใช่แอปเปิ้ลถูกต้องจำนวน 19,136 รูป , ทำนายรูปว่าไม่ใช่แอปเปิ้ลแต่ความจริงเป็นแอปเปิ้ล 523 รูป และทำนายรูปว่าเป็นแอปเปิ้ลแต่ความจริงไม่ใช่แอปเปิ้ลจำนวน 1,418 รูปซึ่งได้ค่าความถูกต้องทั้งหมดร้อยละ 91

จากผลการทดลองพบว่าแบบจำลองป่าแยก (Isolation Forest) จากภาพประกอบ 11 การทำนายภาพแอปเปิ้ลถูกต้องจำนวน 1,657 รูป , ทำนายรูปที่ไม่ใช่แอปเปิ้ลถูกต้องจำนวน 18,924 รูป , ทำนายรูปที่ไม่ใช่แอปเปิ้ลแต่ความจริงเป็นแอปเปิ้ล 477 รูป และทำนายรูปที่เป็นแอปเปิ้ลแต่ความจริงไม่ใช่แอปเปิ้ล จำนวน 1,630 รูปซึ่งได้ค่าความถูกต้องทั้งหมดร้อยละ 91



ภาพประกอบ 10 Confusion matrix ของ One-Class support vector machine โดยกำหนดค่า 0 คือ ไม่ใช่รูปผลไม้แอปเปิ้ล และค่า 1 คือรูปผลไม้แอปเปิ้ล



ภาพประกอบ 11 Confusion matrix ของ Isolation Forest โดยกำหนดค่า 0 คือ ไม่ใช่รูปผลไม้แอปเปิ้ล และค่า 1 คือรูปผลไม้แอปเปิ้ล

หลังจากการประเมินประสิทธิภาพของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ด้วย Confusion Matrix แล้ว ผู้วิจัยได้วัดประสิทธิภาพของแบบจำลองทั้ง 2 แบบด้วยค่าความถูกต้อง (Accuracy), ค่า F1-Score, Macro Avg และค่าพื้นที่ใต้กราฟ (ROC AUC) สรุปผลได้ว่าค่าความถูกต้อง (Accuracy) ของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และแบบจำลองป่าแยก (Isolation Forest) นั้นเท่ากันที่ร้อยละ 91 , ค่า F1-Score, Macro Avg คือ ค่าเฉลี่ยของ Precision และ Recall ของทั้ง 2 class ผลคือแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) ดีกว่าแบบจำลองป่าแยก (Isolation Forest) เล็กน้อยที่ร้อยละ 79 และร้อยละ 78 ตามลำดับ และสุดท้ายคือการวัดประสิทธิภาพด้วยค่าพื้นที่ใต้กราฟ (ROC AUC) ซึ่งแบบจำลองป่าแยก (Isolation Forest) ให้ผลที่ดีกว่าเล็กน้อยที่ร้อยละ 84.85 และร้อยละ 84.29 ตามตารางที่ 4

ตาราง 4 ตารางเปรียบเทียบผลการทดลองของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest)

| Measure             | One-class SVM | Isolation Forest |
|---------------------|---------------|------------------|
| Accuracy            | 91%           | 91%              |
| F1-Score, Macro Avg | 79%           | 78%              |
| ROC AUC             | 84.29%        | 84.85%           |

ตัวอย่างรูปของการทำนายของ แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) ตามภาพประกอบ 11, 12, 13 และ 14

TRUE POSITIVE

oc\_svm\_preds: 1 | actual: 1



ภาพประกอบ 12 True Positive คือเป็นภาพผลไม้แอปเปิ้ล และทำนายว่าเป็นภาพแอปเปิ้ลจริง

TRUE NEGATIVE

oc\_svm\_preds: 0 | actual: 0



ภาพประกอบ 13 True Negative คือไม่ใช่ภาพผลไม้แอปเปิ้ล และทำนายว่าไม่เป็นภาพแอปเปิ้ลจริง

FALSE POSITIVE

oc\_svm\_preds: 1 | actual: 0



ภาพประกอบ 14 False Positive คือไม่ใช่ภาพผลไม้แอปเปิ้ล แต่ทำนายว่าเป็นภาพแอปเปิ้ลจริง

FALSE NEGATIVE

oc\_svm\_preds: 0 | actual: 1



ภาพประกอบ 15 False Negative คือเป็นภาพผลไม้แอปเปิ้ล แต่ทำนายว่าไม่ใช่ภาพแอปเปิ้ลจริง

ตัวอย่างรูปของการทำนายของแบบจำลองป่าแยก (Isolation Forest) ตามภาพประกอบ 15, 16, 17 และ 18

TRUE POSITIVE

if\_preds: 1 | actual: 1



ภาพประกอบ 16 True Positive คือเป็นภาพผลไม้แอปเปิ้ล และทำนายว่าเป็นภาพแอปเปิ้ลจริง

TRUE NEGATIVE

if\_preds: 0 | actual: 0



ภาพประกอบ 17 True Negative คือไม่ใช่ภาพผลไม้แอปเปิ้ล และทำนายว่าไม่เป็นภาพแอปเปิ้ลจริง

FALSE POSITIVE

if\_preds: 1 | actual: 0



ภาพประกอบ 18 False Positive คือไม่ใช่ภาพผลไม้แอปเปิ้ล แต่ทำนายว่าเป็นภาพแอปเปิ้ลจริง

FALSE NEGATIVE

if\_preds: 0 | actual: 1



ภาพประกอบ 19 False Negative คือเป็นภาพผลไม้แอปเปิ้ล แต่ทำนายว่าไม่ใช่ภาพแอปเปิ้ลจริง

ซึ่งการทำนายผิดไม่ว่าจะเกิดจากการที่ความจริงเป็นภาพผลไม้แอปเปิ้ล แต่ทำนายว่าไม่ใช่ภาพแอปเปิ้ล ชนิด False Negative หรือการทำนายว่าไม่ใช่ภาพผลไม้แอปเปิ้ล แต่ทำนายว่าเป็นภาพแอปเปิ้ล ชนิด False Positive เนื่องจากว่าข้อมูลที่เรานำมาเป็นชุดข้อมูลฝึกฝน (Training data) ทั้งหมดอาจจะไม่เพียงพอ, ชุดข้อมูลทดสอบ (Testing data) มีความหลากหลายที่สูงเกินไป หรือบางภาพที่ไม่ใช่แอปเปิ้ลที่นำมาทดสอบคล้ายกับชุดข้อมูลฝึกฝน (Training data) จึงทำนายผิด และสุดท้ายคือการปรับจูนไฮเปอร์พารามิเตอร์ถ้ามีการปรับจูนที่เหมาะสมมากกว่านี้อาจจะลดค่าของ False Negative และ False Positive ลงได้

## บทที่ 5

### สรุปผลการทดลอง

ในการทำวิจัยเรื่องแบบจำลองการคัดแยกผลไม้แบบหนึ่งชนิดด้วยการเรียนรู้ของเครื่อง เพื่อตรวจจํารูปแอปเปิ้ล ผู้วิจัยได้เปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อสรุปผลการทดลอง โดยแบ่งหัวข้อการสรุปผลได้ดังต่อไปนี้

#### 5.1 สรุปผลการวิจัย

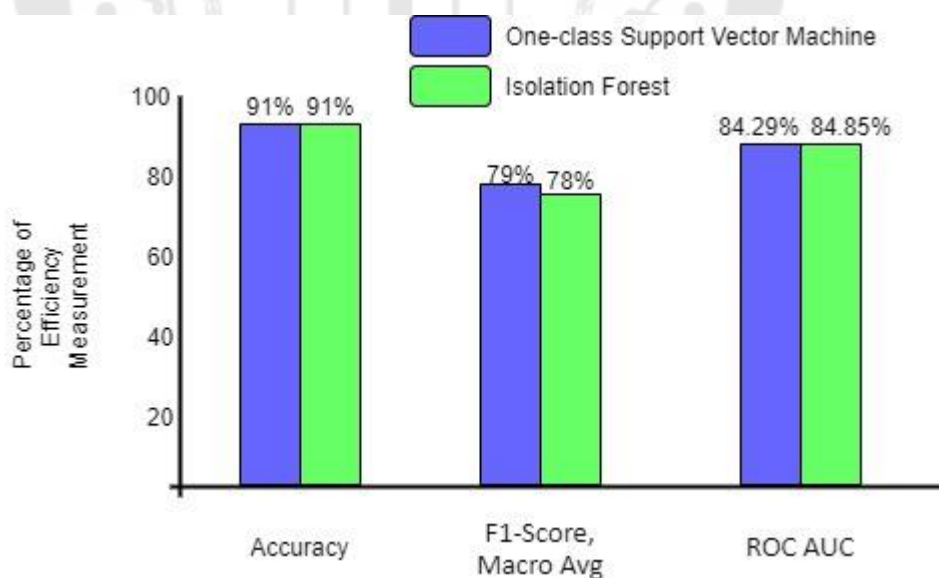
#### 5.2 ข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

งานวิจัยนี้สร้างขึ้นมาจากปัญหาของการคัดแยกผลไม้ชนิดต่างๆในซูเปอร์มาร์เก็ตที่ต้องใช้ทรัพยากรคนในการคัดแยกและอาจเกิดความผิดพลาดในการคัดแยกประเภทของผลไม้ในบางครั้งเนื่องจากบางชนิดนั้นมีรูปร่างที่คล้ายกันแต่ไม่ใช่ชนิดเดียวกัน เมื่อมีการนำผลไม้ขึ้นชั้นวางเพื่อเลือกซื้อ และเมื่อลูกค้าหยิบแล้วคืนในชั้นวางที่ไม่ใช่ชนิดของผลไม้ นั้น

จากปัญหาที่กล่าวมาทั้งหมดผู้วิจัยได้ทำการสร้างแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์คํายันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ประกอบกับนำหลักการของกระบวนการ ResNet50 เพื่อใช้ในการดึงลักษณะเด่นของภาพออกมา (Feature Extraction) ซึ่งส่งผลต่อมิติที่เพิ่มขึ้น และนำหลักการการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis : PCA) เพื่อลดมิติลงและนำข้อมูลเข้าแบบจำลองต่อไป หลังจากที้นำข้อมูลทดสอบ (Testing data) เข้าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์คํายันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) และวัดประสิทธิภาพ สรุปได้ว่าทั้ง 2 แบบจำลองมีประสิทธิภาพที่ใกล้เคียงกัน ประสิทธิภาพของแบบจำลองป่าแยก (Isolation Forest) ดีกว่า แบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์คํายันประเภทเดียว (One-Class Support Vector Machine) เนื่องจากมีการทำนายแอปเปิ้ลได้ถูกต้องมากกว่าอยู่ที่ 1,657 และ 1,611 ตามลำดับ ดังภาพประกอบที่ 9 และ 10 ค่าความถูกต้อง (Accuracy) ของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์คํายันประเภทเดียว (One-Class Support Vector Machine) และแบบจำลองป่าแยก (Isolation Forest) นั้นได้ค่าความถูกต้องที่ร้อยละ 91 เท่ากันตามภาพประกอบ 11 ค่า F1-Score, Macro Avg คือค่าเฉลี่ยของ Precision และ Recall ของรูปภาพที่เป็นแอปเปิ้ลและไม่ใช่แอปเปิ้ล พบว่าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์คํายันประเภทเดียว (One-Class Support Vector

Machine) ดีกว่าแบบจำลองป่าแยก (Isolation Forest) เล็กน้อยที่ร้อยละ 79 และร้อยละ 78 ตามลำดับตามภาพประกอบ 11 ซึ่งเหตุผลที่ F1-Score, Macro Avg มีค่าต่ำกว่าค่าความถูกต้อง (Accuracy) เนื่องจากค่า F1-Score, Macro Avg ดูเรื่องของความแม่นยำของแบบจำลอง แต่ค่าความถูกต้อง (Accuracy) ดูในเรื่องของความถูกต้องของแบบจำลอง จากภาพประกอบที่ 9 และ 10 จะพบว่าที่ค่า F1-Score ของทั้ง 2 แบบจำลองน้อยนั้นเพราะว่าแบบจำลองทั้ง 2 แบบให้ค่าความแม่นยำ (Precision) และการเรียกคืน (Recall) ของการทำนายภาพผลไม้แอปเปิ้ลค่อนข้างน้อย ดังนั้นเมื่อนำมาหาค่า F1-Score, Macro Avg คือการนำมาเฉลี่ยกันจะทำให้ค่า F1-Score, Macro Avg น้อย ในขณะที่ค่าความถูกต้อง (Accuracy) จะคำนวณมาจากจำนวนที่แบบจำลองทำนายรูปภาพผลไม้แอปเปิ้ลและรูปภาพที่ไม่ใช่แอปเปิ้ลถูก ซึ่งเป็นสัดส่วนที่มากเมื่อเทียบกับจำนวนทั้งหมดของรูปภาพทั้งหมดส่งผลให้ค่าความถูกต้อง (Accuracy) มีค่ามาก ค่าพื้นที่ใต้กราฟ (ROC AUC) พบว่าแบบจำลองป่าแยก (Isolation Forest) ให้ผลที่ดีกว่าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) เล็กน้อยที่ร้อยละ 84.85 และร้อยละ 84.29 ตามลำดับ ตามภาพประกอบ 19



ภาพประกอบ 20 เปรียบเทียบผลการทดลองของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ซึ่งประสิทธิภาพของทั้ง 2 แบบจำลองใกล้เคียงกันอย่างมาก

## 5.2 ข้อเสนอแนะ

5.2.1 ในงานวิจัยนี้ได้้นำการใช้ วิธีการเรียนรู้ของเครื่องจักร (Machine Learning) มาประยุกต์ และใช้ข้อมูลจากชุดภาพผลไม้ 360 องศาบางส่วนซึ่งเป็นภาพฉากหลังขาวทั้งหมดใน ทุกชนิดผลไม้ ซึ่งในทางปฏิบัติแล้วผลไม้แอปเปิ้ลอาจจะรวมอยู่กับผลไม้ชนิดอื่นๆ ซึ่งงานวิจัย อนาคตอาจจะมีการนำภาพรูปภาพผลไม้แอปเปิ้ลที่มีฉากหลังเป็นผลไม้ชนิดอื่นๆรวมอยู่ด้วย เพื่อเข้า มาสร้างแบบจำลองให้มีประสิทธิภาพและประยุกต์ใช้ในเชิงพาณิชย์ได้มากกว่านี้

5.2.2 นำข้อมูลจากชุดภาพผลไม้ 360 องศาประยุกต์ใช้กับอัลกอริทึมการเรียนรู้เชิง ลึก (Deep Learning) ประเภทอื่นเช่น แบบจำลองจำแนกข้อมูลด้วยโครงข่ายประสาทเทียม แบบสังวัตนาการ (Convolutional Neural Network) หรือแบบจำลองที่ถูกฝึกฝนมาแล้วเช่น กลุ่ม ภาพเรขาคณิต 16 ( Visual Geometry Group : VGG16) และอัลกอริทึมอื่นๆ เพื่อให้ ทำให้ แบบจำลองมีประสิทธิภาพมากขึ้นและนำมาประยุกต์ใช้กับการคัดแยกผลไม้ในชีวิตจริง

5.2.3 สามารถเพิ่มค่าความถูกต้อง (Accuracy) ให้กับแบบจำลองได้โดยการเพิ่ม จำนวนข้อมูลฝึกฝน (Training data) เพิ่มขึ้น, จำลองข้อมูลเพิ่มมากขึ้นจากข้อมูลที่ได้มาที่อาจจะ ผิดไปในบางภาพเพื่อใช้ในการฝึกฝนแบบจำลอง และการจำลองรูปแบบของทุกรูปภาพที่มีการ หมุน (Rotate) ในทุกแบบ

5.2.4 นำแบบจำลองจากงานวิจัยนี้ทดสอบกับชนิดของผลไม้ชนิดอื่นในงานวิจัย อนาคตเพื่อทดสอบความมีประสิทธิภาพของแบบจำลอง

## ภาคผนวก

1. ข้อมูลชุดฝึกฝน (Training Set) ที่ประกอบด้วยภาพผลไม้ทั้งหมด 6,404 รูป

```
train_img_dir_n = "../input/test-sprite/training_set/Apple"
train_img_paths_n = [join(train_img_dir_n, filename) for filename in os.listdir(train_img_dir_n)]
```

2. นำชุดข้อมูลทดสอบ (Test Set) ที่ประกอบด้วยภาพผลไม้แอปเปิ้ลและไม่ใช่ภาพผลไม้แอปเปิ้ลจำนวน 2,134 รูปและ 20,554 รูปตามลำดับ

```
test_img_dir_n = "../input/test-sprite/test_set/Other"
test_img_paths_n = [join(test_img_dir_n, filename) for filename in os.listdir(test_img_dir_n)]
```

3. หลังจากนำข้อมูลเข้าแล้วทำการเรียกข้อมูลในชุดทดสอบ (Testing data) มาตรวจสอบถึงจำนวนทั้งหมด

```
print('Distribution of Image Types in Test Set')
print(test_path_df['image_type'].value_counts())
```

```
Distribution of Image Types in Test Set
Other      20554
Apple       2134
Name: image_type, dtype: int64
```

4. เมื่อขั้นตอนการนำเข้าข้อมูลในชุดฝึกฝน (Training data) และชุดข้อมูลทดสอบ (Testing) เรียบร้อยแล้ว ขั้นตอนถัดมาคือทำการเตรียมข้อมูลชุดฝึกฝน (Training data) และชุดข้อมูลทดสอบ (Testing) เพื่อนำเข้าโครงข่ายประสาทเทียมแบบลึกที่เรียกว่า ResNet50 โดยในที่นี้กำหนดขนาดของรูปเพื่อนำเข้าที่ 100

```
# prepare images for resnet50
image_size = 100

def read_and_prep_images(img_paths, img_height=image_size, img_width=image_size):
    imgs = [load_img(img_path, target_size=(img_height, img_width)) for img_path in img_paths]
    img_array = np.array([img_to_array(img) for img in imgs])
    #output = img_array
    output = preprocess_input(img_array)
    return(output)

X_train = read_and_prep_images(train_img_paths_n)
X_test = read_and_prep_images(all_test_paths)
```

5. ถัดมาคือขั้นตอนการนำข้อมูลจากการเตรียมในขั้นตอนที่แล้วเข้าโครงข่ายประสาทเทียมแบบลึก (ResNet50) ใช้สำหรับการดึงคุณลักษณะเด่นของภาพออกมาทั้งชุดข้อมูลชุดฝึกฝน (Training data) และชุดข้อมูลทดสอบ (Testing) เมื่อได้ค่าออกมาจะทำให้มิติมีขนาดสูงขึ้น

```
# get features from resnet50

resnet_weights_path = 'imagenet'

# X : images numpy array
resnet_model = ResNet50(input_shape=(image_size, image_size, 3), weights=resnet_weights_path, include_top=False, pooling='avg') # Since top layer is the fc layer used for predictions

X_train = resnet_model.predict(X_train)
X_test = resnet_model.predict(X_test)
```

6. ถัดมาคือการทำ Standard Scaler เพื่อให้ภาพที่ออกมาจาก ResNet50 อยู่ในมาตรฐานเดียวกันสามารถเปรียบเทียบกันได้

```
# Apply standard scaler to output from resnet50
ss = StandardScaler()
ss.fit(X_train)
X_train = ss.transform(X_train)
X_test = ss.transform(X_test)
```

7. ขั้นตอนถัดมาทำการวิเคราะห์องค์ประกอบหลัก (Principles Component Analysis: PCA) เป็นการทำให้องค์ประกอบที่สัมพันธ์กันมาอยู่รวมกันทำให้ข้อมูลเล็กลง ในงานวิจัยนี้ใช้สำหรับลดมิติ (Dimension) ของข้อมูลลงหลังจากผ่านขั้นตอนการเตรียมข้อมูลแล้วขั้นตอนถัดไปคือการนำ

ข้อมูลเข้าแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest) ต่อไป

```
# Take PCA to reduce feature space dimensionality
pca = PCA(n_components=512, whiten=True)
pca = pca.fit(X_train)
print('Explained variance percentage = %0.2f' % sum(pca.explained_variance_ratio_))
X_train = pca.transform(X_train)
X_test = pca.transform(X_test)
```

8. ขั้นตอนถัดมาคือการกำหนดไฮเปอร์พารามิเตอร์ของแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine) และ แบบจำลองป่าแยก (Isolation Forest)

```
# Train classifier and obtain predictions for OC-SVM
oc_svm_clf = svm.OneClassSVM(gamma=0.001, kernel='rbf', nu=0.08) # Obtained using grid search
if_clf = IsolationForest(contamination=0.08, max_features=1.0, max_samples=1.0, n_estimators=40) # Obtained using grid search

oc_svm_clf.fit(X_train)
if_clf.fit(X_train)

oc_svm_preds = oc_svm_clf.predict(X_test)
if_preds = if_clf.predict(X_test)
```

9. ผลลัพธ์ประสิทธิภาพการทำนายแบบจำลองจำแนกข้อมูลด้วยเครื่องเวกเตอร์ค้ำยันประเภทเดียว (One-Class Support Vector Machine)

```
print('roc auc score: oc_svm_preds')
oc_svm_preds=svm_if_results['oc_svm_preds']
actual=svm_if_results['is_apple']
print(roc_auc_score(actual, oc_svm_preds))
print(classification_report(actual, oc_svm_preds))
sns.heatmap(confusion_matrix(actual, oc_svm_preds),annot=True,fmt='2.0f')
plt.show()
```

```
roc auc score: oc_svm_preds
0.842965664106225
```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.97      | 0.93   | 0.95     | 20554   |
| 1            | 0.53      | 0.75   | 0.62     | 2134    |
| accuracy     |           |        | 0.91     | 22688   |
| macro avg    | 0.75      | 0.84   | 0.79     | 22688   |
| weighted avg | 0.93      | 0.91   | 0.92     | 22688   |

10. ผลลัพธ์ประสิทธิภาพการทำนายแบบจำลองป่าแยก (Isolation Forest)

```

print('roc auc score: if_preds')
if_preds=svm_if_results['if_preds']
actual=svm_if_results['is_apple']
print(roc_auc_score(actual, if_preds))
print(classification_report(actual, if_preds))
sns.heatmap(confusion_matrix(actual, if_preds),annot=True,fmt='.2f')
plt.show()

```

```

roc auc score: if_preds
0.8485864812951826

```

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.98      | 0.92   | 0.95     | 20554   |
| 1            | 0.50      | 0.78   | 0.61     | 2134    |
| accuracy     |           |        | 0.91     | 22688   |
| macro avg    | 0.74      | 0.85   | 0.78     | 22688   |
| weighted avg | 0.93      | 0.91   | 0.92     | 22688   |



## บรรณานุกรม

Disease, P. a. (2016). Storage of fresh fruit and vegetables.

<https://www.agric.wa.gov.au/fruit/storage-fresh-fruit-and-vegetables>

Hara, Y., Fukuyama, Y., Murakami, K., Iizaka, T., & Matsui, T. (2020). *Fault Detection of Hydroelectric Generators using Isolation Forest*. Paper presented at the 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE).

Jintawatsakoon, S., & Charoenruengkit, W. (2019, 20-22 Nov. 2019). *Open-Set Bottle Classifying using a Convolution Neural Network*. Paper presented at the 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE).

Karakaya, D., Ulucan, O., & Turkan, M. (2019, 31 Oct.-2 Nov. 2019). *A Comparative Analysis on Fruit Freshness Classification*. Paper presented at the 2019 Innovations in Intelligent Systems and Applications Conference (ASYU).

Kausar, A., Sharif, M., Park, J., & Shin, D. R. (2018). *Pure-CNN: A Framework for Fruit Images Classification*. Paper presented at the 2018 International Conference on Computational Science and Computational Intelligence (CSCI).

Kumari, R. S. S., & Gomathy, V. (2018). *Fruit classification using statistical features in svm classifier*. Paper presented at the 2018 4th International Conference on Electrical Energy Systems (ICEES).

Lehnert, R. (2015). How many apple varieties are too many?

<https://www.goodfruit.com/how-many-apple-varieties-are-too-many/>

Lu, S., Lu, Z., Aok, S., & Graham, L. (2018, 19-21 Nov. 2018). *Fruit Classification Based on Six Layer Convolutional Neural Network*. Paper presented at the 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP).

Perera, P., & Patel, V. M. (2019). *Deep transfer learning for multiple class novelty detection*. Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Ponce, J. M., Aquino, A., & Andujar, J. M. (2019). Olive-Fruit Variety Classification by

- Means of Image Processing and Convolutional Neural Networks. *IEEE Access*, 7, 147629-147641.
- Prati, R. C., Batista, G. E. A. P. A., & Monard, M. C. (2008). Evaluating classifiers using ROC curves. *IEEE Latin America Transactions*, 6(2), 215-222.
- Septiarini, A., Hamdani, H., Hatta, H. R., & Kasim, A. A. (2019, 23-24 Oct. 2019). *Image-based processing for ripeness classification of oil palm fruit*. Paper presented at the 2019 5th International Conference on Science in Information Technology (ICSITech).
- Shahbudin, S., Zamri, M., Kassim, M., Abdullah, S. A. C., & Suliman, S. I. (2017, 9-10 Nov. 2017). *Weed classification using one class support vector machine*. Paper presented at the 2017 International Conference on Electrical, Electronics and System Engineering (ICEESE).
- Sidehabi, S. W., Suyuti, A., Areni, I. S., & Nurtanio, I. (2018). *Classification on passion fruit's ripeness using K-means clustering and artificial neural network*. Paper presented at the 2018 International Conference on Information and Communications Technology (ICOIACT).
- Vaishnav, D., & Rao, B. R. (2018, 15-16 Nov. 2018). *Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool*. Paper presented at the 2018 3rd International Conference on Inventive Computation Technologies (ICICT).
- Wibawa, H. A. (2019). *Snake Fruit Classification by Using Histogram of Oriented Gradient Feature and Extreme Learning Machine*. Paper presented at the 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS).
- Zhang, K., Kang, X., & Li, S. (2019). *Isolation forest for anomaly detection in hyperspectral images*. Paper presented at the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium.
- Zhu, D., Wang, M., Zou, Q., Shen, D., & Luo, J. (2019, 25-27 Oct. 2019). *Research on Fruit Category Classification Based on Convolution Neural Network and Data Augmentation*. Paper presented at the 2019 IEEE 13th International Conference on

Anti-counterfeiting, Security, and Identification (ASID).





## ประวัติผู้เขียน

|                   |                                                                                                                                                    |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| ชื่อ-สกุล         | นายศรัทธัย โชติชาติมาลา                                                                                                                            |
| วัน เดือน ปี เกิด | 22 ตุลาคม 2533                                                                                                                                     |
| สถานที่เกิด       | กรุงเทพมหานคร                                                                                                                                      |
| วุฒิการศึกษา      | พ.ศ.2552<br>บัญชีบัณฑิต สาขาวิชาบัญชี<br>จาก จุฬาลงกรณ์มหาวิทยาลัย<br>วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล<br>จาก มหาวิทยาลัยศรีนครินทรวิโรฒ |

