



การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีน  
โดยใช้เทคนิคการทำเหมืองข้อมูล

THE ANALYSIS OF DURIAN RIPENING LEVELS FROM DIFFUSION OF  
ETHYLENE GASES BY USING DATA MINING TECHNIQUES



โดย  
อุดมศักดิ์ เพียรไพโรจน์

วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตาม  
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
วิทยาลัยนวัตกรรมการผลิตและเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย มหาวิทยาลัยรังสิต

ปีการศึกษา 2561



**THE ANALYSIS OF DURIAN RIPENING LEVELS FROM DIFFUSION OF  
ETHYLENE GASES BY USING DATA MINING TECHNIQUES**

**BY  
UDOMSAK PAEANPAIROJ**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENT FOR  
THE DEGREE OF MASTER OF SCIENCE IN INFORMATION  
TECHNOLOGY  
COLLEGE OF DIGITAL INNOVATION AND INFORMATION  
TECHNOLOGY**

**GRADUATE SCHOOL, RANGSIT UNIVERSITY  
ACADEMIC YEAR 2018**

วิทยานิพนธ์เรื่อง

การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีน

โดยใช้เทคนิคการทำเหมืองข้อมูล

โดย

อุดมศักดิ์ เพ็ชรไพโรจน์

ได้รับการพิจารณาให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยรังสิต

ปีการศึกษา 2561

ผศ.ดร.เชษฐาเนติ ศรีสอาน  
ประธานกรรมการสอบ

รศ.ดร.ชม กัมปาน  
กรรมการ

ผศ.ดร.วุฒิพงษ์ ชินศรี  
กรรมการและอาจารย์ที่ปรึกษา

บัณฑิตวิทยาลัยรับรองแล้ว

(ผศ.ร.ต.หญิง ดร.วรรณิ์ สุขสาตร)

คณบดีบัณฑิตวิทยาลัย

21 มกราคม 2562

Thesis Entitled

**THE ANALYSIS OF DURIAN RIPENING LEVELS FROM DIFFUSION OF  
ETHYLENE GASES BY USING DATA MINING TECHNIQUES**

by

UDOMSAK PAEANPAIROJ

was submitted in partial fulfillment of the requirements  
for the degree of Master of Science Program in Information Technology

Rangsit University  
Academic Year 2018

---

Asst.Prof. Chetneti Srisaad,Ph.D  
Examination Committee Chairperson

Assoc.Prof. Chom Kimpan, Ph.D.  
Member

---

Asst.Prof. Wutthipong Chinnasri, Ph.D.  
Member and Advisor

Approved by Graduate School

(Asst.Prof.Plт.Off. Vannee Sooksatra, D.Eng.)

Dean of Graduate School

January 21, 2019

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความกรุณาของ ผศ.ดร.วุฒิพงษ์ ชินศรี ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ โดยอาจารย์ได้กรุณามอบวิชาความรู้ในด้านการศึกษาและ ด้านการพัฒนาคุณภาพชีวิต ทั้งยังให้ความช่วยเหลือในด้านต่าง ๆ รวมไปถึงเป็นกำลังใจและแรงผลักดันให้ กระผมผ่านพ้นอุปสรรคนานับประการจนสำเร็จได้ด้วยดี

ขอขอบคุณบิดา มารดา ผู้ให้กำเนิด ครอบครัวเพียรไฟโรจน์ และภรรยาผู้วิจัย ที่ยอมเสียสละเวลาส่วนตัวให้การสนับสนุนด้านการศึกษาดำเนินการจนทำให้การทำวิจัยประสบความสำเร็จ

ขอขอบคุณสถาบันวิชาการป้องกันประเทศ กองบัญชาการกองทัพไทย ที่เป็นสื่อกลางในการมอบทุนการศึกษาในระดับปริญญาโท

ขอขอบคุณท่าน ดร.อาทิตย์ อุไรรัตน์ อธิการบดีมหาวิทยาลัยรังสิต ที่มอบทุนการศึกษาเรียนดีให้กับข้าพเจ้าจนสำเร็จการศึกษาระดับปริญญาโทนี้ ซึ่งข้าพเจ้าจะนำความรู้นี้ไปใช้พัฒนากองทัพและประเทศชาติสืบต่อไป

อุดมศักดิ์ เพียรไฟโรจน์  
ผู้วิจัย

**5906158 : MAJOR: INFORMATION TECHNOLOGY; M.S.(INFORMATION TECHNOLOGY)**

**KEYWORDS : ETHYLENE, RIPENING, GAS SENSOR, DATA MINING, CLUSTERING, CLASSIFICATION**

**UDOMSAK PAEANPAIROJ: THE ANALYSIS OF DURIAN RIPENING LEVELS FROM DIFFUSION OF ETHYLENE GASES BY USING DATA MINING TECHNIQUES. THESIS ADVISOR: ASST.PROF.WUTTHIPONG CHINNASRI, Ph.D , 80 p.**

The level of ripening in durian is an important factor for operators or distributors in planning, storing, transporting, as well as distributing. To determine the ripening levels, it is required to penetrate skin of durian. This penetration, however, appears to cause certain damage to durians. As it was evident that ethylene is associated with fruit ripening, this research aims to analyze the ripening levels in durians from ethylene gas diffusion using data mining techniques. The Arduino board connected with MQ3, MQ6 and MQ8 detection sensors was developed as the assessment tool. Once the data had been collected, the data mining procedures were performed for data clustering. The findings indicated that the K-Means clustering algorithm was able to classify four levels of ripening in durian as follows: (1) unripe, (2) pre-ripe, (3) ripe, and (4) overripe. This resonates with what most Thais commonly understand about the ripening levels of durian or banana. In addition, The classification technique was used for the ripening analysis. In terms of classifications, various types of algorithm, for example, Decision Trees, K Nearest Neighbors (KNN), Neural Networks, Naive Bayes, and Support Vector Machine (SVM). were used to compare the accuracy rate of their performance through the cross-validation procedure. It is shown that the SVM and the Neural Network algorithms were the most accurate. The accuracy of the algorithms in this experimental was almost at 100 percent. Therefore, these algorithms can be further developed to analyze the fruit ripening level from analysis of gas diffusion.

Student's Signature ..... Thesis Advisor's Signature .....

5906158 : สาขาวิชาเอก: เทคโนโลยีสารสนเทศ; วท.ม.(เทคโนโลยีสารสนเทศ)  
 คำสำคัญ : ก๊าซเอทิลีน, การสุกของผลไม้, เซ็นเซอร์, การทำเหมืองข้อมูล, การแบ่งกลุ่ม  
 ของข้อมูล, การจำแนกข้อมูล

อุดมศักดิ์ เพียรไพโรจน์ : การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของ  
 ก๊าซเอทิลีนโดยใช้เทคนิคการทำเหมืองข้อมูล (THE ANALYSIS OF DURIAN RIPENING  
 LEVELS FROM DIFFUSION OF ETHYLENE GASES BY USING DATA MINING  
 TECHNIQUES) อาจารย์ที่ปรึกษา: ผศ.ดร.วุฒิพงษ์ ชินศรี, จำนวนหน้าทั้งหมด 80 หน้า.

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของ  
 ก๊าซเอทิลีน โดยใช้เทคนิคการทำเหมืองข้อมูล เนื่องจากระดับการสุกของทุเรียนเป็นปัจจัยสำคัญที่  
 ผู้ประกอบการหรือผู้จัดจำหน่ายใช้ในการวางแผนจัดเก็บ ขนส่ง และจัดจำหน่าย ซึ่งการตรวจสอบ  
 ระดับการสุกจำเป็นต้องเจาะทำลายผิวทุเรียนบางส่วนเพื่อตรวจสอบระดับการสุก ทำให้ผลทุเรียน  
 เกิดความเสียหาย ดังนั้นเมื่อผู้วิจัยทราบว่าก๊าซเอทิลีน มีความสัมพันธ์กับระดับการสุกของผลไม้  
 ผู้วิจัยจึงพัฒนาอุปกรณ์วัดระดับการสุกของทุเรียนโดยใช้ Arduino เชื่อมต่อกับเซ็นเซอร์ MQ3, MQ6  
 และ MQ8 เพื่อตรวจสอบการแพร่กระจายของก๊าซเอทิลีน เมื่อได้ข้อมูลจากเซ็นเซอร์แล้ว ผู้วิจัยได้ใช้  
 เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering) ในการแบ่งกลุ่มทุเรียนตามระดับการสุก โดยใช้อัลกอริทึม  
 K-Mean มาใช้ในการทดลอง ซึ่งผลการทดลองสามารถแบ่งกลุ่มระดับการสุกได้เป็น 4 กลุ่ม คือ  
 1) ดิบ 2) ห่าม 3) สุก และ 4) งอม ซึ่งเป็นไปตามลักษณะการสุกของทุเรียน หรือกล้วยที่คนไทยส่วนใหญ่  
 เข้าใจกัน สำหรับการวิเคราะห์ระดับการสุกของทุเรียนนั้น ผู้วิจัยใช้เทคนิคการจำแนกข้อมูล (Classification)  
 ในการสร้างโมเดล และทำการเปรียบเทียบอัลกอริทึมที่นำมาใช้ในการสร้างโมเดล ได้แก่ Decision Tree,  
 K Nearest Neighbors (KNN), Neural Network, Naive Bayes และ Support Vector Machine (SVM)  
 โดยเปรียบเทียบค่าความแม่นยำในการจำแนกข้อมูล (Accuracy) โดยการแบ่งกลุ่มของข้อมูลเพื่อทำ  
 การทดสอบใช้แบบ Cross Validation ผลการทดลอง พบว่า Support Vector Machine และ Neural  
 Network มีค่าความแม่นยำในการจำแนกข้อมูลสูงที่สุดใกล้เคียงกัน นอกจากนี้ โมเดลที่ได้จากทุก  
 อัลกอริทึมมีค่าความแม่นยำในการจำแนกข้อมูลสูงเกือบ 100% จึงสามารถนำโมเดลเหล่านั้นไปใช้  
 เพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีนได้

## สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อภาษาไทย	ข
บทคัดย่อภาษาอังกฤษ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
<b>บทที่ 1</b>	
<b>บทนำ</b>	<b>1</b>
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์การวิจัย	1
1.3 คำถามการวิจัย	2
1.4 กรอบแนวคิดการวิจัย	3
1.5 นิยามศัพท์	3
<b>บทที่ 2</b>	
<b>ทบทวนวรรณกรรมที่เกี่ยวข้องและทฤษฎีที่เกี่ยวข้อง</b>	<b>5</b>
2.1 ทฤษฎีที่เกี่ยวข้อง	5
2.2 งานวิจัยที่เกี่ยวข้อง	27
<b>บทที่ 3</b>	
<b>ระเบียบวิธีการวิจัย</b>	<b>30</b>
3.1 การพัฒนาเครื่องมือการตรวจจับก๊าซเอทีลิน	30
3.2 เพื่อกำหนดกลุ่มของระดับการสูดของทุเรียน	33
3.3 การจำแนกประเภทการสูดของทุเรียนจากการแพร่กระจายของก๊าซ	34
<b>บทที่ 4</b>	
<b>ผลการวิจัย</b>	<b>37</b>
4.1 ผลการพัฒนาเครื่องมือการตรวจจับก๊าซเอทีลิน	37
4.2 ผลการกำหนดกลุ่มของระดับการสูดของทุเรียน	42
4.3 ผลการจำแนกประเภทการสูดของทุเรียนจากการแพร่กระจายของก๊าซ	45



## สารบัญ (ต่อ)

	หน้า
<b>บทที่ 5</b>	
<b>สรุปผลและข้อเสนอแนะ</b>	<b>52</b>
5.1 สรุปผลการวิจัยและอภิปรายผล	52
5.2 ข้อเสนอแนะ	54
<b>บรรณานุกรม</b>	<b>55</b>
<b>ภาคผนวก</b>	<b>59</b>
<b>ผนวก ก</b> บทความที่ได้รับการตีพิมพ์ในการประชุมวิชาการระดับชาติมหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9	<b>60</b>
<b>ผนวก ข</b> เกียรติบัตรในการแสดงการเข้าร่วมนำเสนอผลงานภาคบรรยายในการประชุมวิชาการระดับชาติมหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9	<b>74</b>
<b>ผนวก ค</b> อุปกรณ์ที่ใช้ในการพัฒนาเครื่องมือสำหรับทดสอบระดับความสุขของทุเรียน	<b>76</b>
<b>ประวัติผู้วิจัย</b>	<b>80</b>

## สารบัญตาราง

ตารางที่	หน้า	
4.1	ตารางแสดงการจัดกลุ่มความเข้มข้นของก๊าซต่างๆ	41
4.2	ตารางแสดงความสัมพันธ์ระหว่างกลุ่มความเข้มข้นปริมาณก๊าซกับระดับการสูดของทุเรียน	43
4.3	ตารางแสดงค่า K ในการจัดกลุ่มด้วย K-Mean	44
4.4	แสดงพารามิเตอร์และค่าพารามิเตอร์ทำการปรับค่า	46
4.5	แสดงการเปรียบเทียบประสิทธิภาพของอัลกอริทึม	48



## สารบัญรูป

รูปที่	หน้า	
2.1	แสดงสูตรทางเคมีของเอทิลีน	9
2.2	ภาพแสดง Arduino	10
2.3	ภาพแสดงโปรแกรมเริ่มต้นของ Arduino	12
2.4	แสดงโครงสร้างของแก๊สเซนเซอร์	13
2.5	แสดงตัวอย่างภาษา C	15
2.6	ภาพแสดงระยะห่างยูคลิดีเนียน	18
2.7	ภาพแสดงโครงสร้างของเครือข่ายประสาทเทียม	23
2.8	ภาพแสดงการเรียนรู้ของ Support Vector Machine	26
3.1	ภาพแสดงสูตรทางเคมีของเอทิลีน	31
3.2	ภาพแสดง MQ Gas Sensor	32
3.3	ภาพแสดงอาคูโน	32
4.1	ภาพแสดงอุปกรณ์วัดระดับการสุกของทุเรียน	37
4.2	ภาพแสดงโปรแกรมกำหนดค่าและสั่งงานอุปกรณ์	38
4.3	ภาพแสดงปริมาณก๊าซออกมาเป็นลักษณะกราฟ และตัวเลข	39
4.4	ภาพแสดงปริมาณข้อมูลที่เก็บได้จากทุเรียน	40

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ประเทศไทยเป็นประเทศอยู่ในโซนเอเชียและเป็นพื้นที่เขตร้อนชื้นซึ่งเอื้ออำนวยต่อการเจริญเติบโตของผลไม้จึงมีผลไม้มากมายหลายชนิด ซึ่งผลไม้นั้นสามารถจัดแบ่งประเภทได้หลายรูปแบบขึ้นอยู่กับ การเลือกปัจจัยที่จะนำมาใช้แบ่งประเภท ซึ่งในงานวิจัยนี้สนใจเกี่ยวกับการวิเคราะห์ระดับการสุกของผลไม้

ปัญหาในการจัดเก็บ การขนส่ง และการซื้อขายผลไม้นั้น มักเกิดกับผลไม้นั้นที่สามารถบ่มสุกได้ เพราะถ้าเกษตรกร ผู้ประกอบการ และผู้จำหน่ายวางแผนในการจัดเก็บและดำเนินการขนส่งได้ไม่สมบูรณ์ จะทำให้ผลไม้นั้นนำไปจำหน่ายให้กับผู้บริโภคไม่ตรงตามความต้องการ ทำให้ต้องสูญเสียเวลาและค่าใช้จ่ายในการจัดเก็บสินค้าเพิ่มเติม ซึ่งมีงานวิจัยหลายชิ้นที่ศึกษาเกี่ยวกับการควบคุมระดับการสุกของผลไม้ในขั้นตอนการจัดเก็บและการขนส่งสินค้า (Wills, Harris, Spohr, & Golding, 2014; Blanke, 2014; Noiwan, Kumpoun, Whangcha, Saengnil, & Uthaibutra, 2006) อย่างไรก็ตามในมุมมองของผู้วิจัยนั้นผลไม้นั้นที่สามารถบ่มสุกได้และสามารถมองเห็นการเปลี่ยนแปลงสีผิวของผลไม้ได้ด้วยเอกลักษณ์ (Non-Scientific) เกษตรกรหรือผู้จำหน่ายผลไม้ยังพอที่จะระบุได้ว่าผลไม้นั้นมีระดับการสุกมากน้อยเพียงใด โดยอาศัยประสบการณ์และการสังเกตสีผิวของผลไม้ แต่ผลไม้นั้นที่สามารถบ่มสุกได้และสีผิวไม่เปลี่ยนแปลงหรือเปลี่ยนแปลงเล็กน้อย เช่นทุเรียน ขนุน เกษตรกรหรือผู้จำหน่ายต้องอาศัยประสบการณ์และการสังเกตเป็นอย่างมาก ผู้วิจัยจึงมีแนวคิดที่จะทำการวิเคราะห์ระดับการสุกของผลไม้จากการแพร่กระจายของก๊าซโดยใช้เทคนิคการทำเหมืองข้อมูล โดยเลือก ทุเรียน มาเป็นตัวแทนของผลไม้นั้นที่สามารถบ่มสุกได้และไม่สามารถสังเกตจากภายนอกได้ว่ามีระดับการสุกมากน้อยเพียงใดมาใช้ในการทดลอง

### 1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อพัฒนาเครื่องมือตรวจจับก๊าซเอทิลีนจากทุเรียน

1.2.2 เพื่อทดลองแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทีลิน โดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering)

1.2.3 เพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลินโดยใช้เทคนิคการจำแนกข้อมูล (Classification)

### 1.3 คำถามการวิจัย

ในการวิจัยนี้ ผู้วิจัยมีความสนใจที่จะวิเคราะห์ระดับการสุกของผลไม้จากการแพร่กระจายของก๊าซและใช้ปริมาณก๊าซจากทุเรียนเป็นตัวแทนผลไม่บ่มสุกจากนั้นจะใช้เทคนิคการทำเหมืองข้อมูล โดยได้กำหนดขอบเขตในการวิจัยไว้ ดังนี้

#### 1.3.1 กลุ่มข้อมูลที่ใช้เพื่อการวิจัย

- 1.3.1.1 ทุเรียนสายพันธุ์หมอนทอง ในโซนภาคตะวันออกเฉียงเหนือของประเทศไทย
- 1.3.1.2 ลักษณะทุเรียนที่ใช้ในการจำหน่ายจากร้านค้าผู้บริโภคนับจำนวน 160 ลูก
- 1.3.1.3 ข้อมูลจากก๊าซที่เกี่ยวข้องกับกระบวนการสุกของทุเรียน
- 1.3.1.4 การวิจัยนี้วิเคราะห์แยกกลุ่ม ประเภทการสุกของทุเรียนจากการแพร่กระจายของก๊าซ

#### 1.3.2 กลุ่มอุปกรณ์ประเภทฮาร์ดแวร์ เพื่อการวิเคราะห์งานวิจัย

- 1.3.2.1 อุปกรณ์ในกลุ่ม MQ Gas Sensor
- 1.3.2.2 Arduino
- 1.3.2.3 พัดลมดูดอากาศ
- 1.3.2.4 กล่องพลาสติกสำหรับใส่ทุเรียนเพื่อการวิเคราะห์

#### 1.3.3 กลุ่มเครื่องมือประเภทซอฟต์แวร์ เพื่อการวิเคราะห์งานวิจัย

- 1.3.3.1 Rapid Miner สำหรับวิเคราะห์ข้อมูลเพื่อค้นหาความสัมพันธ์ จัดกลุ่มและแบ่งประเภทของข้อมูล
- 1.3.3.2 ภาษา C สำหรับสั่งงานให้อุปกรณ์ทำงานตามระบบและรูปแบบของงานวิจัย

## 1.4 กรอบแนวคิดการวิจัย

- 1.4.1 ศึกษาข้อมูลระดับปริมาณก๊าซเอทิลีนที่มีความสัมพันธ์กับอาการสุกของทุเรียน
- 1.4.2 ศึกษาเทคนิคและงานวิจัยที่เกี่ยวข้อง
- 1.4.3 พัฒนาเครื่องมือตรวจจับก๊าซเอทิลีนและเขียนโปรแกรม
- 1.4.4 เก็บรวบรวมข้อมูลปริมาณก๊าซจากกล่องเครื่องมือตรวจจับก๊าซ
- 1.4.5 ทำการทดลองกำหนดกลุ่มของระดับการสุกของทุเรียน
- 1.4.6 ทำการเปรียบเทียบประสิทธิภาพของ Algorithms ในการจำแนกประเภทการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีน
- 1.4.7 ทำการวิเคราะห์และสรุปผลการทดลอง
- 1.4.8 จัดทำเอกสารและสรุปผล

## 1.5 นิยามศัพท์เฉพาะ

ประเภทของผลไม้ สามารถแบ่งเป็น 2 ประเภท (Kaushlendra, Saurabh, Murtaza, & Tanushri, 2015) ตามอัตราการหายใจ ได้แก่ ผลไม้ที่มีอัตราการหายใจเปลี่ยนแปลงตามอายุหลังเก็บเกี่ยวผลไม้ที่แก่แล้ว (Climacteric Fruit) หรือสามารถบ่มให้สุกได้ เช่น กัลย มะม่วง มะละกอ ทุเรียน และผลไม้ที่มีอัตราการหายใจไม่เปลี่ยนแปลงตามอายุหรือเปลี่ยนแปลงตามอายุในอัตราที่น้อยมากหลังเก็บเกี่ยวผลไม้ที่แก่แล้ว (Non-Climacteric Fruit) หรือไม่สามารถบ่มให้สุกได้ เช่น ส้ม สับปะรด มะนาว และเงาะ

**Ethylene Gas** คือ ก๊าซที่ปนอยู่ในผลไม้ทุกชนิดและถือเป็นปัจจัยหลักของการดำรงชีวิตของผลไม้ ดังนั้นเมื่อผลไม้อยู่ในระหว่างอาการเปลี่ยนแปลงตามสภาวะจากอาการดิบไปจนถึงอาการอม(ชราภาพ) นั้นเอทิลีนจะเป็นส่วนสำคัญในการเปลี่ยนแปลงสภาพนี้โดยสามารถพิจารณาจากจำนวนความเข้มข้นของก๊าซได้ แต่ผลไม้แต่ละชนิดจะปล่อยก๊าซเอทิลีนออกมาแตกต่างกันขึ้นอยู่กับชนิดและสายพันธุ์ของผลไม้

**MQ Gas Sensor** เป็นเครื่องมืออิเล็กทรอนิกส์ด้วยวิธีการออกไซด์ของโลหะทำหน้าที่เป็นชั้นตรวจจับปริมาณก๊าซ โดยเซนเซอร์ MQ3 MQ6 และ MQ8 นี้จะทำให้ทราบถึงปริมาณความเข้มข้นของก๊าซนั้น ๆ ตามลักษณะความต้องการใช้งาน

**Arduino** คือ ไมโครคอนโทรลเลอร์ตระกูลต่าง ๆ มาใช้ร่วมกันในภาษา C ซึ่งภาษา C นี้เป็นลักษณะเฉพาะ คือมีการเขียนไวยากรณ์ของ Arduino ขึ้นมาเพื่อให้การสั่งงานไมโครคอนโทรลเลอร์ที่แตกต่างกันสามารถใช้งานโค้ดตัวเดียวกันได้ โดยตัวโครงการ ได้ออกบอร์ดทดลองมาหลาย ๆ รูปแบบ เพื่อใช้งานกับเครื่องมือที่ช่วยในการพัฒนาโปรแกรม (Integrated Development Environment) โดยมีสิ่งอำนวยความสะดวกต่าง ๆ เช่น คำสั่ง Compile, Run ตัวอย่างเช่น NetBeans Editplus, JCreator และ Eclipse สาเหตุหลักที่ทำให้ Arduino เป็นนิยมมาก เป็นเพราะซอฟต์แวร์ที่ใช้งานร่วมกันสามารถ โหลดได้ฟรี ใช้งานง่าย และกระบวนการไม่ซับซ้อน

**การทำเหมืองข้อมูล (Data Mining)** เป็นเทคนิคการค้นหารูปแบบหรือเปรียบเทียบ (Set) ของกลุ่มข้อมูลจำนวนมาก โดยใช้หลักการคำนวณหาความคล้ายคลึง แบ่งประเภท จัดกลุ่มจากวิชา สถิติ คณิตศาสตร์ จากการเรียนรู้ของเครื่อง และการจดจำแบบหรือเกิดขึ้นบ่อยครั้ง การทำเหมืองข้อมูล คือ กระบวนการที่นำข้อมูลมาเพื่อค้นหาแนวทาง ความสัมพันธ์ที่มีอยู่ในข้อมูลและนำไปประยุกต์ใช้ ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม อาจเป็นวิวัฒนาการของการจัดเก็บข้อมูล ในรูปแบบฐานข้อมูลที่ สามารถดึงข้อมูลสารสนเทศมาใช้ และจะนำผลลัพธ์ที่ได้ไปใช้ประโยชน์ การทำเหมืองข้อมูลจึงมีรูปแบบ วิธีการ และกระบวนการให้เลือกใช้ได้ตามความเหมาะสม หรือตามเงื่อนไขทั้งในลักษณะ การทำแบบอัตโนมัติและการปรับจูน Parameter เอง ในการทดลองนี้ผู้วิจัยจะนำเทคนิคของ Clustering มาใช้แบ่งกลุ่มและ Classification มาใช้วิเคราะห์ โดยเลือก Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naïve Bayes และ Support Vector Machine (SVM) ในการเปรียบเทียบ และจำแนกประเภท เป็นต้น



## บทที่ 2

### ทบทวนวรรณกรรมและทฤษฎีที่เกี่ยวข้อง

การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซโดยใช้เทคนิคการทำเหมืองข้อมูล เป็นการศึกษาข้อมูลระดับการแพร่กระจายของก๊าซที่มีอยู่ในทุเรียนและงานวิจัยที่เกี่ยวข้องในการทำนายผลด้วยเทคนิคการทำเหมืองข้อมูล ประกอบด้วยทฤษฎี และงานวิจัยที่เกี่ยวข้อง ดังนี้

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 ประวัติและความเป็นมา

ในกระบวนการเติบโตของพืชและผลไม้ตามธรรมชาตินั้น เริ่มต้นตั้งแต่กำเนิดจากเมล็ดเข้าสู่กระบวนการทางธรรมชาติด้วยการแบ่งเซลล์ เพิ่มขนาดของเซลล์ เปลี่ยนรูปร่างของเซลล์ ตลอดจนงอกแทงรากฝอยลงสู่ดินเพื่อดูดธาตุต่าง ๆ นำมาสังเคราะห์เป็นอาหารนั้นโดยในงานวิจัยวัฏจักรชีวิตพืชดอกนั้นกล่าวไว้ว่าพืชที่เริ่มจากเมล็ดพันธุ์ที่งอกมาเป็นต้นพืช เจริญเติบโต ออกดอก ออกผลสร้างเมล็ด ตลอดจนดำเนินถึงขั้นตอนสุกและแก่ในเวลาต่อมาพร้อมให้เก็บเกี่ยว นั้น เป็นแค่ผิวเผินที่มนุษย์เห็นคุณค่าแต่ในความเป็นจริงแล้วยังมีรายละเอียดอื่น ๆ ที่น่าสนใจแทรกอยู่ในระยะการเจริญเติบโตของพืชในช่วงระยะการเจริญแก่ต่าง ๆ ของพืชและผลไม้ ซึ่ง กิตติวงศ์พิเชษฐ (2559) ได้กล่าวว่าสิ่งหนึ่งที่สามารถบอกถึงวัฏจักรของพืชผลไม้ได้นั้นคืออายุ โดยมนุษย์จะทราบอายุของพืชผลไม้ได้นั้นจะต้องได้มาจากการคำนวณหรือจากการนับวันโดยต้องเริ่มตั้งแต่วันที่มีการเพาะปลูก ตลอดจนผลไม้เริ่มติดดอกออกผลอ่อน นั้นหมายถึงมนุษย์เราต้องเห็นเหตุการณ์การกำเนิดผลไม้ทั้งหมด สิ่งเหล่านี้ถือเป็นอุปสรรคที่จะเข้าถึงอายุผลไม้ได้ แต่สิ่งหนึ่งที่จะทำให้มนุษย์สามารถเข้าถึงอายุของผลไม้ได้ คือการค้นหาพัฒนาการทางสรีรวิทยาของผลไม้ หรือ การหาอายุการสุกของผลไม้ (Ripening) ซึ่งเป็นกระบวนการแปรสภาพของผลไม้จากลักษณะหนึ่ง ไปสู่อีกลักษณะหนึ่ง ซึ่งผู้วิจัยจะขอเรียกว่า ชราภาพ ในช่วงการเปลี่ยนแปลงนั้นจะมีขั้นตอนและองค์ประกอบหลายปัจจัยเกิดขึ้นทั้งภายในและภายนอกของตัวของผลไม้เอง เช่น ระดับของการหายใจ ระดับปริมาณความเข้มข้นของก๊าซเอทิลีน สีของผิวหน้าด้านนอก ความเหนียวของเปลือก



ความนุ่มของเนื้อด้านใน สิ่งเหล่านี้ล้วนแล้วเป็นปัจจัยสำคัญในการเปลี่ยนแปลงของลักษณะดิบ กลายเป็นลักษณะสุกทั้งสิ้น และอาจมีเรื่องของช่วงเวลาเข้ามาเกี่ยวข้อง เพราะผลไม้บางชนิดมีการเปลี่ยนแปลงรวดเร็วหรือช้าลงต่างกันั้นขึ้นอยู่กับอุณหภูมิ ความสมบูรณ์ซึ่งรวมไปถึงชนิดของผลไม้เองที่มีความแก่เต็มที่หรืออยู่ในช่วงของผลบริบูรณ์ (Maturity) ผลไม้จะดำเนินการแปรสภาพเช่นนี้ไปตามวัฏจักรจนถึงที่สุดจนถึงการย่อยสลาย หากมนุษย์สามารถรู้ขั้นตอนการเปลี่ยนแปลงทางสรีรวิทยาจากการตรวจหาสารเคมี ก๊าซ หรือปัจจัยอื่น ๆ ที่มีอิทธิพลต่อพืชและผลไม้ที่ใช้ในการเจริญเติบโตได้แล้วนั้นก็จะทำให้ทราบถึงอายุ วันเวลาในการเจริญเติบโต รวมไปถึงการวางแผนในการเก็บเกี่ยวผลผลิตและนำไปจัดจำหน่ายได้อย่างมีประสิทธิภาพและเกิดข้อผิดพลาดน้อยที่สุด ทฤษฎีนี้ จะทำให้มนุษย์สามารถนำผลไม้ที่มีลักษณะแก่จัด หรือ ใกล้เคียงอายุมาจัดจำหน่ายได้ก่อน หรือ อาจกล่าวอีกนัยยะหนึ่งได้ว่ามนุษย์อาจจะอยู่เหนือกฎเกณฑ์ทางธรรมชาติของพืชผลไม้ก็เป็นได้ เป็นต้น

## 2.1.2 การแบ่งประเภทของผลไม้

ในช่วงที่ผลไม้อยู่ในลักษณะสุกนั้นจะมีอัตราการหายใจสูง และพบว่ามีปริมาณของก๊าซเอทิลีน (Ethylene) แพร่กระจายอยู่จำนวนมาก นั่นคือลักษณะสำคัญที่ส่งผลจากการพิสูจน์ได้ว่าอัตราการหายใจและปริมาณก๊าซเอทิลีน มีความสัมพันธ์กับลักษณะการสุกของผลไม้ เช่น เมื่อผลไม้ อยู่ในลักษณะดิบปริมาณก๊าซเอทิลีนและการหายใจจะมีต่ำแต่ในทางกลับกันเมื่อผลไม้อยู่ในลักษณะ อากาศสุกจะมีอัตราการหายใจและก๊าซเอทิลีนสูง และสูงขึ้นเรื่อย ๆ เมื่อก้าวเข้าสู่วัยชราภาพ การวิเคราะห์ระดับการสุกของผลไม้จึงขอแบ่งประเภทของผลไม้เป็น 2 ประเภท (Kaushlendra, et al., 2015) ตามอัตราการหายใจ ดังนี้

### 2.1.2.1 ผลไม้บ่มสุก

เป็นผลไม้ที่มีอัตราการหายใจเปลี่ยนแปลงตามอายุหลังเก็บเกี่ยว กล่าวคือผลไม้ที่แก่แล้ว (Climacteric Fruit) หรือสามารถบ่มให้สุกได้ เช่น กล้วย มะม่วง มะละกอ ทุเรียน ผลไม้บ่มสุกนั้นในช่วงก่อนบริบูรณ์จะมีการปล่อยหรือแพร่กระจายของก๊าซเอทิลีนจำนวน น้อย แต่ในช่วงผลบริบูรณ์ ก๊าซเอทิลีนจะเพิ่มจำนวนและแพร่กระจายจำนวนมาก ก๊าซเอทิลีนที่ ผลไม้สร้างขึ้นมีผลต่อกระบวนการสุกของผลไม้ในกลุ่มนี้ หากพิสูจน์ทางชีวเคมี กระบวนการทาง วิทยาศาสตร์จะพบว่าก๊าซเอทิลีนส่งผลต่อการเปลี่ยนแปลงตามสภาวะของผลไม้โดยจะทอดแทรก เข้าไปทำปฏิกิริยาต่อเปลือกผลไม้ทำให้เปลือกอ่อนนุ่มลง เนื้อนุ่มขึ้น กลิ่นที่เพิ่มขึ้นหรือลดลง

ในด้านรสชาติ ก๊าซเอทิลินจะเปลี่ยนแปลงให้เป็นน้ำตาล ซึ่งถือว่าก๊าซเอทิลินมีผลต่อการกระจายพันธุ์ของพืชและผลไม้เป็นอย่างมาก

### 2.1.2.1 ผลไม้บ่มไม่สุก

เป็นผลไม้ที่มีอัตราการหายใจไม่เปลี่ยนแปลงตามอายุหรือเปลี่ยนแปลงตามอายุในอัตราที่น้อยมากหลังเก็บเกี่ยวผลไม้ที่แก่แล้ว (Non-Climacteric Fruit) หรือ ไม่สามารถบ่มให้สุกได้ เช่น ส้ม สับปะรด มะนาว เงาะ เป็นผลไม้ที่อัตราการหายใจของผลไม้เพิ่มขึ้นระหว่างที่ผลไม้แปรสภาพในช่วงก่อนบริบูรณ์ อัตราการหายใจของผลไม้บ่มไม่สุกนั้นจะลดลงเมื่อผลไม้แก่ขึ้น หากเกษตรกรเก็บเกี่ยวผลผลิตในช่วงนั้นๆ หากนำไปบ่มต่อ ผลไม้จะไม่มีอาการไม่สุกต่อได้โดยอาศัยก๊าซเอทิลิน เหตุผลสำคัญคือก๊าซเอทิลินจะไม่สามารถกระตุ้นให้ผลไม้มีการสุกมากขึ้นในทางกลับกันเกษตรกรควรเก็บเกี่ยวผลไม้ในช่วงที่ผลไม้มีการเปลี่ยนแปลงทางสรีรหรือแก่เต็มที่ตามที่กล่าวมาข้างต้น ข้อควรระวังในผลไม้บ่มไม่สุกคือ ก๊าซเอทิลินจะมีประโยชน์ต่อผลไม้ในจำนวนที่พอเหมาะเท่านั้นหากมีปริมาณมากจะส่งผลทำให้ผลไม้เสื่อมสภาพได้

## 2.1.3 ทูเรียน

### 2.1.3.1 ความเป็นมาของทูเรียน

ทูเรียนจัดเป็นไม้ผลยืนต้นขนาดใหญ่ ใบเลี้ยงเดี่ยวในตระกูล Bombacaceae ชื่อสามัญ Durian ชื่อวิทยาศาสตร์ Durio zibethinus Murray ทูเรียนถูกค้นพบครั้งแรกในแถบเอเชียตะวันออกเฉียงใต้ บริเวณหมู่เกาะอินเดีย ทูเรียนจัดอยู่ในกลุ่มพืชใบเลี้ยงเดี่ยวที่มีปลายใบเรียวแหลม ด้านบนสีเขียว ด้านล่างนั้น เป็นสีน้ำตาล เป็นพืชที่มีดอกเกิดเป็นช่อตามลำต้นมีสีขาว ลูกรูปทรงรี มีเปลือกหนามแหลมในลูกเนื้อเป็นสีเหลืองเป็นพดู ด้านในเมล็ดมีสีขาว ไม่มียาง จากการศึกษาของผู้วิจัยค้นพบว่าสารานุกรมไทยสำหรับเยาวชนตามโครงการพระราชประสงค์ในพระบาทสมเด็จพระเจ้าอยู่หัว ซึ่ง หิรัญ หิรัญประดิษฐ์ (2554) ได้กล่าวถึงประวัติของทูเรียนที่เริ่มมีในประเทศไทยคือบันทึกของนายเมอร์ซีเออร์ เดอลาลูแบร์ (Monsieur De La Loubre) นักบวชนิกายเยซูอิต ซึ่งเป็นหัวหน้าคณะราชทูตในพระเจ้าหลุยส์ที่ 14 ของประเทศฝรั่งเศสในขณะนั้นว่ามีอยู่ในประเทศไทยตั้งแต่ในสมัยอยุธยาช่วงปี พ.ศ.2228 ซึ่งเป็นช่วงที่คณะราชทูตฯ เดินทางมาทำสนธิสัญญาทางการค้า โดยในบันทึกเขียนเกี่ยวกับเรื่องของเกษตรกรในประเทศไทยที่ระบุถึงทูเรียนว่า “ดูเรียน (Durion) หรือที่ชาวสยามเรียกว่า ทูลเรียน (Tourrion) เป็นผลไม้ที่นิยมมากในแถบนี้” ต่อมาในสมัยรัตนโกสินทร์ พระยาแพทย์พงศาวิสุทธาธิบดี (สุนทรเวช) ได้กล่าวถึงการแพร่กระจายพันธุ์ของ

ทุเรียนจากจังหวัดนครศรีธรรมราช มายังกรุงเทพมหานคร ตั้งแต่ประมาณ พ.ศ.2318 และมีการทำสวนทุเรียนในตำบลบางกร่างคลองบางกอกน้อย ตอนใน ช่วงปี พ.ศ.2397 จากหลักฐานที่ค้นพบนี้สามารถบอกได้ว่าทุเรียนเป็นผลไม้ที่มีความนิยมของประชาชนคนไทยมาตั้งแต่สมัยอดีตกาล ประกอบกับทุเรียนจัดเป็นผลไม้ชนิดหนึ่งที่ทำให้ประเทศไทยมีชื่อเสียงโด่งดังไปทั่วโลกและได้รับการขนานนามว่าเป็นราชาแห่งผลไม้ (Subhadrabandhu & Ketsa, 2001) โดยผู้บริโภคจะเลือกทานทุเรียนในระดับการสุกที่แตกต่างกันเพราะมีเอกลักษณ์ในเรื่องของรสชาติที่หวานหอมอุดมไปด้วยธาตุอาหารต่าง ๆ ที่มีประโยชน์ต่อร่างกาย เช่น ไนโตรเจน แคลเซียม ฟอสฟอรัส แมงกานีส โพแทสเซียม และกำมะถัน จากการค้นคว้าเพิ่มเติมพบว่า ทุเรียนมีทั้งหมด 227 พันธุ์ สำหรับประเทศไทยแบ่งออกได้เป็น 6 กลุ่ม คือ กลุ่มกบ กลุ่มลวงกลุ่มกำยาน กลุ่มกำป็น กลุ่มทองย้อย กลุ่มเบ็ดเตล็ด และมีทุเรียนที่ประเทศไทย นิยมปลูกเพื่อการส่งออกอีกจำนวน 4 สายพันธุ์ คือ สายพันธุ์หมอนทอง สายพันธุ์กำยาน สายพันธุ์กระดุมทอง และสายพันธุ์ชะนี (มหาวิทยาลัยราชภัฏรำไพพรรณีและเครือข่ายนักวิจัยสาขาพืชในจังหวัดจันทบุรี, 2556)

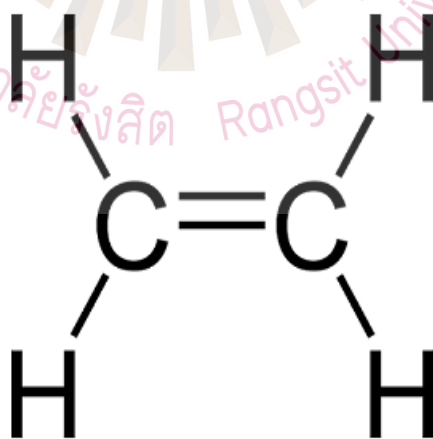
### 2.1.3.2 พฤติกรรมมนุษย์กับการบริโภคทุเรียนในประเทศไทย

ประเทศไทยอยู่ในพื้นที่เขตร้อนชื้นสมบูรณ์ไปด้วยอาหารที่มีรสชาติเป็นเอกลักษณ์ เช่น ส้มตำ ต้มยำกุ้ง อาหารจำพวกเผ็ดร้อน และมีผลไม้ทั้งในฤดูกาลและนอกฤดูกาลให้เลือกรับประทานหลากหลายชนิด อีกทั้งยังสร้างชื่อเสียงให้แก่ประเทศไทยโด่งดังไปทั่วโลก ทำให้สามารถดึงดูดนักท่องเที่ยวจากพื้นที่ต่าง ๆ เดินทางมาท่องเที่ยวประเทศไทยเพื่อเลือกรับประทานอาหารและผลไม้ทั้งในฤดูกาลและนอกฤดูกาลเหล่านี้ แต่สิ่งที่สำคัญและถือว่าเป็นเอกลักษณ์คือ มีตลาดเปิดท้าย (ตลาดนัด) มีลักษณะเป็นเสรีสามารถเดินจับจ่ายใช้สอยได้อย่างเป็นเอกเทศไม่จำกัดเพศ อายุ หรือระยะเวลา สร้างความเป็นกันเอง สนุกสนาน และบันเทิง ซึ่งเป็นที่นิยมของคนไทยเองและเป็นเสน่ห์ดึงดูดนักท่องเที่ยวเดินทางมาเลือกซื้อสินค้า และสิ่งหนึ่งที่เป็นชื่อเสียงในตลาดนัดด้านผลไม้ก็คือ ทุเรียน จากข้อมูลสถานการณ์สินค้าเกษตรที่สำคัญและแนวโน้มปี 2560 พบว่าอัตราการบริโภคของทุเรียนในประเทศไทยในช่วงปี 2555 – 2559 มีอัตราการบริโภคจาก 152,524 ตัน ในปี 2555 เป็น 160,444 ตัน ในปี 2559 หรือเพิ่มขึ้นร้อยละ 2.60 ต่อปี ซึ่งการบริโภคส่วนใหญ่จะอยู่ในรูปผลสด (สำนักงานเศรษฐกิจการเกษตร สำนักวิจัยเศรษฐกิจการเกษตร, 2559) ด้วยเหตุนี้ทุเรียนจึงจัดอยู่ในประเภทผลไม้ในฤดูกาลที่นิยมทั้งคนไทยและต่างประเทศ และด้วยชื่อที่เป็นราชาของผลไม้จึงทำให้ผู้บริโภคมีความอยากทดลองรับประทานหรือพิสูจน์รสชาติของทุเรียน ซึ่งยังสามารถรับประทานได้ในหลายรูปแบบจึงทำให้ผู้บริโภคมีพฤติกรรมการรับประทานที่แตกต่างกันตามความชื่นชอบ เช่น การทานทุเรียนทอดกรอบที่ผลิตจากทุเรียนดิบ ทุเรียนงอมนำไปแปรรูปเป็นทุเรียนกวน

หรือมูลข้าวเหนียวทุเรียนในส่วนของทุเรียนสุกนั้นสามารถเริ่มรับประทานได้ตั้งแต่ห่ามกรอบ ห่ามสุก สุกกลาง ไปจนถึงสุกอม (เกือบอม) ขึ้นอยู่กับความชื่นชอบของผู้บริโภคเลือกรับประทานเอง นี่คือเหตุผลสำคัญดังที่กล่าวมาว่าทุเรียนเป็นผลไม้ที่ได้รับความนิยมทั้งในและนอกประเทศ

#### 2.1.4 Ethylene Gas

ส่วนประกอบของก๊าซเอทิลีนนั้น หากพิจารณาทางกระบวนการทางวิทยาศาสตร์ สูตรทางเคมีของเอทิลีน คือ  $C_2H_4$  (Eberhardt, Schmitt, & Wöllenstein, 2017) ซึ่งหมายถึงมีคาร์บอน 2 อะตอม และมี ไฮโดรเจน 4 อะตอม ตามรูปที่ 2.1 เอทิลีนคือก๊าซที่ปนอยู่ในผลไม้ทุกชนิดและถือเป็นปัจจัยหลักของการดำรงชีวิตของผลไม้เพราะเป็นส่วนควบคุมการเจริญเติบโต การหลุดร่วงของใบ การผลิตดอก รวมไปถึงการเจริญเติบโตของผลพืชหรือผลไม้ทุกชนิดจะสามารถผลิตเอทิลีนขึ้นมาใช้เองได้โดยสกัดจากเนื้อเยื่อทุกสรีระของร่างกาย การผลิตได้จำนวนมากหรือน้อยนั้นขึ้นอยู่กับสภาพแวดล้อม เช่น ความชื้นของดิน ปริมาณแสง หรืออุณหภูมิปัจจัยเหล่านี้ย่อมส่งผลต่อปริมาณมากหรือน้อยต่อการผลิตเอทิลีน ของพืชหรือผลไม้ดังกล่าว ดังนั้นเมื่อผลไม้อยู่ในระหว่างอาการเปลี่ยนแปลงตามสภาวะจากอาการดิบไปสู่อาการอม (ชราภาพ) นั้นเอทิลีนย่อมถือว่าเป็นส่วนสำคัญหรือเป็นปัจจัยหลักในการเปลี่ยนแปลงสภาพนี้โดยสามารถวิเคราะห์พิจารณาจากจำนวนความเข้มข้น หรือปริมาณของก๊าซได้ ในผลไม้แต่ละชนิดจะปล่อยก๊าซเอทิลีนออกมาแตกต่างกันขึ้นอยู่กับชนิดและสายพันธุ์ของผลไม้

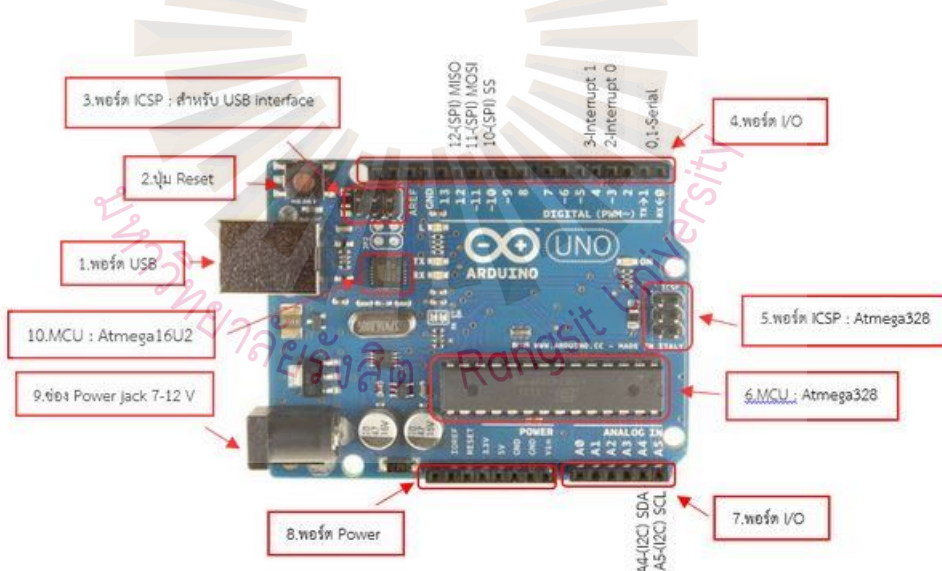


รูปที่ 2.1 แสดงสูตรทางเคมีของเอทิลีน

ที่มา : ผู้วิจัย

### 2.1.5 อาดูโน่ (Arduino)

เนื่องจาก Arduino เป็นภาษาอิตาลี ซึ่งมีการอ่านออกเสียงเป็นรูปแบบเฉพาะและยังไม่มีกำหนดเป็นคำภาษาไทยขึ้นมาอย่างเป็นทางการ ดังนั้นเพื่อไม่ให้เกิดความสับสน ผู้วิจัยจึงขอใช้การทับศัพท์ตามชื่อเรียกที่เขียนเป็นภาษาอังกฤษ Arduino เป็นแพลตฟอร์มอิเล็กทรอนิกส์แบบโอเพนซอร์ส บนพื้นฐานของฮาร์ดแวร์และซอฟต์แวร์ที่ง่ายต่อการใช้งานในลักษณะที่นำชิปไอซีไมโครคอนโทรลเลอร์ตระกูลต่าง ๆ มาใช้งาน (Arduino, 2018 a) หากผู้ประดิษฐ์เขียนไลบรารีของ Arduino ขึ้นมาเพื่อใช้ในการสั่งงานกับไมโครคอนโทรลเลอร์ที่แตกต่างกันแต่สามารถใช้งานโค้ดตัวเดียวกันได้โดยตัว Arduino ได้ออกบอร์ดทดลองมาหลาย ๆ รูปแบบ เพื่อให้มีความเหมาะสม สอดคล้องกับภาษาต่าง ๆ ที่มีโดยเน้นไปที่การใช้งานง่ายและให้นักประดิษฐ์รุ่นใหม่เข้าใช้งานในลักษณะเริ่มต้นได้ตามวัตถุประสงค์ สาเหตุหลักที่ทำให้ Arduino เป็นนิยมนั้นเพราะซอฟต์แวร์ที่ใช้งานร่วมกันได้และง่ายต่อการใช้งานสำหรับผู้เริ่มต้นและยังมีความยืดหยุ่นมากเพียงพอสำหรับผู้ใช้งานในระดับสูงและยังรองรับทั้งระบบ Mac, Windows และ Linux และสามารถโหลดได้ฟรีมีราคาต่ำกว่าเมื่อเทียบกับแพลตฟอร์มไมโครคอนโทรลเลอร์ชนิดอื่น ๆ



รูปที่ 2.2 รูปแสดง Arduino

ที่มา : Arduino, 2018

จากรูปที่ 2.2 เป็นการแสดงบอร์ดของ Arduino ที่ผู้ประดิษฐ์สามารถเลือกนำมาใช้งานให้เหมาะสมกับงานของตนเอง โดยในส่วนใหญ่การตั้งค่าบอร์ดของ Arduino จะมี 2 ลักษณะคือการ



ตั้งค่าพารามิเตอร์ที่ใช้เพื่อการรวบรวมอัปเดต และการตั้งค่าไฟล์ที่ใช้คำสั่งกระบวนการโหลดเข้าระบบหน่วยความจำ (Bootloader) ซึ่งถือเป็นส่วนสำคัญของการโหลดข้อมูลเพื่อนำไปใช้งาน โดยผู้วิจัยจะขออธิบายส่วนต่าง ๆ ของ Arduino ดังนี้

### 2.1.5.1 ส่วนที่เป็นฮาร์ดแวร์ (Hardware)

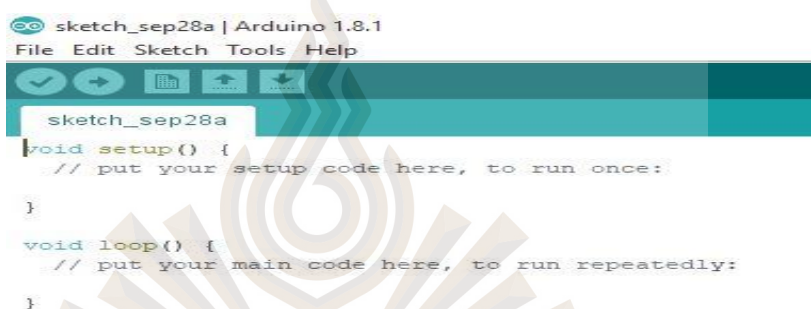
บอร์ดอิเล็กทรอนิกส์ที่มีขนาดเล็กถูกประกอบด้วยชิ้นส่วนหลักที่มีอุปกรณ์ควบคุมที่มีขนาดเล็ก หรือไมโครคอนโทรลเลอร์ (MCU) เป็นตัวควบคุมร่วมกับอุปกรณ์อิเล็กทรอนิกส์อื่น ๆ โดยทั่วไปนักพัฒนาจะเรียกว่าบอร์ด Arduino ซึ่งจะมีลักษณะการใช้งานที่แตกต่างกันไปตามความต้องการของชิ้นงานในเรื่องของขนาดสเปค เช่น ประสิทธิภาพของไมโครคอนโทรลเลอร์ แรงดันไฟ และความเร็วในการรับส่งสัญญาณ ตามรูปที่ 2.2 ดังนี้

- 1) ช่อง USB ใช้สำหรับต่อกับคอมพิวเตอร์เพื่ออัปเดตโปรแกรมเข้าไมโครคอนโทรลเลอร์ และจ่ายไฟให้กับบอร์ด
- 2) ปุ่มรีเซ็ตใช้เพื่อทำการรีเซ็ตระบบไมโครคอนโทรลเลอร์ให้เริ่มการทำงานใหม่
- 3) ICSP Port เป็นช่องที่ใช้ติดต่อกับ Visual Com port
- 4) ช่องรับ-ส่งสัญญาณดิจิทัลตั้งแต่ขา D 0 ถึง D 13 นอกจากนี้ บาง Pin จะทำหน้าที่อื่นๆ เพิ่มเติมด้วย เช่น Pin0,1 เป็นขา Tx,Rx Serial, Pin3,5,6,9,10 และ 11 เป็นขา PWM
- 5) ICSP Port เป็นช่องที่ใช้โหลดระบบเข้าสู่หน่วยความจำ
- 6) MCU เป็นไมโครคอนโทรลเลอร์ที่ใช้บนบอร์ด Arduino
- 7) เป็นช่องที่นอกจากจะเป็นทำหน้าที่ส่งข้อมูลแล้วยังเป็นช่องรับสัญญาณอนาล็อกด้วย โดยเริ่มตั้งแต่ขา A0-A5
- 8) ไฟเลี้ยงของบอร์ดเมื่อต้องการจ่ายไฟให้กับวงจรภายนอก ประกอบด้วยไฟเลี้ยง +3.3 V, +5V, GND, Vin
- 9) Power Jack รับไฟจากเครื่องดัดแปลง โดยที่แรงดันอยู่ระหว่าง 7-12 V
- 10) MCU ที่ทำหน้าที่ติดต่อกับคอมพิวเตอร์

### 2.1.5.2 ส่วนที่เป็นซอฟต์แวร์ (Software)

สิ่งที่สำคัญสำหรับการใช้งานที่ควบคู่กับเครื่องมือที่ช่วยในการพัฒนาโปรแกรม หรือ Integrated Development Environment (IDE) โดยมีสิ่งอำนวยความสะดวกต่าง ๆ เช่น คำสั่ง Compile, Run ตัวอย่างของ IDE เช่น NetBeans, Editplus, JCreator และ Eclipse

เครื่องมือช่วยพัฒนาโปรแกรมนี้ถือว่าเป็นส่วนสำคัญในการใช้พัฒนาเพราะถือว่าเป็นแพลตฟอร์มออนไลน์ที่ช่วยให้นักพัฒนาสามารถเขียนโปรแกรมติดต่อฟังชันต่าง ๆ และสามารถเข้าถึงได้ในแดชบอร์ดเดียว ในส่วนของโมดูล Arduino รุ่นที่มีราคาสูงนั้นสามารถประกอบขึ้นได้ด้วยตนเอง เพราะมีกระบวนการไม่ซับซ้อน ภาษาที่ใช้ในการพัฒนานั้นได้แก่ ภาษา C/C++ แต่สำหรับผู้ที่มีความเชี่ยวชาญสามารถเขียนเป็นภาษาโปรแกรม AVR-C ลงในโปรแกรม Arduino ได้โดยตรง ตามรูปที่ 2.3



```
sketch_sep28a | Arduino 1.8.1
File Edit Sketch Tools Help

sketch_sep28a
void setup() {
  // put your setup code here, to run once:
}

void loop() {
  // put your main code here, to run repeatedly:
}
```

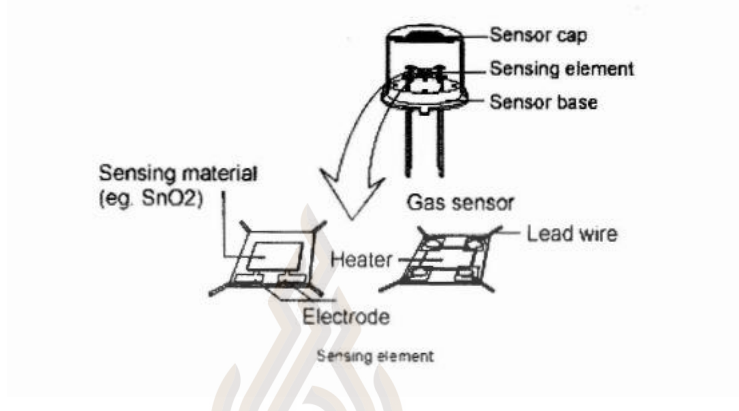
รูปที่ 2.3 รูปแสดงโปรแกรมเริ่มต้นของ Arduino

ที่มา : ผู้วิจัย

### 2.1.6 เซนเซอร์ (Sensor)

เป็นอุปกรณ์ตรวจจับสัญญาณ อุณหภูมิ แสง เสียง แรงทางกล ความดันบรรยากาศ ความเร็ว อัตราเร่ง ระดับของเหลว ระยะขจัด รวมไปถึงอัตราการไหล (อิพิงค์ สุริยา และ มงคล ปุษยตานนท์, 2549) มีลักษณะการทำงานในรูปแบบที่เป็นกระบวนการโดยจะเปลี่ยนแปลงสัญญาณออกที่มีรูปแบบ และนำไปประมวลผลต่อโดยใช้ตัวแปรของระบบทำการวัดค่าต่าง ๆ ที่ได้รับเข้ามาตามสภาพลักษณะที่ต้องการทราบหรือควบคุม ส่วนใหญ่แล้วจะอยู่ในรูปของฟิสิกส์ ดังนั้นจากเหตุผลในความจำเป็นที่ต้องนำข้อมูลที่ได้จากการตรวจจับปริมาณทางฟิสิกส์เหล่านั้นด้วยเซนเซอร์ไปใช้ข้อมูลเพื่อทำการตรวจสอบหรือเพื่อแสดงสถานภาพกระบวนการของระบบ ซึ่งโดยทั่วไปพบว่าเทคโนโลยีของเซนเซอร์ได้ถูกนำไปใช้ประกอบหลักที่สำคัญในลักษณะงาน 2 ประเภท คือ ใช้ตรวจปริมาณทางฟิสิกส์เพื่อนำไปจัดเก็บหรือตรวจวัดและใช้ตรวจสอบ สภาพ

กระบวนการ โดยมีโครงสร้างของเซนเซอร์ตรวจจับปริมาณความเข้มข้นของแก๊ส (Gas Sensor) ตามรูปที่ 2.4 (อิพิงส์ สุริยา และมงคล ปุชยานนท์, 2549)



รูปที่ 2.4 แสดง โครงสร้างของแก๊สเซนเซอร์  
ที่มา : อิพิงส์ สุริยา และ มงคล ปุชยานนท์, 2549

เซนเซอร์แก๊สจะมีหลักการทำงานหลัก ๆ คือจะเปลี่ยนแก๊สที่ถูกดักจับได้ในอากาศ โดยมีโครงสร้างภายในเป็นลักษณะพิเศษในเชิงฟิสิกส์ คือเซนเซอร์จะเปลี่ยนความหนาแน่นของแก๊สเป็นสัญญาณไฟฟ้า โดยที่เราจะสามารถทราบค่าความเข้มข้นของแก๊สที่ต้องการค้นหาได้นั้น จะต้องได้มาจากการทำปฏิกิริยาในเชิงเคมีของตัวเซนเซอร์ ระหว่างสารกึ่งตัวนำกับค่าทางเคมีของแก๊สที่ต้องการค้นหาขดลวดภายในตัวเซนเซอร์จะทำหน้าที่นำพาความร้อนให้สารกึ่งตัวนำที่มีการเคลือบด้วยสารดีบุกหรือการออกไซด์ของดีบุก เมื่อขดลวดโดนความร้อนจะมีการตอบสนองด้วยการถ่ายอิเล็กตรอนไปผสมกับออกซิเจนที่มีอยู่ในอากาศรอบ ๆ วัตถุหรือกลุ่มเป้าหมายที่ต้องการค้นหาการที่โลหะอยู่ในสถานะละลายอนุภาคออกจากโลหะหากพิสูจน์ทางกระบวนการทางวิทยาศาสตร์จะพบว่าการเคลื่อนไหวของโมเลกุลตลอดเวลาจากนั้นจะเกิดการเปลี่ยนแปลงขึ้นเมื่ออุณหภูมิของโมเลกุลลดลงอนุภาคนั้นจะชลอความเร็วลงตามด้วยเช่นกัน จนทำให้อนุภาคนั้นเกิดการดึงดูดกันและแปลสภาพเป็นนิวเคลียส อนุภาคจะจับเรียงตัวกันอย่างไรทิศทางไม่เป็นระเบียบแต่ในบางครั้งอาจมีการจับตัวกันเป็นระเบียบและสม่ำเสมอก็เป็นได้ เมื่อมีการเย็นตัวลงของโลหะจะพบว่าเป็นนิวเคลียสนั้นเพิ่มขนาดไปเป็นผลึก (Crystal) เล็ก ๆ และมีการขยายตัวจนเปลี่ยนเป็น เป็นผลึกใหญ่ในที่นี้จะเรียกว่า เกรน (Grains) การเติบโตจะสิ้นสุดลงเมื่อมีจำนวนปริมาณของโลหะหลอมเหลวสูญเสียไป หรือมีอีกนัยยะหนึ่งว่าเกิดจากปฏิกิริยาของเกรนที่ขยายตัวไปชนกับเกรนตัวหนึ่งจนทำให้เกิดขอบของเกรน (Grain Boundary) เมื่อขดลวดมีการกระทำตอบสนองต่อออกซิเจนจะมีผลทำให้



เกิด Grain Boundary ซึ่งจะทำให้กลุ่มแก๊สที่ต้องการค้นหาผ่านเข้ามาส่งผลต่อการกระทำนั้นให้เกิดเปลี่ยนแปลงค่าความต้านทานในตัวเซนเซอร์จึงทำให้ทราบข้อมูลค่าความเข้มข้นของปริมาณแก๊สที่ต้องการค้นหาที่สามารถนำไปใช้งานได้ตามความต้องการของกลุ่มเซนเซอร์ต่าง ๆ ได้ ในส่วนการทดลองนั้นผู้วิจัยนำ MQ Gas Sensor (Arduino playground, 2018 b) มาใช้เป็นตัวตรวจจับก๊าซได้ ผู้วิจัยได้ศึกษาข้อมูลเพิ่มเติมซึ่งสามารถอธิบายได้ว่ากระบวนการทำงานของ เพราะ MQ Gas Sensor จะมีลักษณะการทำงานที่คล้ายกับเซนเซอร์ทั่วไปคือ การตรวจจับก๊าซที่มีปะปนอยู่ในอากาศโดยรอบของกลุ่มเป้าหมายหรือวัตถุอื่น ๆ ตัวอย่างเช่น MQ-6 จะตรวจจับก๊าซของ LPG และ MQ-8 ตรวจจับไฮโดรเจน ขึ้นอยู่กับความต้องการของข้อมูลก๊าซชนิดใด MQ Gas Sensor มีกระบวนการทำงานคือจะเปลี่ยนค่าแรงดันของกระแสไฟฟ้า (Electromotive Force : EMF) ที่ไหลผ่านขดลวดขนาดเล็กที่มีปฏิกิริยาต่อเคมีที่อยู่ภายในตัวเซนเซอร์เป็นค่าความต้านทาน ซึ่งเรียกรกระทำนี้ว่าการออกไซค์กับโลหะปฏิกิริยาการเพิ่มหรือลดลงนี้จะทำให้ทราบถึงข้อมูลความเข้มข้นของก๊าซในอากาศเป้าหมายโดยปริมาณแรงดันจะขึ้นอยู่กับจำนวนความหนาแน่นของก๊าซนั้น ด้วยเหตุนี้จึงทำให้เราสามารถนำค่าที่ได้มานั้นไปใช้ประมวลผลตามลักษณะการใช้งานตามความต้องการได้

### 2.1.7 ทฤษฎีเบื้องต้นภาษา C

ภาษา C เป็นภาษาคอมพิวเตอร์ที่ใช้ในการเขียนโปรแกรมที่มีโครงสร้าง โดยการกำหนดขอบเขตของตัวแปร การเรียกใช้ตัวเอง (Recursion) ที่มีรูปแบบการเขียนโปรแกรมตามลำดับ โดยนักพัฒนาชื่อว่า Dennis Ritchie ในช่วงปี 1969 -1973 ที่ Bell Labs และใช้สำหรับพัฒนาระบบปฏิบัติการ Unix จากนั้นจึงได้รับความนิยมอย่างแพร่หลาย และถูกบริษัทผู้พัฒนานำไปใช้สำหรับพัฒนาในสถาปัตยกรรมคอมพิวเตอร์และระบบปฏิบัติการจนเป็นที่ยอมรับและนิยมมาก (ดวงแก้ว สวามิภักดิ์, 2531) โดยมี American National Standards Institute (ANSI) เป็นผู้กำหนดมาตรฐานตั้งแต่ปี 1989 และ International Organization for Standardization (ISO) ในช่วงเวลาต่อมา ลักษณะการทำงานของภาษา C คือ คอมพิวเตอร์จะไม่เข้าใจภาษาขั้นสูงที่มนุษย์เขียนขึ้นจึงจำเป็นต้องมีคอมไพเลอร์ (Compiler) ที่ทำหน้าที่แปลภาษาขั้นสูงให้เป็นภาษาคอมพิวเตอร์ จากนั้นคอมไพเลอร์จะทำหน้าที่แปลภาษามาเป็นภาษาเป้าหมาย (Object Program) เพื่อติดต่อสื่อสารกับเครื่องคอมพิวเตอร์ต่อไป และตัวเชื่อม (Linker) ทำหน้าที่ตรวจสอบโปรแกรมว่ามีการเรียกใช้ฟังก์ชันมาตรฐานใดจาก Library บ้าง ขั้นตอนนี้เรียกว่าการรวมฟังก์ชันสำเร็จรูปแล้วสร้างไฟล์ที่ทำงานได้ ในการเขียนโปรแกรมโดยใช้ภาษา C นั้น ผู้พัฒนาจะต้องใช้เครื่องมือเพื่อช่วยในการพัฒนาโปรแกรมนั้นคือ IDE (Integrated Development Environment) ในการอำนวยความสะดวกในคำสั่ง Compile, Run เช่น

NetBeans, Editplus, JCreator, Eclipse ซึ่งเครื่องมือช่วยในการพัฒนานี้จะรวมคำสั่ง เมนู และ GUI เพื่อใช้ในการเขียนโปรแกรมหรือเป็นการสร้างสภาพแวดล้อมให้เอื้ออำนวยต่อการเขียนโปรแกรมนั้นเอง โดย IDE นั้นจะทำหน้าที่เปิดไฟล์โปรแกรมและเก็บข้อมูลการคอมไพล์ไว้ในโปรเจกเป็นต้น ตัวอย่างแสดงภาษา C ตามรูปที่ 2.5

```
// variables will change:
int buttonState = 0;           // variable for reading the pushbutton status

void setup() {
  // initialize the LED pin as an output:
  pinMode(ledPin, OUTPUT);
  // initialize the pushbutton pin as an input:
  pinMode(buttonPin, INPUT);
}

void loop() {
  // read the state of the pushbutton value:
  buttonState = digitalRead(buttonPin);

  // check if the pushbutton is pressed.
  // if it is, the buttonState is HIGH:
  if (buttonState == HIGH) {
    // turn LED on:
    digitalWrite(ledPin, HIGH);
  } else {
    // turn LED off:
    digitalWrite(ledPin, LOW);
  }
}
```

รูปที่ 2.5 แสดงตัวอย่างภาษา C

ที่มา : ผู้วิจัย

### 2.1.8 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลคือการวิเคราะห์ข้อมูลเพื่อหารูปแบบ (Patterns) หรือกฎ (Rules) ที่เกิดขึ้นในฐานข้อมูลขนาดใหญ่ โดยการหาความสัมพันธ์กันระหว่างข้อมูล (Relation) โดยการดึงข้อมูลและข่าวสารที่สำคัญน่าสนใจที่ซ่อนอยู่ในข้อมูลหรือตามที่คุณค้นหาต้องการนำไปใช้ประโยชน์จากจำนวนข้อมูลหรือขนาดใหญ่ ด้วยวิธีการทางสถิติ การเรียนรู้ของเครื่อง การรู้จำแบบ (Pacharawongsakda, 2014) โดยในการทำเหมืองข้อมูลนั้น สายชล สตินสมบูรณ์ (2558) ได้กล่าวไว้ว่าเป็นกระบวนการทำงาน (Process) ที่มีลักษณะสกัดข้อมูล (Extract Data) จากฐานข้อมูลที่มีขนาดใหญ่ (Large Information) เพื่อให้ได้มาซึ่งแหล่งข้อมูลของสารสนเทศที่มีประโยชน์ (Useful Information) ที่เราอาจจะยังหรือไม่สามารถทราบข้อมูลที่ถูกซ่อนอยู่ (Unknown Data) โดยเป็นสารสนเทศที่มี

เหตุผล (Valid Information) และสามารถนำไปใช้ได้ (Actionable) ซึ่งเป็นสิ่งสำคัญที่ช่วยในการตัดสินใจ

### 2.1.8.1 ประเภทข้อมูลที่สามารถทำเหมืองข้อมูล

ประเภทข้อมูลที่สามารถนำมาใช้ในการทำเหมืองข้อมูลนั้นได้แก่ฐานข้อมูลประเภทตาราง ฐานข้อมูลคลังสินค้า ฐานข้อมูลรายการ และฐานข้อมูลขั้นสูง อย่างไรก็ตาม เอกสิทธิ์ คุณานันทกุล (2557) ได้อธิบายไว้ว่าในข้อมูลเหล่านี้เมื่อนำมาใช้ในการวิเคราะห์จริงแล้วจะเป็นข้อมูล 2 ลักษณะ คือ ข้อมูลแบบมีโครงสร้าง (Structured Data) และข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) โดยสามารถอธิบายได้ดังนี้

#### 1) ข้อมูลแบบมีโครงสร้าง (Structured Data) มีลักษณะดังนี้

1.1) ฐานข้อมูลตาราง (Relational Database) เป็นข้อมูลที่ถูกจัดเก็บข้อมูลในรูปแบบของตารางและมีลักษณะที่มีความสัมพันธ์แบบมีอยู่จริง (Entity-Relationship Model) ฐานข้อมูลประเภทตารางนี้จะในรูปแบบ Excel ในการประมวลผลด้วย Data mining จะเรียกข้อมูลในส่วนของแต่ละตัวอย่าง (Example) ในส่วนของคอลัมน์เรียก Attribute หรือ Feature ข้อมูลตารางนี้ช่วยลดความซ้ำซ้อนของข้อมูลได้และสามารถจัดเก็บเพิ่มแก้ไข และลบข้อมูลได้

1.2) ฐานข้อมูลคลังสินค้า (Data Warehouses) คลังข้อมูลรวบรวมข้อมูลจากหลาย ๆ ฐานข้อมูล แปลงข้อมูลให้มีความเหมือนกันเหมาะสำหรับการเรียกดู (View) เพื่อสร้างรายงานสรุป

1.3) ฐานข้อมูลรายการ (Transactional Database) เป็นฐานข้อมูลที่มีลักษณะของการเปลี่ยนแปลงของข้อมูลจากสถานะหนึ่ง ไปยังอีกสถานะหนึ่ง โดยภายในหนึ่งข้อมูลที่เกี่ยวข้องกับเหตุการณ์ที่เกิดขึ้น และจะต้องจดบันทึกไว้เพื่อนำไปปรับปรุงเพิ่มข้อมูลหลัก (Transactional) จะประกอบไปด้วยคำสั่งแบบประมวลผลเสร็จสมบูรณ์ (Commit) และคำสั่งแบบการประมวลผลในการดำเนินการถ้าไม่สมบูรณ์ นั้นจะถูกยกเลิกและคืนค่าข้อมูลย้อนกลับไปเป็นค่ารูปแบบเดิม (Rollback) ตัวอย่างเช่น การถอนเงินแต่ยอดเงินไม่เพียงพอในการถอนนั้น คำสั่งจะเกิดการคืนค่าข้อมูลย้อนกลับและยกเลิกการดำเนินการนั้นไป

2) ข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) เป็นฐานข้อมูลขั้นสูง (Advanced Database) เป็นฐานข้อมูลที่ถูกจัดเก็บในรูปแบบอื่นเช่น ข้อมูลจากสังคมออนไลน์ในลักษณะรูปแบบชนิดไฟล์ตัวอักษร (Text file) ในรูปแบบข้อมูลข่าวสารทั้งรูปและเสียง หรือจากเว็บไซต์

### 2.1.8.2 เทคนิคการวิเคราะห์ข้อมูลด้วย Data Mining

เทคนิคการวิเคราะห์ข้อมูลด้วย Data Mining สามารถวิเคราะห์ได้ 2 ประเภทหลัก (Pacharawongsakda, 2014) ดังนี้

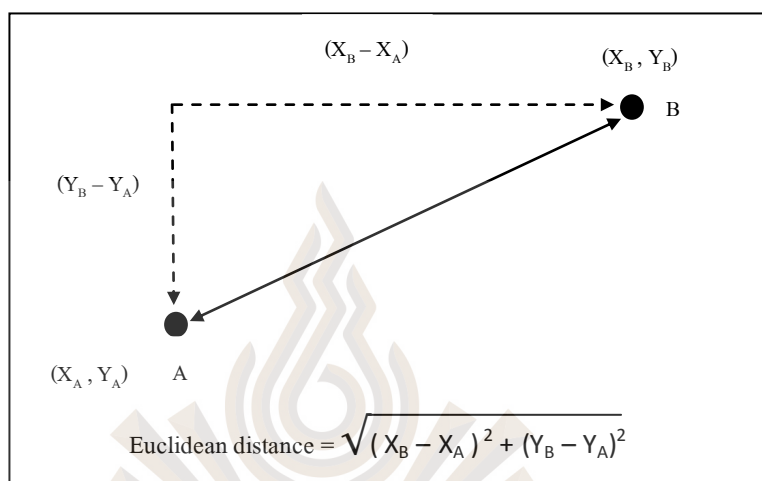
1) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการพิจารณาข้อมูลว่ามีความสัมพันธ์กันในลักษณะใด เทคนิคนี้จะค้นหากฎความสัมพันธ์ (Association Rule) และการแบ่งกลุ่ม (Clustering)

2) การเรียนรู้แบบมีผู้สอน (Supervised Learning) เทคนิคนี้จะนำข้อมูลที่มีในอดีตมาวิเคราะห์โดยการสร้างโมเดล (Model) เพื่อทำนายสิ่งที่จะเกิดขึ้น โดยใช้หลักการทางสถิติ คณิตศาสตร์ รวมไปถึงกฎต่าง ๆ เทคนิคนี้จะค้นหาข้อมูลโดยการจำแนกข้อมูล (Classification) และการประมาณค่าข้อมูล (Regression) ซึ่งลักษณะทั้งสองนี้มีขั้นตอนกระบวนการทำงานที่มีลักษณะคล้ายคลึงกันแต่จะแตกต่างกันตรงที่การค้นหาคำตอบในกลุ่มข้อมูลที่ซ่อนอยู่ การจำแนกข้อมูลจะมีการทำนายค่าที่เป็นตัวอักษร ส่วนการทำนายข้อมูลแบบการวิเคราะห์การถดถอยจะเป็นลักษณะค่าที่เป็นตัวเลข เป็นต้น

### 2.1.8.3 การแบ่งกลุ่มของข้อมูล (Clustering)

ในความเป็นจริงแล้วนั้นข้อมูลที่มีอยู่ในปัจจุบันทั้งในรูปแบบของโซเชียล หรือในรูปแบบของฐานข้อมูล จะมีลักษณะที่แตกต่างกันไปตามกลุ่มของประเภท ชนิด รวมไปถึงลักษณะของข้อมูล และในแต่ละกลุ่มข้อมูลนั้นจะมีความแตกต่างจากวัตถุในกลุ่มอื่น ๆ ด้วยเหตุนี้จึงจำเป็นต้องมีการแบ่งกลุ่มของข้อมูล (Clustering) ด้วยเทคนิคการแบ่งกลุ่มนี้มีลักษณะการทำงานด้วยการแบ่งกลุ่มข้อมูลที่มีลักษณะคล้ายกันให้อยู่รวมกัน โดยจัดข้อมูลให้อยู่ในกลุ่มด้วยวิธีการวัดความคล้ายคลึง (Similarity) หรือการวัดระยะห่าง (Distance) ระหว่างข้อมูลแต่ละตัวเพื่อให้ทราบถึงขนาดและจำนวนข้อมูลที่มีอยู่เพื่อใช้ในการจำแนกกลุ่มต่อไปค่าการคำนวณวัดระยะห่างมีหลากหลายชนิด เช่น Euclidean, Person Correlation, Superman Rank Correlation แต่ที่นิยมใช้กันคือ ระยะห่างยูคลิดีเนียน (Euclidean Distance) ตามรูปที่ 2.6 (Pacharawongsakda, 2014) ในการทำการแบ่งกลุ่มส่วนใหญ่จะมีเทคนิคที่นิยมคือ K-Mean Agglomerative clustering และ DBSCAN – Density Based Spatial Clustering of Application with Noise ซึ่งในแต่ละเทคนิคจะมีลักษณะการทำงานที่แตกต่างกันไป ตัวอย่างเช่น K-Mean เป็นเทคนิคประเภท Partitional Clustering คือต้องกำหนดจำนวนคลัสเตอร์ที่ต้องการจะแบ่งในตัวแปรของ K มีลักษณะการทำงานคือ กำหนดจุดศูนย์กลางของแต่ละคลัสเตอร์ และหาระยะห่างระหว่างข้อมูลแต่ละตัวกับจุดศูนย์กลาง (Mean) ของแต่ละคลัสเตอร์ เมื่อได้จุดศูนย์กลางแล้วนั้นจะทำการกำหนดให้ข้อมูลที่อยู่ในแต่ละคลัสเตอร์ที่อยู่ใกล้ที่สุด เพื่อคำนวณหาจุด

ศูนย์กลางของแต่ละคลัสเตอร์ใหม่อีกครั้ง ซึ่งจะซ้ำในลักษณะนี้จนข้อมูลที่อยู่ในแต่ละคลัสเตอร์ไม่มีการเปลี่ยนแปลง เป็นต้น



รูปที่ 2.6 รูปแสดงระยะห่างยูคลิเดียน

ที่มา : Pacharawongsakda, 2014

#### 2.1.8.4 การจำแนกประเภทข้อมูล (Classification)

การจำแนกประเภทข้อมูล คือการนำข้อมูลในอดีตมาเป็นพื้นฐานหรือตัวอย่างในการเรียนรู้หรือรู้จำด้วยการสร้างแบบจำลอง โมเดล (Model) ขึ้นมาเพื่อที่จะให้ข้อมูลชุดใหม่ทำการค้นหาคำตอบ ตัวอย่างเช่นการค้นหาการส่งข้อความที่ผู้รับไม่ได้ร้องขอผ่านทางระบบอิเล็กทรอนิกส์ โดยส่วนมากจะทำให้เกิดความไม่พอใจต่อผู้รับข้อความ (Spam) จะต้องมีตัวอย่างจดหมายอิเล็กทรอนิกส์หรือไปรษณีย์อิเล็กทรอนิกส์ที่เป็นหรือไม่เป็นข้อความที่ทำให้เกิดความไม่พอใจก่อนเพื่อให้ระบบเรียนรู้จากนั้นจึงสร้าง โมเดลขึ้นมาโดยกำหนดค่าไว้ที่ตารางจัดการหรือข้อความที่เป็นหัวข้อ (Label) เพื่อให้ระบบทำการจำแนกประเภทของข้อมูลว่ามีลักษณะเป็นเช่นไรในการประยุกต์ใช้งานสามารถแบ่งเป็น 3 ขั้นตอน ดังนี้

1) ขั้นตอนการสร้างโมเดล เป็นขั้นตอนแรกในการจัดเตรียมของข้อมูล และสร้างข้อมูลให้เหมาะสมในการนำไปใช้งานได้จากนั้นนำข้อมูลดิบไปทำการทดสอบ (Training Data) ในการทำเทรนนิ่งนี้จะได้คำตอบมา 2 ส่วนคือแอตทริบิวต์ที่เป็นคำตอบน่าสนใจ และแอตทริบิวต์ที่เป็นแบบทั่วไป เช่นในกลุ่มข้อมูลที่เป็นอีเมลจะได้คำตอบที่เป็นอีเมลปกติและอีเมลแปลกปลอม



หรือไม่ตรงตามวัตถุประสงค์สิ่งสำคัญคือข้อมูลแต่ละแอตทริบิวต์ (Attributes) จะต้องมี ความสัมพันธ์กับแต่ละคำตอบจึงจะทำให้โมเดลที่เราสร้างมีความน่าเชื่อถือและ ประสิทธิภาพมาก

2) ขั้นตอนการทดสอบประสิทธิภาพของ โมเดลในการวัดประสิทธิภาพ นี้สามารถใช้เทคนิคต่างๆ มาทำการเปรียบเทียบกันในแต่ละ โมเดล โดยหาค่าความถูกต้อง (Accuracy) โดยพิจารณาทุกคลาสรวมกัน เพื่อให้เกิดความสัมพันธ์ของข้อมูลและคำตอบที่ถูกต้อง จนเป็นที่น่าพอใจในการทดสอบนี้อาจกระทำ ได้ 2 ลักษณะคือ การวัดประสิทธิภาพของ โมเดลจาก การจำแนกข้อมูล และการแบ่งข้อมูลเพื่อใช้ในการวัด ประสิทธิภาพของ โมเดลจากการจำแนก ประเภทข้อมูล

3) ขั้นตอนการนำ โมเดลไปใช้งานเพื่อทำนายข้อมูลใหม่หลังจากวัด ประสิทธิภาพและได้คำตอบที่มีความถูกต้องน่าเชื่อถือแล้ว เราสามารถนำข้อมูลที่ได้นี้ไปทำการ ทำนายข้อมูลที่เข้ามาใหม่เพื่อให้ได้คำตอบที่แม่นยำมากขึ้นตัวอย่างเช่น การพยากรณ์สภาพอากาศ การแข่งขันฟุตบอลที่มีคำตอบเป็น มีแสงแดดแต่มีความชื้นแบบธรรมดา แต่จะยังไม่สามารถสรุปได้ เพียงใด แต่หากมีการทำนายใหม่อีกครั้งจะได้คำตอบที่เป็น มีแสงแดดจัดความชื้นแบบธรรมดา และสามารถแข่งขันได้ เมื่อทราบถึงขั้นตอนการจำแนกประเภทข้อมูลว่ามีลักษณะแบบใดนั้นขั้นตอนต่อไป คือการนำข้อมูลมาวัดประสิทธิภาพดังนี้

### 3.1) การวัดประสิทธิภาพของ โมเดลจากการจำแนกข้อมูล

โดยทั่วไปแล้วจะเป็นที่นิยมในการใช้วัดประสิทธิภาพของ โมเดลจากการจำแนกข้อมูลอยู่ 5 ค่า คือ

Precision เป็นการคำนวณวัดความแม่นยำของ โมเดล โดยจะมีการพิจารณาข้อมูลที่ใช้ในการทดสอบทีละชุดข้อมูล

Recall เป็นการคำนวณวัดแม่นยำต้องของ โมเดลอีกครั้ง โดยจะมีการพิจารณาที่ใช้ในการทดสอบทีละชุดข้อมูล

F-measure เป็นการวัด Precision และ Recall พร้อมกันของ โมเดล โดยพิจารณาข้อมูลที่ใช้ในการตรวจสอบทีละชุดข้อมูล

Accuracy เป็นการวัดความถูกต้องของ โมเดล โดยพิจารณารวม ข้อมูลที่ใช้ในการตรวจสอบทุกชุดข้อมูล

### 3.2) การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของ โมเดลการจำแนกประเภทข้อมูล

ขั้นตอนนี้เป็นกรวัดประสิทธิภาพที่แบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนที่ใช้ในการสร้างโมเดล และส่วนที่ใช้ในการทดสอบประสิทธิภาพของโมเดล ในการแบ่งข้อมูลนี้พออธิบายได้ 3 ขั้นตอน คือ

Self Consistency Test เป็นวิธีที่ง่ายสะดวกและรวดเร็วที่สุดคือการใช้ข้อมูลที่ใช้ในการสร้างโมเดลและข้อมูลที่ใช้ในการทดสอบเป็นข้อมูลชุดเดียวกัน การวัดประสิทธิภาพลักษณะนี้อาจจะให้ค่าความถูกต้องสูงมาก เนื่องจากเป็นข้อมูลชุดเดียวกันที่ระบบได้ทำการเรียนรู้มาแล้ว การวัดประสิทธิภาพลักษณะนี้เหมาะกับการดูแนวโน้มเบื้องต้นของการสร้างโมเดล หากได้ค่าวัดน้อยอาจบอกได้ว่า ข้อมูลนี้ไม่เหมาะสมกับโมเดล

Split Test เป็นการแบ่งข้อมูลด้วยวิธีการสุ่มเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยจะนำข้อมูลในส่วนมาก (70% และ 80%) ไปเป็นส่วนของโมเดล และในส่วนน้อย (30% และ 20%) เป็นในส่วนของการทดสอบประสิทธิภาพของโมเดล การทดสอบลักษณะนี้เป็นการสุ่มข้อมูลเพียงครั้งเดียวซึ่งอาจทำให้ข้อมูลที่ใช้สร้างโมเดลและวัดประสิทธิภาพอาจเป็นข้อมูลที่มีลักษณะหรือชุดเดียวกัน จึงอาจทำให้ผลการวัดมีค่าสูงหรือค่าน้อย จึงควรทำการสุ่มข้อมูลหลาย ๆ ครั้ง ซึ่งการวัดประสิทธิภาพลักษณะนี้มีผลดีคือสร้างง่ายและรวดเร็วเหมาะสมกับข้อมูลใหญ่

Cross Validation Test เป็นวิธีที่นิยมในการวัดประสิทธิภาพของโมเดลเป็นอย่างมากเนื่องจากมีความน่าเชื่อถือ วิธีการทำคือการแบ่งข้อมูลออกเป็นหลาย ๆ ส่วน โดยแต่ละส่วนจะมีชุดข้อมูลที่เท่าๆ กัน อาจแบ่งเป็น 5 หรือ 10 Fold หลังจากนั้นจะเก็บข้อมูล 1 ส่วนเพื่อใช้ทดสอบประสิทธิภาพของโมเดล ทำวนจนครบตามจำนวนที่แบ่งไว้ เช่น แบ่งข้อมูลเป็น 5 - Fold Cross – Validation จะทำการทดสอบ 5 รอบ โดยชุดแรกจะใช้ข้อมูล 2 – 5 สร้างโมเดลและใช้ข้อมูล 1 ทดสอบ ชุดสองใช้ 1 3 4 และ 5 สร้างโมเดล และใช้ข้อมูล 2 ทดสอบ ทำวนจนครบทุกชุดจำนวนข้อมูลทดสอบ

#### 2.1.8.5 อัลกอริทึมที่ใช้ในการจำแนกประเภทของข้อมูล

การจำแนกประเภทของข้อมูลเป็นเทคนิคต่างๆ ของการทำเหมืองข้อมูล โดยการใช้ข้อมูลในอดีตมาเป็นแบบจำลองในการทำนายข้อมูลใหม่ในอนาคตด้วยการเรียนรู้จำ

ของเครื่อง (Machine Learning) สำหรับการวิเคราะห์ข้อมูลที่มีขนาดใหญ่จึงจำเป็นต้องมีอัลกอริทึมเป็นเครื่องมือในการช่วยวิเคราะห์โดยมีรายละเอียดดังนี้

#### 1) เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้การตัดสินใจถูกพัฒนาโดย (Quinlan, 1993) เป็นเทคนิคยอดนิยมในการทำการจำแนกประเภทกลุ่มข้อมูลเพราะต้นไม้การตัดสินใจนั้นมีการเรียนรู้เหตุการณ์ที่จะเกิดขึ้นโดยใช้เงื่อนไขเป็นตัวช่วยในการตัดสินใจและมีอัลกอริทึมในการเลือกแอตทริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดมาใช้งาน โดยมุ่งเน้นที่การคำนวณค่าปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบ โดยผู้วิจัยจะขอทับศัพท์เป็นภาษาอังกฤษว่า Entropy และค่าปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบในสถานะปัจจุบันกับปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบของสถานะก่อนหน้า (Information Gain) โดยผู้วิจัยจะขอเรียกว่า IG เป็นหลักและจะแบ่งข้อมูลเป็นสองส่วนเพื่อหาค่ากึ่งกลางระหว่างค่าที่แบ่งได้ จากนั้นระบบจะนำค่า IG ที่สูงสุ่มมาใช้งาน จากนั้นก็จะหาแอตทริบิวต์ถัดไปเรื่อย ๆ โดยมีสูตรการคำนวณตามสมการที่ 2-1

$$\text{Entropy}(c_1) = -p(c_1) \log p(c_1)$$

(2-1)

$$\text{Information Gain}(\text{parent}, \text{child}) = \text{Entropy}(\text{parent}) - [P(c_1) \times \text{Entropy}(c_1) + P(c_2) \times \text{Entropy}(c_2) + \dots]$$

จากรูปปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบการคำนวณสมการ Entropy ( $c_1$ ) =  $-p(c_1) \log p(c_1)$  และ  $p(c_1)$  คือค่าความน่าจะเป็นของ  $c_1$  ซึ่งค่า Entropy นี้จะเป็นตัววัดความแตกต่างข้อมูล เช่น ข้อมูลมีความแตกต่างกันอย่างมาก ค่า Entropy จะมีค่าสูง แต่ในทางกลับกันหากข้อมูลมีความแตกต่างกันน้อย ค่า Entropy จะมีค่าต่ำ ดังนั้นหากข้อมูล Entropy ของโหนดลูก (Child) แบ่งข้อมูลได้ดี จะมีค่าต่ำแต่ค่าผลต่างระหว่างปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบในสถานะปัจจุบันกับปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบของสถานะก่อนหน้า (Information Gain) จะสูงเมื่อเปรียบเทียบกับ โหนดบน (Parent) ในขั้นตอนการสร้าง โมเดลของต้นไม้ตัดสินใจนั้นจะมุ่งเน้นความสำคัญไปที่การคำนวณค่า IG ของแต่ละแอตทริบิวต์เป็นหลักเพื่อนำมาเปรียบเทียบกับคลาสเพื่อค้นหาแอตทริบิวต์ที่มีค่า IG มากที่สุดเป็น Root ของต้นไม้ตัดสินใจ (เอกสิทธิ์ พัทรวงศ์ศักดิ์, 2557)



## 2) เทคนิคการจำแนกข้อมูลด้วยวิธีการค้นหาเพื่อนบ้านที่ใกล้ที่สุด

( K Nearest Neighbors )

เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุด เป็นการจัดแบ่งคลาสด้วยวิธีการวัดระยะห่างของแต่ละแอตทริบิวต์ (Attribute) ในชุดของข้อมูลทีนำมาใช้ในการทดสอบ และจะนำไปคำนวณค่าออกมา เทคนิคการค้นหาเพื่อนบ้านที่ใกล้ที่สุดนั้นจะมีข้อพิเศษคือจะนำข้อมูลที่มีลักษณะเหมือนกันมาเก็บไว้ในชุด Class เดียวกันด้วยการวัดระยะห่างของระหว่างข้อมูลที่ต้องการค้นหา หรือต้องการทำนายกับข้อมูลที่อยู่ใกล้เคียงกัน ด้วยจำนวน K และคำตอบที่ทำนายได้คือ Class ที่ถูกค้นพบมากที่สุดจากข้อมูลที่อยู่ใกล้เคียงกันหรือเป็นเพื่อนบ้านทั้ง K ตัว โดยจะวัดระยะห่างระหว่างข้อมูลแบบ Euclidean ซึ่งเกิดจากรากที่สองของผลต่างระหว่างแอตทริบิวต์ยกกำลังสอง ตามสมการที่ 2-2

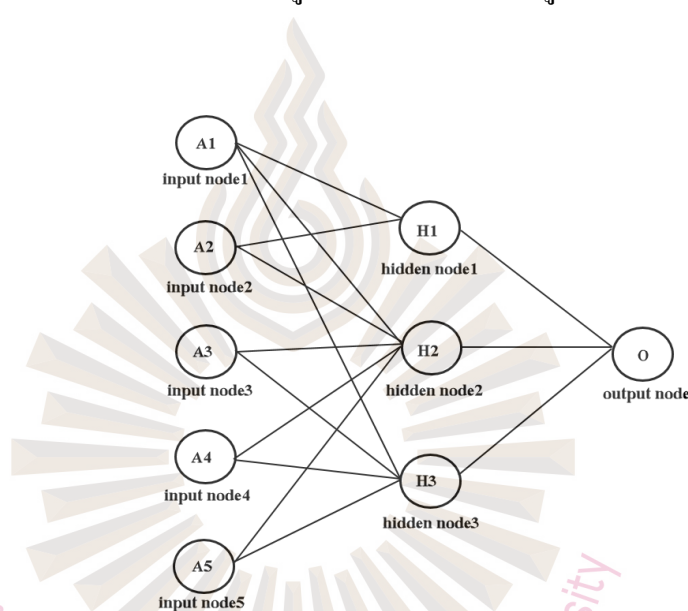
$$D_{\text{Euclidean}} = \sqrt{(X_1 - y_1)^2 + (X_2 - y_2)^2 + \dots + (X_L - y_L)^2} \quad (2-2)$$

จากรูป  $X_1$  คือแอตทริบิวต์ที่ 1 ของข้อมูลจุดที่ 1 และ  $y_1$  คือแอตทริบิวต์ที่ 1 ของข้อมูลจุดที่ 2 โดยข้อมูลทั้งสองตัวมีจำนวนแอตทริบิวต์เท่ากับ L นอกจากนี้การทำงานของ KNN ไม่มีการสร้างโมเดลแต่จะใช้ค่าตัวทั้งหมดมาทำการเทรนนิ่งเพื่อการทำนายชุดข้อมูล เรียกวิธีการนี้ว่า Lazy Mode (Aha, 1997)

## 3) เทคนิคการจำแนกข้อมูลด้วยวิธีเครือข่ายประสาทเทียม (Neural Network)

เครือข่ายประสาทเทียมเป็นโมเดลทางคณิตศาสตร์ที่เลียนแบบการทำงานของสมองมนุษย์ เปรียบเสมือนว่ามีเส้นประสาทเชื่อมต่อกันหลายจุดเป็นจุดเชื่อมต่อและโยงกันจนเกิดเป็นระบบประสาทเทียมขึ้นมาการทำงานก็จะนำค่าข้อมูลขาเข้า ไปกำหนดน้ำหนัก (Weight) ระหว่างโหนดและปรับเปลี่ยนค่าน้ำหนักด้วยการนำค่าน้ำหนัก ไปคูณกับข้อมูลขาเข้าของแต่ละขาผลที่ได้จากทุกขาข้อมูลนั้นจะนำมาเปรียบเทียบกับเกณฑ์การแบ่งค่า (Threshold) ที่กำหนดไว้ หากมีมากกว่าก็จะส่งผลเป็นข้อมูลขาออกไปและนำข้อมูลที่ไปยังเส้นประสาทอื่น และเปรียบเทียบจำแนกเข้าตามกลุ่ม แต่ในทางกลับกันถ้ามีค่าน้อยกว่าเกณฑ์การแบ่งค่าก็จะไม่ส่งผลได้ออกมา สิ่งที่สำคัญคือ ต้องทราบค่าปริมาณน้ำหนักของชุดข้อมูลและเกณฑ์การแบ่งค่าเพื่อต้องการเพื่อให้คอมพิวเตอร์รู้จักซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้ โดยการสอนให้คอมพิวเตอร์รู้จักการเรียนรู้แบบการจดจำรูปแบบของสิ่งที่ต้องการให้รู้จักนี้เรียกว่า อัลกอริทึม

การแพร่ย้อนกลับ (Back Propagation) ซึ่ง เอกสิทธิ์ พัทธวงษ์ศักดิ์ (2557) ได้อธิบายไว้ว่าเป็นกระบวนการย้อนกลับของการรู้จำในการฝึก Feed-forward Neural Networks จะมีการใช้อัลกอริทึมแบบการแพร่ย้อนกลับเพื่อใช้ในการปรับปรุงน้ำหนักคะแนนของเครือข่าย (Network Weight) หลังจากใส่รูปแบบข้อมูลสำหรับฝึกให้แก่เครือข่ายในแต่ละครั้งแล้วค่าที่ได้รับจากเครือข่ายจะถูกนำไปเปรียบเทียบกับผลที่คาดหวังแล้วทำการคำนวณหาค่าความผิดพลาด ซึ่งค่าความผิดพลาดนี้จะถูกส่งกลับเข้าสู่เครือข่ายเพื่อใช้แก้ไขค่าน้ำหนักคะแนนต่อไป อัลกอริทึมนี้เหมาะสมสำหรับการคาดการณ์อนาคต หรือแยกกลุ่มข้อมูลที่เป็นชนิด มีข้อดีคือสามารถทำนายข้อมูลที่เป็นคลาสและข้อมูลที่เป็นตัวเลขได้ดี



รูปที่ 2.7 รูปแสดง โครงสร้างของเครือข่ายประสาทเทียม

ที่มา : เอกสิทธิ์ พัทธวงษ์ศักดิ์, 2557

จากรูปที่ 2.9 แสดง โครงสร้างของเครือข่ายประสาทเทียมได้ว่ามีลักษณะการทำงานประกอบด้วยโหนดและเส้นใยเชื่อมโยงแต่ละโหนดต่อกันเป็นจุด ๆ เมื่อพิจารณาตามรูปที่สามารถแบ่งออกเป็น 3 ชั้นดังนี้

ชั้นที่ 1 จะเรียกว่าชั้นของข้อมูลชุดขาเข้าประกอบด้วยโหนดที่มีแอดทริบิวต์แสดงอยู่ในข้อมูลทดสอบที่ใช้ในการเรียนรู้ระบบ จำนวนโหนดจะเท่ากับแอดทริบิวต์ ในชั้นนี้จะกำหนดค่าน้ำหนักที่เชื่อมต่อระหว่างโหนดชั้นข้อมูลขาเข้ากับโหนดชั้นในแต่ละชั้นของชั้นซ่อนตัว

(Hidden Layer) แล้วนำค่าน้ำหนักของข้อมูล ไปคูณกับค่าของข้อมูลที่เชื่อมโยงเข้ามาในส่วนของการคำนวณและค่าเป็นข้อมูลขาส่งออกที่ผ่านฟังก์ชัน Activate ด้วยสัญญาณขาเข้าที่มาจากโหนดในชั้นก่อนหน้า และจะส่งผ่านไปข้างหน้าจากโหนดหนึ่ง ไปสู่อีกโหนดหนึ่ง (Sigmoid Function)

ชั้นที่ 2 จะเรียกว่าชั้นของชั้นซ่อนตัว ประกอบด้วยโหนดต่าง ๆ ที่มาจากการแปลงของช่วงข้อมูลด้วยการไหลผ่านของข้อมูลเพื่อลดความซ้ำซ้อน (Normalization) จำนวนของโหนดในชั้นนี้จะเท่ากับแอดทริบิวต์หารสอง ในช่วงโหนดชั้นซ่อนตัวนี้จะพยายามแยกความแตกต่างข้อมูลโดยใช้เส้นตรงเส้นเดียว (Linearly Separable) ส่งค่าข้อมูลขาออกที่ได้จากการแปลงข้อมูลจากชั้นที่ 1 แล้วจะส่งไปยัง Activate Function เพื่อทำการคำนวณแต่ละโหนดในโมเดลและส่งค่าข้อมูลขาออกไปชั้นที่ 3

ชั้นที่ 3 จะเรียกว่าชั้นของข้อมูลขาออก (Output Layer) ประกอบด้วยโหนดที่เป็นคลาส หลังจากได้ค่าข้อมูลขาออกที่ได้รับจากชั้นซ่อนตัวก็จะคำนวณค่าของแอดทริบิวต์ ในชั้นนี้ด้วย Active Function ของโหนดข้อมูลขาออกจะเป็นฟังก์ชันแบบเชิงเส้นตรงหากมีความคลาดเคลื่อนจากการทดสอบเครือข่ายประสาทเทียมนี้จะมีการแก้ไขด้วยวิธีการปรับเปลี่ยนค่าน้ำหนักของเส้นระหว่างโหนดอีกครั้งจนได้ค่าผิดพลาดน้อยที่สุดเรียกเทคนิคนี้ว่าการทำย้อนหลังจากโหนดข้อมูลขาออกย้อนกลับไปที่โหนดชั้นซ่อนตัวเพื่อหาค่าผิดพลาดด้วยเทคนิค Back Propagation (เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา, 2557) ตามสมการที่ 2-3

$$\text{Error}_i = \text{Output}_i (1 - \text{Output}_i) (\text{Actual}_i - \text{Output}_i) \quad (2-3)$$

จากรูปที่ 2.10  $\text{Error}_i$  คือความคลาดเคลื่อนของโหนดที่  $i$  ในชั้นข้อมูลขาออกค่า  $\text{Actual}_i$  คือค่าของคลาสที่อยู่ในการทดสอบข้อมูลและ  $\text{Output}_i$  คือค่าที่คำนวณได้จากโมเดลเครือข่ายประสาทเทียม

#### 4) เทคนิคการจำแนกข้อมูลด้วยวิธี Naive Bayes

Naive Bayes เป็นโมเดลการจัดการข้อมูลเป็นกลุ่มด้วยการใช้หลักความน่าจะเป็น (Probability) ซึ่งเป็นพื้นฐานและสมมุติฐานของ Bayes' Theorem ตัวอย่างเช่นความน่าจะเป็นที่ B เกิดก่อนและ A เกิดตามมาความน่าจะเป็นที่ A เกิดก่อน B ความน่าจะเป็นที่ A และ B เกิดร่วมกัน และความน่าจะเป็นที่ไม่มีทั้ง A และ B เกิด ดังสมการที่ 2-4 (เอกสิทธิ์ พัทธวงษ์ศักดิ์ดา, 2557)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2-4)$$

ตัวอย่างเช่นจะทำนายว่า ถ้ามีแดดออก จะมีเหตุการณ์คือฝนจะไม่ตก และกำหนดให้ A เป็นแดดออก ส่วน B เป็นเหตุการณ์ฝนไม่ตก จากสมการจะคำนวณได้

$$\text{โดยที่} \quad P(\text{แดดออก}|\text{ฝนจะไม่ตก}) = \frac{P(\text{แดดออก} \cap \text{ฝนจะไม่ตก})}{P(\text{ฝนจะไม่ตก})}$$

P (แดดออก | ฝนจะไม่ตก) คือค่า Conditional Probability คือค่าความน่าจะเป็นที่ฝนจะไม่ตกเกิดขึ้นก่อนแล้วแดดออกจึงตามมา

P (แดดออก  $\cap$  ฝนจะไม่ตก) คือค่า Joint Probability คือค่าความน่าจะเป็นที่แดดออกและฝนจะไม่ตกเกิดร่วมกัน

P (ฝนจะไม่ตก) คือค่าความน่าจะเป็นที่ฝนจะไม่ตกเกิดขึ้นก่อน

ในลักษณะเดียวกันนั้นเราจะเขียน  $P(\text{ฝนจะไม่ตก}|\text{แดดออก})$  หรือค่าความน่าจะเป็นที่แดดออกก่อนแล้วฝนจะไม่ตกตามมาก็ได้ตามสมการคือ

$$P(\text{ฝนจะไม่ตก}|\text{แดดออก}) = \frac{P(\text{แดดออก} \cap \text{ฝนจะไม่ตก})}{P(\text{แดดออก})}$$

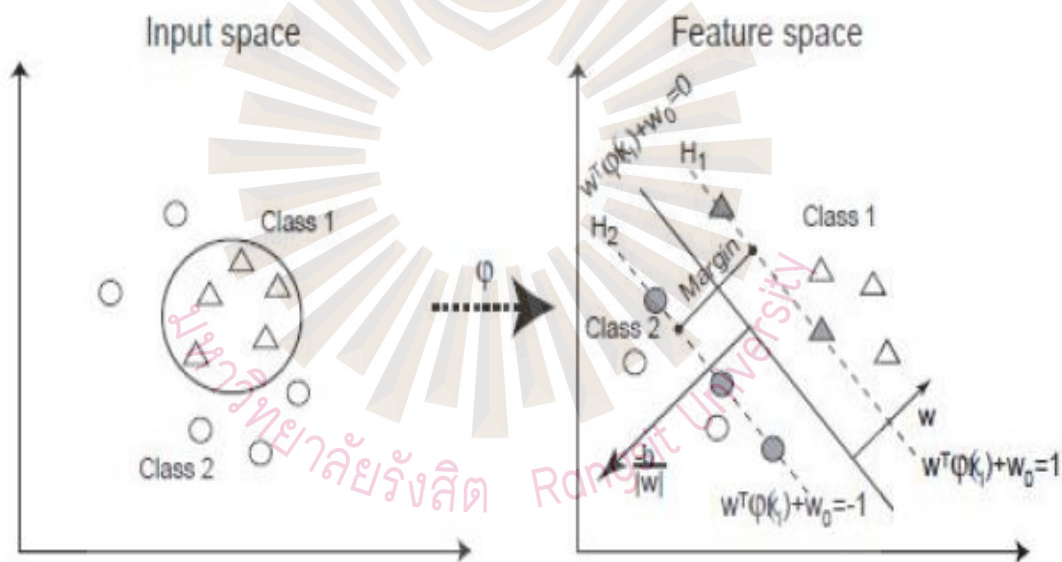
จากทั้ง 2 แบบ เราจะเห็นได้ว่ามี  $P(\text{แดดออก} \cap \text{ฝนจะไม่ตก})$  เหมือนกันจึงเขียนสมการได้ดังนี้

$$\begin{aligned} P(\text{แดดออก} \cap \text{ฝนจะไม่ตก}) &= P(\text{แดดออก} | \text{ฝนจะไม่ตก}) \times P(\text{ฝนจะไม่ตก}) = P(\text{ฝนจะไม่ตก} | \text{แดดออก}) \times P(\text{แดดออก}) \\ P(\text{ฝนจะไม่ตก} | \text{แดดออก}) &= \frac{P(\text{แดดออก} | \text{ฝนจะไม่ตก}) \times P(\text{ฝนจะไม่ตก})}{P(\text{แดดออก})} \end{aligned}$$

ดังตัวอย่างที่กล่าวมาเรียกว่าทฤษฎี Baye Theorem ในการนำไปใช้การจำแนกประเภทข้อมูลเป็นต้น

### 5) เทคนิคการจำแนกข้อมูลด้วยวิธี Support vector machine (SVM)

Support vector machine (SVM) นั้นเป็นเทคนิคที่ถูกคิดค้นขึ้นมาใหม่ และมีประสิทธิภาพสูง โดยใช้หลักการของโมเดลเชิงเส้นที่แบ่งข้อมูลเป็นสอง Class จากการหาสัมประสิทธิ์ของสมการเพื่อป้อนข้อมูลเข้าระบบเพื่อเรียนรู้จากการสร้างเส้นแบ่งแยกกลุ่มตามรูปที่ 2.8 แต่ด้วยรากฐานเดิมของ Support Vector Machine (SVM) ถูกออกแบบมาเพื่อใช้กับข้อมูลเชิงเส้น แต่ในการนำข้อมูลมาทำการเทรนนิ่งนั้นส่วนใหญ่จะเป็นข้อมูลที่ไม่ใช่เชิงเส้นตรงจึงต้องมีการแก้ปัญหาด้วยวิธีการ Kernel Function เพื่อทำการแปลงข้อมูลเพื่อให้โมเดลเชิงเส้นไปอยู่ในมิติที่สูงขึ้น (Dimension) วิธีนี้จะเป็นการเรียงตัวของข้อมูล เพื่อให้ได้พื้นที่มิติสูงขึ้น (Higher Dimensional Space) ในปัจจุบัน Kernel Function ที่นิยมใช้มีอยู่ 3 ชนิด (Hsu, Chang & Lin, 2003.) คือ โพลลิโนเมียล (Polynomial) เรเดียลเบสฟังก์ชัน (Radial Basis Function - RBF) และซิกมอยด์ (Sigmoid)



รูปที่ 2.8 รูปแสดงการเรียนรู้ของ Support vector machine (SVM)

ที่มา : เอกสิทธิ์ พัชรวงศ์ศักดิ์, 2557

## 2.2 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องเกี่ยวกับกระบวนการสุกของผลไม้ และการกำหนดพารามิเตอร์เพื่อเป็นเทคนิคในการทำเหมืองข้อมูลและนำมาใช้เป็นแนวทางในการออกแบบการทดลองให้เหมาะสมกับปัญหาการแยกแยะระดับการสุกของทุเรียน โดยวัตถุประสงค์ของผู้วิจัยคือ เพื่อให้เกษตรกรหรือผู้จำหน่ายทุเรียนสามารถจำหน่ายสินค้าได้ตรงตามความต้องการของผู้บริโภค และยังเป็น การช่วยเหลือเกษตรกรในด้านการจัดเก็บและด้านการขนส่ง ที่จะส่งผลในเรื่องของต้นทุนผลผลิต ซึ่งง่ายต่อการนำไปใช้งานอีกทั้งยังสามารถทดแทนการวิเคราะห์ระดับการสุกของทุเรียนในรูปแบบเดิม

ในกระบวนการสุกของผลไม้ นั้นสิ่งหนึ่งที่ต้องถือว่าเป็นปัจจัยและมีบทบาทที่สำคัญทั้งในการพัฒนาและการเสื่อมถอยของพืช นั่นคือ ก๊าซเอทิลีน เป็นก๊าซที่พืชหรือผลไม้ผลิตขึ้นมาเองตามตามธรรมชาติเป็นฮอร์โมนที่มีสถานะเป็นก๊าซ โดยเกิดขึ้นทุกเนื้อเยื่อของพืชและผลไม้โดยในงานวิจัยของ (Protasio Pereira, Galvão, Kobayashi, B.Caçaõ & Esteves Vieira, 2005) ได้กล่าวว่าเอทิลีนมีความสำคัญในการแบ่งระดับการสุกของกาแฟสายพันธุ์อาราบิก้าเป็นระดับต่าง ๆ โดยสามารถวัดระดับความเข้มข้นของก๊าซเอทิลีนได้

ในส่วนของการทำเหมืองข้อมูลนั้น ได้มีงานวิจัยหลากหลายงานวิจัยที่นำเทคนิคการวิเคราะห์ข้อมูลจากการทำเหมืองข้อมูลมาใช้ในการทำนาย เช่น งานทางกรบริหารเพื่อกำหนดกลุ่มลูกค้า งานทางด้านกรแพทย์เพื่อใช้วิเคราะห์แยกกลุ่มของอาการที่น่าจะเกิดโรค หรือเพื่อการคัดกรองผู้ป่วยที่อาจจะสงสัยว่าเป็นโรคต่าง ๆ ใช้ในส่วนของการพยากรอากาศเพื่อกำหนดสภาพภูมิอากาศ เป็นต้น เมื่อผู้วิจัยเล็งเห็นถึงความสามารถของเหมืองข้อมูลจึงคิดพัฒนาอุปกรณ์เพื่อใช้ตรวจจับก๊าซเอทิลีนแล้วผลลัพธ์ของจำนวนข้อมูลในลักษณะของตัวเลขนำมาใช้ทำเหมืองข้อมูลและวิเคราะห์โดยใช้เทคนิคต่างเปรียบเทียบกัน ผู้วิจัยได้ศึกษาเกี่ยวกับการวิเคราะห์โครงสร้างการตัดสินใจของ (Bhargava, Sharma, Bhargava & Mathuria, 2013) ได้กล่าวไว้ว่าต้นไม้ตัดสินใจเป็นเทคนิคที่สามารถใช้จัดหมวดหมู่ที่มีการกำกับดูแล โดยเฉพาะอย่างยิ่งเมื่อผลลัพธ์ถูกตีความโดยมนุษย์ และต้นไม้การตัดสินใจหลายตัวแปรที่ใช้ความสัมพันธ์ของแอตทริบิวต์ที่ปฏิบัติตามเงื่อนไขสามารถให้วิธีและผลลัพธ์ที่ดีที่สุด ซึ่งสอดคล้องกับงานวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของเทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับการเลือกตัวแทนที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม โดยเดช ธรรมศิริ และ พยุง มีสัง (2554) ได้กล่าวว่าการใช้เทคนิคต้นไม้ตัดสินใจสามารถวิเคราะห์ผลลัพธ์ที่ดีมาก โดยการเลือกใช้ตัวจำแนกข้อมูลที่เหมาะสมในการรวมกลุ่มด้วยวิธี



ขั้นตอนเชิงพันธุกรรมนี้จะทำให้โมเดลที่มีประสิทธิภาพมากยิ่งขึ้น ในส่วนของงานวิจัยที่ใช้เทคนิคโครงข่ายประสาทเทียม (Neural Network) นั้น (Rowley; Baluja & Kanade, 1998) ได้ทำการวิจัยเกี่ยวกับการตรวจจับใบหน้าด้วยเครือข่ายประสาทเทียมซึ่งสามารถใช้ตรวจจับใบหน้าเบื้องต้นได้ดีและมีประสิทธิภาพซึ่งสามารถพัฒนาต่อไปในเรื่องการจับเท็จในแง่ของกระบวนการพิสูจน์หลักฐานด้านนิติวิทยาได้ และในงานวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อน ภูริธยา อามฤครัตน์ และ พยุง มีสัง (2553) ได้ใช้เทคนิคการทำเหมืองข้อมูลโดยใช้โมเดล Support Vector Machines ในการทดลองซึ่งผลการทดลองสามารถแยกประเภทของผู้ป่วยที่เป็นโรคลมร้อนได้ผลดีมาก

ในส่วนของการพัฒนาอุปกรณ์ที่มีลักษณะความคล้ายคลึงและเกี่ยวข้องกับงานของผู้วิจัย นั่นคืออุปกรณ์วัดความสูงของทุเรียน ของวิทยาลัยการอาชีพบางสะพาน อ.บางสะพาน จ.ประจวบคีรีขันธ์ ซึ่งถูกพัฒนาโดย เมธี จำพวง, พันทิวา คำแก้ว, สมทรง แจ่มอักษร และ จิตวัฒนา บุญเลิศ (2561) ซึ่งกระบวนการทำงานของเครื่องวัดคุณรูปทุเรียนนี้จะมีลักษณะคล้ายกับปากกาโดยมีเข็มฉีดยาเสียบปลายเพื่อแทงเข้าไปวัดเนื้อทุเรียนที่อยู่ภายในลูกเมื่อกดเข็มลงไปเครื่องวัดจะตรวจระดับความดันหรือแรงหนืดด้านของเนื้อทุเรียนและส่งผลไปที่ด้านปลายอุปกรณ์ผู้ใช้งานจะสามารถอ่านค่าของทุเรียนได้คือเมื่อทุเรียนดิบจะเป็นสีแดง ทุเรียนห่ามจะขึ้นเป็นสีเหลือง และเมื่อทุเรียนสุกจะขึ้นเป็นสีเขียว ในการพัฒนาอุปกรณ์นี้กลุ่มผู้พัฒนาได้แบ่งกลุ่มของทุเรียนเป็น 3 ลักษณะคือ ดิบ ห่าม และสุก ซึ่งมีความสอดคล้องกับงานของผู้วิจัยในแง่ดิบ ห่าม สุก และแตกต่างกันคือ ผู้วิจัยแบ่งกลุ่มของทุเรียนเป็น 4 ลักษณะคือ ดิบ ห่าม สุก และงอม เป็นต้น

จากการศึกษางานวิจัยที่เกี่ยวข้อง โดยรวมนี้ทำให้ผู้วิจัยได้ทราบถึงข้อมูลต่าง ๆ เช่น กลุ่มของผลไม้ที่มีผลต่อกระบวนการสุก ทั้งในชนิดบ่มสุกและบ่มไม่สุก ซึ่งมีผลเกี่ยวเนื่องกับระดับปริมาณของก๊าซที่มาจากฮอร์โมนสำคัญรวมไปถึงสารเคมีที่พืชหรือผลไม้ปล่อยออกมาตามกระบวนการของธรรมชาติ ผู้วิจัยจะได้นำมาประกอบกับการพัฒนาอุปกรณ์ทั้งส่วนของฮาร์ดแวร์และในส่วนซอฟต์แวร์เพื่อใช้ในการตรวจจับก๊าซนั้น และนำไปถึงขั้นตอนสำคัญคือการทำเหมืองข้อมูลที่จะใช้ในการวิเคราะห์เพื่อหาการพยากรณ์ การคาดการณ์ และการทำนายผลที่สามารถนำมาประยุกต์ใช้กับการทดลองงานวิจัยดังกล่าว ซึ่งทำให้ทราบถึงกลุ่มข้อมูล ประเภทของข้อมูล โดยอาศัยเทคนิคการทำเหมืองข้อมูลในการวิเคราะห์ด้วยการใช้อัลกอริทึมที่แตกต่างกันไป เช่น แผนรูปต้นไม้เพื่อการตัดสินใจ (Decision Tree) การจัดแบ่งคลาสจากเทคนิคการค้นเพื่อนใกล้บ้าน (K-NN) โครงข่ายประสาทเทียม (Neural Network) การจำแนกประเภทข้อมูลด้วยวิธี (Naive Bayes)

และการวิเคราะห์ข้อมูลด้วยหลักการทางคณิตศาสตร์โดยการใช้สมการเส้นตรง (Support Vector Machine) เพื่อนำมาวิเคราะห์ทั้งในรูปแบบของการทดลองด้วยวิธีการเริ่มต้น และเพิ่มประสิทธิภาพ เพื่อให้ได้ผลลัพธ์ที่แม่นยำ และถูกต้องที่สุด





## บทที่ 3

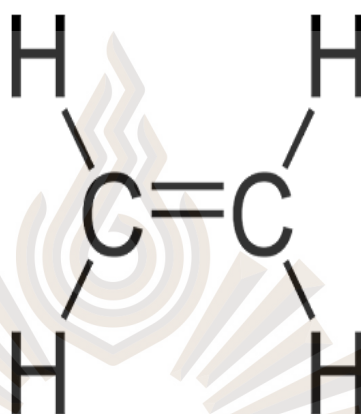
### ระเบียบวิธีการวิจัย

ปัญหาในการจัดเก็บ การขนส่ง และการซื้อขายผลไม้ นั้น มักเกิดกับผลไม้ที่สามารถบ่มสุกได้ เพราะถ้าเกษตรกร ผู้ประกอบการ และผู้จำหน่ายวางแผนในการจัดเก็บหรือดำเนินการขนส่งได้ไม่สมบูรณ์ จะทำให้ผลไม้ที่นำไปจัดจำหน่ายนั้นเกิดการเสียหายไม่สมบูรณ์หรือไม่ตรงตามความต้องการของผู้บริโภค ทำให้ต้องสูญเสียเวลาและค่าใช้จ่ายในการจัดเก็บสินค้าเพิ่มมากยิ่งขึ้น ผู้วิจัยจึงมีแนวคิดที่จะทำการวิเคราะห์ระดับการสุกของผลไม้จากการแพร่กระจายของก๊าซเอทิลีนโดยใช้เทคนิคการทำเหมืองข้อมูล และเลือกทุเรียนมาเป็นตัวแทนของผลไม้ที่สามารถบ่มสุกได้ โดยไม่สามารถสังเกตเห็นจากภายนอกได้ว่ามีระดับการสุกมากน้อยเพียงใด มาใช้ในการทดลอง และผู้วิจัยยังเล็งเห็นถึงความสำคัญของทุเรียน ที่เป็นผลไม้เศรษฐกิจของประเทศไทยที่ได้รับความนิยมไปทั่วโลกหากนำการทดลองนี้ไปใช้ในธุรกิจการค้าทุเรียน จะส่งผลดีให้กับผู้ประกอบการในด้านความแม่นยำในการเลือกจำหน่ายตรงตามความต้องการของผู้บริโภคเป็นอย่างมาก โดยผู้วิจัยจะแบ่งขั้นตอนการทดลอง 3 ขั้นตอนตามวัตถุประสงค์ได้แก่ เพื่อพัฒนาเครื่องมือตรวจจับก๊าซเอทิลีนจากทุเรียน เพื่อทดลองแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทิลีน โดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering) และเพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีนโดยใช้เทคนิคการจำแนกข้อมูล (Classification) โดยจะอธิบายดังต่อไปนี้

#### 3.1 การพัฒนาเครื่องมือการตรวจจับก๊าซเอทิลีน

เมื่อผู้วิจัยใช้แนวความคิดที่จะทำการวิเคราะห์ระดับการสุกของทุเรียนโดยการดักจับการแพร่กระจายของก๊าซเอทิลีน ผู้วิจัยจึงพัฒนากล่องเครื่องมือเพื่อทำการตรวจจับก๊าซดังกล่าวเพื่อนำข้อมูลระดับปริมาณความเข้มข้นของก๊าซนี้ไปใช้ทดลองด้วยเทคนิคการทำเหมืองข้อมูล โดยผู้วิจัยจึงศึกษาเกี่ยวกับอุปกรณ์ทั้งในส่วนที่เป็นฮาร์ดแวร์ ซอฟต์แวร์ โครงสร้างของสารเคมี รวมไปถึงปัจจัยหรือสิ่งต่างๆ ที่เกี่ยวข้องกับกระบวนการสุกของผลไม้ เพื่อให้เกิดความเข้าใจและสามารถเลือกใช้อุปกรณ์ที่เกี่ยวข้องกับการพัฒนาเครื่องมือได้ตรงตามความต้องการและเหมาะสมสอดคล้องตามวัตถุประสงค์ของผู้วิจัยในส่วนของก๊าซเอทิลีนนั้น หากพิจารณาตามสูตรทางเคมีคือ  $C_2H_4$  (Eberhardt, Schmitt & Wöllenstein, 2017) ซึ่งหมายถึง คาร์บอน 2 อะตอม และไฮโดรเจน 4 อะตอม

ตามรูปที่ 3.1 และเมื่อพิจารณาทางกระบวนการวิทยาศาสตร์แล้วนั้น (Scientific) จะพบว่าก๊าซเอทิลีน (Ethylene) มีความสัมพันธ์กับกระบวนการสุกของผลไม้ คือปริมาณของก๊าซเอทิลีนจะต่ำเมื่อผลไม้ยังดิบอยู่ แต่จะมีปริมาณสูงขึ้นเมื่อผลไม้เริ่มมีการเปลี่ยนแปลงสภาพหรือมีอายุการสุกขึ้น จึงถือได้ว่าปริมาณความเข้มข้นของเอทิลีนเป็นปัจจัยสำคัญในการสุกของผลไม้ (Liu, Pirrello, Chervin, Roustan, & Bouzayen, 2015 ; Nath, Bouzayen, Mattoo, & Pech, 2014)



รูปที่ 3.1 แสดงสูตรทางเคมีของเอทิลีน

ที่มา : ผู้วิจัย

จากนั้นผู้วิจัยจึงพัฒนาเครื่องมือโดยการจำลองการทำงานจากลักษณะการดมกลิ่นของมนุษย์ที่มีจมูกสำหรับการรับกลิ่นด้วยวิธีการสุดดมหายใจเอาอากาศที่ห่อหุ้มรอบวัตถุเป้าหมายผ่านเข้าทางช่องโพรงจมูกที่มีจุดรับกลิ่น และนำอากาศที่ห่อหุ้มรอบวัตถุเป้าหมายจะผ่านจุดรับกลิ่นส่งไปยังกระบวนการวิเคราะห์กลิ่น และจึงส่งผลเป็นข้อมูลไปประมวลผลในสมองเพื่อบอกว่าวัตถุนั้นมีกลิ่นแบบใดหรือเป็นกลิ่นของอะไร ผู้วิจัยจะสร้างกล่องขึ้นโดยมีการปิดกั้นมิให้มีอากาศจากภายนอกผ่านเข้าไปได้และแบ่งภายในเป็นสองช่อง ในช่องหนึ่งใช้สำหรับใส่ทุเรียนที่เป็นวัตถุเป้าหมาย ส่วนในช่องสองใช้สำหรับติดตั้งอุปกรณ์ต่าง ๆ ซึ่งมี MQ Gas Sensor ตามรูปที่ 3.2 ตามคุณสมบัติของแต่ละเซ็นเซอร์ แล้วใช้พัดลมเป็นตัวดึงเอาอากาศรอบวัตถุเป้าหมายจากช่องหนึ่งผ่านไปส่งช่องสองที่เป็นจุดรับกลิ่นหรือจุดตรวจวัดก๊าซเอทิลีนจากนั้นจะใช้อาณูโน้เป็นจุดประมวลผลและนำข้อมูลที่เก็บได้ไปใช้วิเคราะห์ต่อไป ในส่วนการทำงานของอาณูโน้ที่เป็นไมโครคอนโทรลเลอร์บอร์ดนั้น ตามรูปที่ 3.3



รูปที่ 3.2 แสดง MQ Gas Sensor

ที่มา : Arduino, 2018



รูปที่ 3.3 แสดงอาดูโน่

ที่มา : Arduino, 2018

ในการทดลองนี้ผู้วิจัยจะนำมาเขียนโปรแกรมเพื่อเชื่อมต่อกับก๊าซเซ็นเซอร์ (Gas Sensor) ตามคุณสมบัติของเซ็นเซอร์ในลักษณะต่าง ๆ หนึ่งหากนำเอทิลีนก๊าซเซ็นเซอร์ (Ethylene Gas Sensor) มาเชื่อมต่อกับอาดูโน่ก็จะสามารถตรวจจับก๊าซเอทิลีนได้เช่นกัน แต่อย่างไรก็ตามเอทิลีนก๊าซเซ็นเซอร์ นั้นไม่มีจำหน่ายโดยทั่วไปในประเทศไทย อีกทั้งยังมีราคาสูงเมื่อเทียบกับก๊าซเซ็นเซอร์อื่น ๆ เช่น ไฮโดรเจนก๊าซเซ็นเซอร์ (Hydrogen Gas Sensor) แอลกอฮอล์ก๊าซเซ็นเซอร์ (Alcohol Gas Sensor) ดังนั้นผู้วิจัยจึงออกแบบพัฒนาเครื่องมือ โดยใช้ MQ3 Gas Sensor จับแอลกอฮอล์;

เอทานอล (Ethanol;  $C_2H_5OH$ ) ซึ่งเป็นองค์ประกอบระหว่าง เอทิลีน ( $C_2H_4$ ) และน้ำ ( $H_2O$ ) (Adams & Yang, 1978) จากนั้นผู้วิจัยจะใช้ MQ8 Gas Sensor จับก๊าซไฮโดรเจน (Hydrogen) และในส่วนของ ก๊าซคาร์บอน (C) นั้น ยังไม่มีเซ็นเซอร์ที่ใช้ตรวจจับได้ ผู้ศึกษาจึงใช้ MQ6 Gas Sensors ที่เป็น เซ็นเซอร์ตรวจจับ LPG ตรวจจับแทนได้เนื่องจากก๊าซ LPG มีส่วนประกอบด้วยของ โพรเพน และบิวเทน ซึ่งมีคาร์บอนเป็นธาตุประกอบหลัก (Hashim, 2008) ผู้วิจัยจะนำเอาอุปกรณ์ทำการวัดระดับ ก๊าซแอลกอฮอล์ ไฮโดรเจนและคาร์บอนเมื่อนำทุเรียนใส่ในอุปกรณ์จะ ได้ข้อมูลที่แสดงถึงปริมาณ ของก๊าซที่คลายออกมา และจะนำเอาข้อมูลมาเก็บเป็นข้อมูลในรูปแบบของสถิติโดยใช้เก็บข้อมูล จากทุเรียนที่มีลักษณะต่าง ๆ ตามรูปแบบของธรรมชาติ เช่น คีบ สุก และงอม หลังจากเก็บข้อมูลชุด แรกมาได้แล้ว ผู้วิจัยจึงนำไปกรองข้อมูลด้วยสูตรการคำนวณทางคณิตศาสตร์ ตามสมการที่ 3-1 โดย การนำจำนวนปริมาณก๊าซแต่ละเซ็นเซอร์ที่เริ่มทำการดักจับตั้งแต่วันที่ 1 – 40 มารวมกันแล้วหาร ด้วยจำนวนระยะเวลา จะทำให้ได้ถึงข้อมูลที่มีค่าเฉลี่ยเคลื่อนที่จากจำนวนทุเรียน ทั้งหมดเพื่อนำ ข้อมูลไปทดสอบด้วยเทคนิคการทำเหมืองข้อมูล

$$\begin{aligned}
 MA(40) &= \sum_{i=1}^{40} P_i \\
 &= (P_1 + P_2 + P_3 + \dots + P_{38} + P_{39} + P_{40})
 \end{aligned}
 \tag{3-1}$$

### 3.2 เพื่อกำหนดกลุ่มของระดับการสุกของทุเรียน

หลังจากที่ผู้วิจัยได้พัฒนาเครื่องมือที่ใช้ตรวจจับก๊าซเอทิลีนได้แล้วนั้น ในขั้นต่อไปผู้วิจัย ได้ศึกษางานวิจัยที่นำเสนอเกี่ยวกับระดับการสุกของทุเรียนจะแบ่งออกเป็น 3 ระดับ (Avila, et al., 2008) คือ คีบ (Unripe) สุก (Ripe) และงอม (Overripe) แต่ในสภาพความเป็นจริงแล้วคนไทย บางส่วนชอบรับประทานทุเรียนห่าม (Pre Ripe) และในงานวิจัยที่ศึกษาเกี่ยวกับการเปลี่ยนแปลง ทางกายภาพของมะละกอ ได้กล่าวถึงระดับการสุกที่เรียกว่า ห่าม ไว้ว่าสีผิวของมะละกอ จะเป็นสี เหลืองประมาณ 10-20% (Krongyut, et al., 2011) ประกอบกับยังมีผลไม้อื่น ๆ เช่น กกล้วย มะม่วง ที่มีระดับการสุกที่เรียกว่าห่าม ซึ่งผลไม้นี้เหล่านี้จะมีการเปลี่ยนสีผิวจากสีเขียวเป็นสีเหลืองใน

บางส่วน ในการวิจัยนี้ผู้วิจัยจึงขอแบ่งระดับการสุขของทุเรียนเป็น 4 ระดับ คือ ดิบ ห้าม สุข และ งาม เนื่องจากทุเรียนดิบสามารถนำไป ทำเป็นทุเรียนทอดได้ ส่วนทุเรียนงามสามารถนำไปเป็น ทุเรียนกวนได้ แต่สำหรับทุเรียนห้าม และสุข มักนำไปจำหน่ายให้กับผู้บริโภค โดยไม่ต้องแปรรูป ซึ่งผู้จำหน่ายจะต้องใช้อุปกรณ์ไม้เคาะร่วมกับการสังเกตสีของหนามเพื่อระบุว่าทุเรียนสุกในระดับใด แต่ก็ไม่สามารถมั่นใจได้จึงต้องมีการผ่าออกบางส่วนเพื่อให้ผู้บริโภคลองใช้นี้ว่าตกลงไปที่เนื้อทุเรียน เพื่อตรวจสอบอีกครั้งว่ามีระดับความสุกตามที่ต้องการหรือไม่ ซึ่งวิธีการดังกล่าวยังไม่มีความแม่นยำ และไม่สามารถสร้างความน่าเชื่อถือให้กับผู้บริโภคได้ อีกทั้งยังจำเป็นต้องผ่าทุเรียนออกบางส่วนจึง จะสามารถตรวจสอบระดับความสุกของทุเรียนได้ เมื่อผู้วิจัยได้ศึกษาเกี่ยวกับกระบวนการสุขของ ผลไม้ที่สามารถบ่มสุกได้ เมื่อนำมาบ่มต่อหรือปล่อยให้ทิ้งไว้ในสภาพที่เหมาะสมจะเกิดการเปลี่ยนแปลง หลายประการ อาทิเช่น รสชาติ สี กลิ่น แล้วนั้นผู้วิจัยจะนำข้อมูลที่ผ่านการคัดกรองไปใช้วิเคราะห์ ด้วยเทคนิคการทำเหมืองข้อมูล (Data Mining) และใช้เทคนิคการแบ่งกลุ่ม (K-Mean) ในการตรวจสอบ เบื้องต้นว่าข้อมูลที่จัดเก็บไว้นั้นสามารถแบ่งทุเรียนออกได้เป็นกี่กลุ่ม ซึ่งผู้วิจัยได้ศึกษาจากงานวิจัย ที่เกี่ยวข้อง พบว่า K-Mean สามารถใช้ในการจัดกลุ่มเมื่อทราบจำนวนกลุ่มที่แน่นอน (สุรวัชร ศรีเปารยะ และ สายชล ดินสมบูรณ์ทอง, 2560) และ (วิระยุทธ พิมพากรณ์ และ พยุง มีสัง, 2557) ในการออกแบบ การทดลองด้วยเทคนิคการแบ่งกลุ่มนั้น ผู้วิจัยได้กำหนดค่า K ในรูปแบบต่าง ๆ และใช้เทคนิค Classification เพื่อสร้างโมเดลในการตรวจสอบระดับการสุขของทุเรียน

### 3.3 การจำแนกประเภทการสุขของทุเรียนจากการแพร่กระจายของก๊าซ

เมื่อทราบถึงกลุ่มระดับการสุขของทุเรียนแล้วนั้น ผู้วิจัยจะใช้เทคนิค Classification เพื่อสร้างโมเดลในการตรวจสอบระดับการสุขของทุเรียน โดยทำการเปรียบเทียบประสิทธิภาพ เพื่อให้ได้ผลลัพธ์ ผู้วิจัยเริ่มต้นการทดลองด้วยวิธีการของ Cross Validation (RapidMiner, 2017) มาทำการแบ่งข้อมูลเพื่อใช้ในการทดลอง 80% และเก็บไว้ใช้ทดสอบ 20% จากนั้นจะใช้อัลกอริทึม ที่แตกต่างกันในการสร้างโมเดล โดยเริ่มจากการสร้างรูปแบบที่ใช้ในการประเมินระดับการสุขจาก การสร้างโมเดลในการทำนายผลได้แก่ Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naïve Bayes และ Support Vector Machine (SVM) โดยในการทดสอบนั้น โดยเบื้องต้น ผู้วิจัยกำหนดให้ทุกอัลกอริทึม ใช้การทดสอบแบบค่าเริ่มต้น (Default Parameter set) เพื่อสร้าง โมเดลและนำโมเดลที่ได้มาเปรียบเทียบกันว่า โมเดลไหนมีประสิทธิภาพที่สุด จากนั้นผู้วิจัยได้ ทดลองทำการทดสอบแบบเพิ่มประสิทธิภาพ (Optimize Parameter Set) และทำการทดลองอีกครั้ง



เพื่อเปรียบเทียบประสิทธิภาพกัน เมื่อ ได้ผลจะนำ โมเดลที่ได้มาเปรียบเทียบกันเพื่อดูว่า โมเดลใดมีความแม่นยำในการจำแนกกลุ่มข้อมูล (Accuracy) สูงที่สุด การนี้ผู้วิจัยได้ออกแบบการทดลองแบบการเพิ่มประสิทธิภาพโดยใช้ RapidMiner ในการศึกษา อัลกอริทึมแต่ละวิธีดังนี้

### 3.3.1 ต้นไม้ตัดสินใจ (Decision Tree)

ผู้วิจัยจะใช้ข้อมูลแต่ละคุณลักษณะเพื่อให้ระบบพิจารณาค่าความแตกต่างของปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบตัดสินใจแบ่งเป็นกลุ่มย่อย ๆ และพิจารณาเลือกกลุ่มตามคุณลักษณะต่าง ๆ จากผลลัพธ์ของค่าปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบในสถานะปัจจุบันกับปริมาณที่บอกถึงความไม่เป็นระเบียบของระบบของสถานะก่อนหน้า หรือ Normalized Information Gain (IG) ที่สูงสุด โดยที่ผู้วิจัยจะเลือกปรับใช้พารามิเตอร์ Decision Tree.criterion, Decision Tree.maximal\_depth, Decision Tree.confidence และ Decision Tree.apply\_pruning ในการทดลองเนื่องจากเป็นพารามิเตอร์หลักของอัลกอริทึม

### 3.3.2 เพื่อนบ้านที่ใกล้ที่สุด (K Nearest Neighbors (KNN))

ผู้วิจัยทดลองใช้พารามิเตอร์ โดยการใช้ k-NN.kernel\_type, k-NN.k, k-NN.weighted\_vote, k-NN.numerical\_measure และ k-NN.measure\_types เนื่องจากเป็นฟังก์ชันหลักและยังเป็นพื้นฐานของอัลกอริทึม ซึ่งสามารถกำหนดน้ำหนักและระยะห่าง เพื่อกำหนดค่า K โดยใช้ Cross – Validation โดยในการทดลองนี้ผู้วิจัยจะกำหนดค่า K = 1 เพื่อให้เกิดความแม่นยำมากที่สุด

### 3.3.3 เครือข่ายประสาทเทียม (Neural Network)

ผู้วิจัยจะนำค่าข้อมูลชุดขาเข้าไปกำหนดน้ำหนัก (Weight) ระหว่างโหนด และปรับเปลี่ยนค่าน้ำหนักด้วยการนำค่าน้ำหนักคูณกับข้อมูลชุดขาเข้าของแต่ละขา ผลที่ได้จากทุก ๆ ขานั้นจะนำมาเปรียบเทียบกับกับเกณฑ์การแบ่งค่า (Threshold) ที่กำหนดไว้โดยผู้วิจัยจะทำการทดลองกำหนดพารามิเตอร์ด้วย Neural Net.learning\_rate เพื่อคูณกับ Error Rate และ โหนดที่เข้ามาแล้วบวกเข้ากับน้ำหนักข้อมูลเพื่อทำให้เกิดน้ำหนักของข้อมูลชุดใหม่ Neural Net.training\_cycles เพื่อระบุจำนวนรอบการทดสอบในการส่งค่าย้อนกลับ (Back-Propagation) ค่าข้อมูลชุดขาออกจะถูกเปรียบเทียบกับคำตอบที่ถูกต้องเพื่อคำนวณค่าของฟังก์ชัน Error-Defined ที่กำหนดไว้ล่วงหน้า



ข้อผิดพลาดจะถูกส่งกลับผ่านเครือข่าย การใช้ข้อมูลนี้อัลกอริทึมจะปรับน้ำหนักของแต่ละการเชื่อมต่อเพื่อลดค่าของฟังก์ชันข้อผิดพลาดด้วยจำนวนเล็กน้อย `Neural.Net.momentum` เพื่อใช้ในการปรับค่าน้ำหนักและใช้เพื่อเหนี่ยวนำค่าข้อมูลชุดขาออกที่ได้ในแต่ละรอบ เพื่อป้องกันการเกิดข้อผิดพลาดจากการทดสอบข้อมูลและใช้ `Neural.Net.error_epsilon` ใช้เพื่อการเพิ่มประสิทธิภาพและจะหยุดทำงานหากค่าการทดสอบข้อมูลต่ำกว่าค่า `Epsilon`

### 3.3.4 Naïve Bayes

ผู้วิจัยจะทำการทดลองจัดกลุ่มโดยใช้หลักความน่าจะเป็นซึ่งเป็นพื้นฐานและสมมุติฐานของ Bayes' Theorem โดยการใช้ความสัมพันธ์ของแต่ละโหนดและกำหนดให้มีความสัมพันธ์กันตามแบบที่แบบจำลอง โดยจะนำค่าตัวแปรที่เกี่ยวข้องเก็บไว้ในแต่ละโหนดโดยใช้เงื่อนไขความน่าจะเป็น หากภายในข้อมูลการทดลอง มีค่าแอตทริบิวต์ที่ระบุไม่เคยเกิดขึ้นในบริบทของคลาสที่ระบุความเป็นไปได้เชิงเงื่อนไขจะถูกตั้งค่าเป็นศูนย์เมื่อค่าศูนย์นี้คูณพร้อมกับความน่าจะเป็นอื่น ๆ ค่าเหล่านี้จะถูกตั้งค่าเป็นศูนย์และผลลัพธ์จะทำให้เข้าใจผิด ผู้วิจัยจึงทดลองปรับพารามิเตอร์ `Naive Bayes.laplace_correction` เพื่อหลีกเลี่ยงปัญหานี้ให้เพิ่มหนึ่งรายการในแต่ละจำนวนเพื่อหลีกเลี่ยงการเกิดค่าเป็นศูนย์ ในการเพิ่มจำนวนนี้นั้นมีผลเพียงน้อยนิดต่อความน่าจะเป็นในส่วนของ `Naive Bayes (Kernel).estimation_mode` ใช้เพื่อระบุความหนาแน่นของเคอร์เนลโดยใช้ตัวเลือกสองแบบคือ `Full`, `Greedy` และใช้ `Naive Bayes (Kernel).bandwidth_selection` เพื่อระบุความหนาแน่น ความเร็ว ของเคอร์เนลในแอปพลิเคชันแบบจำลอง

### 3.3.5 Support Vector Machine (SVM)

ผู้วิจัยจะทำการแปลงข้อมูลให้อยู่ในสมการเส้นตรงด้วยการใช้วิธีการ `Kernel Function` เพื่อแปลงข้อมูลให้โมเดลเชิงเส้น ไปอยู่ในมิติที่สูงขึ้นด้วยการเรียงตัวของข้อมูลนี้จะได้พื้นที่มิติสูงขึ้น (`Higher Dimensional Space`) ผู้วิจัยจะทดลองปรับพารามิเตอร์ `SVM.kernel_type` โดยจะทดสอบแบบ `Linear`, `Polynomial`, `Radial Basis Function - RBF` และ `Sigmoid` จากนั้นจะใช้ `SVM.gamma` เพื่อใช้ตรวจสอบเคอร์เนล `SVM.kernel_type` และใช้ `SVM.C` เพื่อป้องกันข้อผิดพลาดของ `SVM.svm_type` ที่เป็น `c-SVC`, `epsilon-SVR`, `nu-SVR`

## บทที่ 4

### ผลการวิจัย

ผู้วิจัยขอเสนอผลการวิจัยตามวัตถุประสงค์ ได้แก่ เพื่อพัฒนาเครื่องมือตรวจจับก๊าซเอทีลีนจากทุเรียน เพื่อทดลองแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทีลีนโดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล และเพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลีน โดยใช้เทคนิคการจำแนกข้อมูล ซึ่งผลการวิจัยที่ได้มีรายละเอียด ดังนี้

#### 4.1 ผลการพัฒนาเครื่องมือการตรวจจับก๊าซเอทีลีน

หลังจากที่ผู้วิจัยพัฒนาอุปกรณ์เพื่อทำการตรวจก๊าซเอทีลีน ตามรูปที่ 4.1 นั้น ผู้วิจัยได้ติดตั้งเซ็นเซอร์ชนิดต่างๆ เข้ากับอุปกรณ์ ได้แก่ MQ3, MQ6 และ MQ8 และทำการเขียนโปรแกรมเพื่อกำหนดค่าและสั่งงานอุปกรณ์ในรูปแบบต่าง ๆ ตามรูปที่ 4.2 ก่อนนำไปใช้ทำการทดลองวัดปริมาณก๊าซของทุเรียนที่นำมาใช้เป็นตัวอย่างเป็นจำนวน 160 ลูก ซึ่งจากการพัฒนาอุปกรณ์ดังกล่าวสามารถเก็บข้อมูลปริมาณก๊าซของแต่ละเซ็นเซอร์ที่ปล่อยออกมาจากทุเรียน และสามารถนำเอาข้อมูลไปใช้ในรูปแบบของสถิติได้ต่อไป



รูปที่ 4.1 แสดงอุปกรณ์วัดระดับการสุกของทุเรียน

ที่มา : ผู้วิจัย

```

Arrays | Arduino 1.8.1
File Edit Sketch Tools Help

int timer = 100;           // The higher the number, the slower the timing.
int ledPins[] = {
  2, 7, 4, 6, 5, 3
};           // an array of pin numbers to which LEDs are attached
int pinCount = 6;        // the number of pins (i.e. the length of the array)

void setup() {
  // the array elements are numbered from 0 to (pinCount - 1).
  // use a for loop to initialize each pin as an output:
  for (int thisPin = 0; thisPin < pinCount; thisPin++) {
    pinMode(ledPins[thisPin], OUTPUT);
  }
}

void loop() {
  // loop from the lowest pin to the highest:
  for (int thisPin = 0; thisPin < pinCount; thisPin++) {
    // turn the pin on:
    digitalWrite(ledPins[thisPin], HIGH);
    delay(timer);
    // turn the pin off:
    digitalWrite(ledPins[thisPin], LOW);
  }

  // loop from the highest pin to the lowest:
  for (int thisPin = pinCount - 1; thisPin >= 0; thisPin--) {

```

```

ArduinoISP | Arduino 1.8.1
File Edit Sketch Tools Help

}
parameter:
parameter param:
// this provides a heartbeat on pin 9, so you can tell the software is running.
uint8_t hbval = 128;
int8_t hbdelta = 8;
void heartbeat() {
  static unsigned long last_time = 0;
  unsigned long now = millis();
  if ((now - last_time) < 40)
    return;
  last_time = now;
  if (hbval > 192) hbdelta = -hbdelta;
  if (hbval < 32) hbdelta = -hbdelta;
  hbval += hbdelta;
  analogWrite(LED_HB, hbval);
}

static bool rst_active_high;

void reset_target(bool reset) {
  digitalWrite(RESET, ((reset && rst_active_high) || (!reset && !rst_active_high)) ? HIGH : LOW);
}

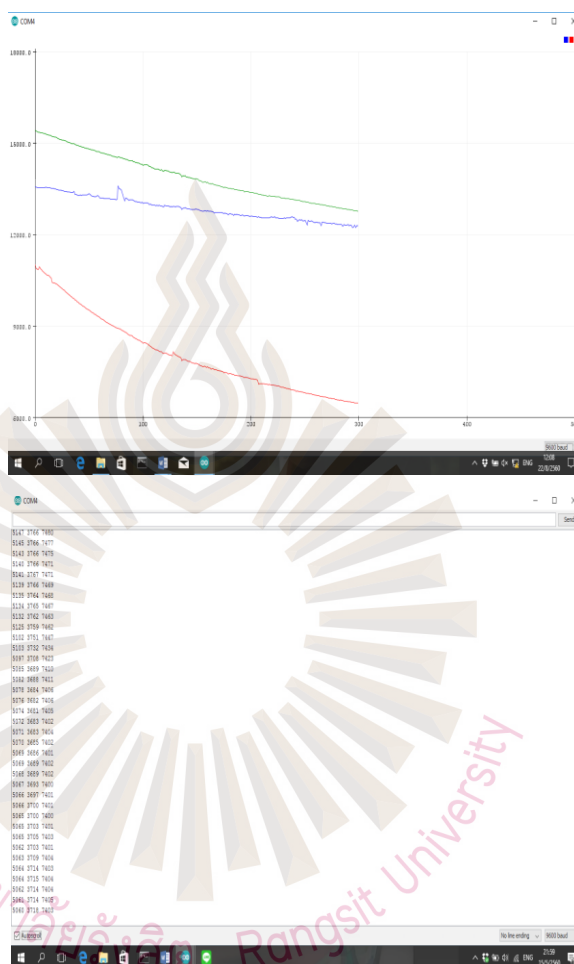
void loop(void) {
  // in mode continue

```

รูปที่ 4.2 แสดง โปรแกรมกำหนดค่าและตั้งงานอุปกรณ์

ที่มา : ผู้วิจัย

ในการเก็บข้อมูลจากอุปกรณ์วัดก๊าซนั้น ผลของข้อมูลจะแสดงผลเป็นลักษณะกราฟและตัวเลข ผู้วิจัยจึงเริ่มเก็บข้อมูลเมื่อกราฟของข้อมูลขึ้นสูงที่สุดและเริ่มเป็นเส้นตรง (Saturation Point) จึงเริ่มเก็บข้อมูลที่เป็นตัวเลขตามรูปที่ 4.3



รูปที่ 4.3 แสดงปริมาณก๊าซออกมาเป็นลักษณะกราฟ และตัวเลข  
ที่มา : ผู้วิจัย

หลังจากได้ข้อมูลที่ยังไม่ผ่านกระบวนการใด ๆ มาได้แล้วนั้นผู้วิจัยจึงนำปริมาณก๊าซแต่ละเซ็นเซอร์ที่เริ่มทำการปักจับนำไปกรองข้อมูลด้วยสูตรการคำนวณทางคณิตศาสตร์ ผลที่ได้ก็จะทำให้ได้ข้อมูลที่มีค่าเฉลี่ยเคลื่อนที่ตามตามรูปที่ 4.4 และนำไปทดสอบแบ่งกลุ่มด้วยเทคนิคการทำเหมืองข้อมูล

No	set 2			avg	set 3			avg	set 4			avg	set 5			avg
	Blue	Red	Green		Blue	Red	Green		Blue	Red	Green		Blue	Red	Green	
1	10891	5486	11232	9139.67	10250	5048	10800	8733.00	10133	5073	10680	8630.07	9779	4908	10500	8386.00
2	10897	5488	11230	9137.67	10345	5047	10799	8730.33	10144	5099	10680	8633.00	9778	4908	10500	8386.67
3	10704	5480	11227	9137.00	10319	5055	10796	8723.33	10132	5088	10684	8627.33	9784	4911	10499	8386.00
4	10709	5472	11225	9135.33	10339	5059	10792	8730.00	10187	5084	10677	8630.00	9782	4911	10499	8387.33
5	10715	5473	11226	9136.00	10340	5062	10792	8733.33	10186	5084	10683	8637.67	9783	4914	10500	8386.00
6	10721	5473	11225	9139.67	10274	5060	10788	8707.33	10184	5082	10682	8630.00	9790	4913	10497	8400.00
7	10725	5473	11223	9140.33	10269	5061	10787	8702.33	10182	5059	10679	8623.00	9797	4914	10499	8390.00
8	10727	5486	11221	9139.00	10331	5062	10790	8728.33	10189	5059	10679	8633.33	9797	4915	10500	8404.33
9	10730	5487	11220	9139.00	10188	5059	10782	8676.33	10189	5059	10671	8633.33	9796	4916	10500	8400.00
10	10730	5485	11217	9137.33	10180	5051	10778	8659.67	10184	5059	10679	8623.00	9799	4916	10500	8400.00
11	10731	5481	11213	9135.00	10124	5047	10778	8649.67	10186	5054	10672	8631.67	9846	4921	10500	8423.33
12	10733	5453	11212	9132.67	10131	5048	10774	8661.00	10181	5062	10672	8628.33	9835	4926	10500	8425.33
13	10731	5450	11208	9129.67	10131	5043	10789	8647.67	10188	5060	10693	8626.67	9878	4930	10500	8472.33
14	10733	5445	11209	9129.00	10141	5048	10771	8652.67	10188	5048	10693	8624.33	9866	4932	10500	8480.67
15	10735	5445	11209	9128.67	10144	5043	10788	8651.67	10148	5048	10670	8622.00	9873	4933	10510	8472.00
16	10736	5442	11202	9128.67	10149	5044	10788	8653.67	10132	5047	10670	8623.00	9872	4934	10510	8472.00
17	10739	5442	11199	9129.67	10150	5048	10795	8653.67	10131	5048	10670	8623.00	9868	4933	10510	8470.33
18	10738	5440	11198	9129.00	10151	5050	10793	8657.00	10134	5047	10683	8623.33	9872	4927	10500	8392.00
19	10730	5444	11198	9134.00	10184	5051	10795	8658.00	10180	5047	10693	8628.33	9872	4934	10511	8472.33
20	10737	5453	11197	9130.67	10229	5053	10791	8691.00	10148	5038	10695	8614.33	9854	4934	10500	8465.33
21	10738	5449	11195	9130.67	10229	5049	10795	8677.67	10180	5041	10681	8621.67	9890	4936	10500	8464.67
22	10758	5448	11194	9133.33	10377	5051	10784	8734.00	10156	5042	10695	8619.67	9844	4938	10510	8463.33
23	10730	5448	11192	9122.67	10388	5063	10783	8731.33	10184	5041	10684	8619.67	9836	4938	10500	8480.33
24	10760	5452	11191	9134.33	10356	5065	10786	8728.33	10148	5041	10683	8613.33	9827	4938	10507	8458.67
25	10751	5458	11189	9132.67	10347	5064	10784	8725.00	10145	5040	10681	8616.67	9824	4938	10503	8457.00
26	10736	5459	11188	9129.67	10323	5067	10783	8719.33	10139	5039	10680	8612.67	9825	4938	10507	8458.67
27	10791	5459	11180	9140.67	10347	5068	10789	8724.67	10137	5037	10683	8610.67	9817	4938	10508	8454.33
28	10833	5460	11191	9140.33	10334	5067	10797	8719.33	10138	5037	10682	8610.33	9855	4932	10487	8370.67
29	10833	5461	11190	9170.00	10320	5061	10795	8710.33	10140	5036	10681	8610.67	9859	4923	10467	8371.67
30	10888	5459	11184	9169.33	10332	5072	10791	8721.67	10110	5030	10686	8599.67	9842	4928	10454	8379.00

รูปที่ 4.4 แสดงปริมาณข้อมูลที่เก็บได้จากทุเรียน  
ที่มา : ผู้วิจัย

จากนั้นผู้วิจัยนำข้อมูลจากรูปที่ 4.4 ไปวิเคราะห์โดยใช้การทดสอบแบ่งกลุ่มด้วยเทคนิคการทำเหมืองข้อมูลเมื่อได้ข้อมูลที่ผ่านกริดคัดกรองแล้วนั้นผู้วิจัยได้เชิญผู้เชี่ยวชาญทำการประเมินระดับการสุกและทำการจัดเก็บค่าต่าง ๆ เรียบร้อยนั้น ผู้วิจัยจึงนำข้อมูลที่จัดเก็บได้นำไปทำเหมืองข้อมูล ซึ่งผลการวิจัยแสดงผลความสัมพันธ์ระหว่างระดับปริมาณก๊าซกับระดับการสุกของทุเรียนตามตารางที่ 4.1 และผลการทดลองกำหนดกลุ่มได้ดังนี้

ตารางที่ 4.1 แสดงการจัดกลุ่มความเข้มข้นของก๊าซต่าง ๆ

ID	กลุ่ม	ลูกที่	แอลกอฮอล์	ไฮโดรเจน	คาร์บอน
1	cluster_3	1	11538.450	5709.650	11828.975
2	cluster_3	2	10756.075	5456.300	11199.325
3	cluster_3	3	11116.700	6101.350	10519.950
.					
.					
48	cluster_2	48	8384.200	4123.400	9183.225
49	cluster_2	49	8375.500	4128.850	9199.500
50	cluster_2	50	8384.725	4110.600	9182.200
.					
.					
83	cluster_1	83	9545.625	4973.200	10272.150
84	cluster_1	84	9867.425	4928.050	10502.525
85	cluster_1	85	9136.275	4774.225	10015.600
.					
.					
158	cluster_0	158	9580.325	6241.725	9731.600
159	cluster_0	159	9425.975	6434.275	9819.050
160	cluster_0	160	10028.125	6209.325	9794.225

จากตารางที่ 4.1 ทำให้ได้กลุ่มของปริมาณความเข้มข้นของก๊าซแอลกอฮอล์ ไฮโดรเจน และคาร์บอน ซึ่งก๊าซนี้มีผลและความสัมพันธ์ต่อระดับการสุขของทุเรียน จากข้อมูลทำให้ได้กลุ่มข้อมูลจำนวน 4 กลุ่มโดยผู้วิจัยจะนำกลุ่มข้อมูลความเข้มข้นไปวิเคราะห์ต่อ โดยจะให้ระบบจัดกลุ่มเข้ากับลักษณะการสุขของทุเรียน



## 4.2 ผลการกำหนดกลุ่มของระดับการสุกของทุเรียน

เมื่อผู้วิจัยนำข้อมูลค่าของกลุ่มความเข้มข้นมาทดสอบด้วยวิธีการจัดกลุ่มและให้ระบบจัดกลุ่มเข้ากับตารางชื่อเรื่อง (Label) โดยใช้ข้อมูลที่ผ่านการคัดกรองหรือผ่านการหาค่าเฉลี่ยเคลื่อนที่ตามรูปที่ 4.4 นั้นผู้วิจัยได้ทำการจัดกลุ่มโดยใช้ K-Mean เพื่อกำหนดกลุ่มของระดับการสุก ซึ่งผู้วิจัยพบว่าสามารถใช้ในการจัดกลุ่มได้ดีถ้าเมื่อทราบจำนวนกลุ่มที่แน่นอน โดยใช้เทคนิคการกำหนดศูนย์กลางของแต่ละกลุ่มข้อมูล และทำการหารระยะห่างระหว่างข้อมูลแต่ละชุดข้อมูลกับจุดศูนย์กลาง (Mean) ของแต่ละกลุ่มข้อมูล โดยกำหนดให้ข้อมูลอยู่ในกลุ่มที่ใกล้ที่สุด จากนั้นคำนวณหาจุดศูนย์กลางของแต่ละกลุ่มข้อมูลใหม่ และทำซ้ำจนข้อมูลที่อยู่ในกลุ่มข้อมูลเดิมไม่มีการเปลี่ยนแปลง ในการทดลองนี้ผู้วิจัยจะไม่กำหนดค่าตารางชื่อเรื่องเพื่อให้อัลกอริทึมทำการจัดกลุ่มของข้อมูลก๊าซทั้ง 3 ชนิด เข้าตามกลุ่มระดับอาการสุกของทุเรียน โดยเบื้องต้นผู้วิจัยตั้งใจกำหนดกลุ่มของระดับการสุกเป็น 4 ระดับ คือ ดิบ ห้าม สุก และงอม ตามลักษณะการสุกของทุเรียนหรือกล้วย ที่คนไทยส่วนใหญ่เข้าใจกัน หรือตามที่คนไทยนิยมนำมาบริโภค แต่อย่างไรก็ตามผู้วิจัยจะทำการทดลองกำหนดค่า K ตั้งแต่ 2 ถึง 6 เพื่อดูผลการจัดกลุ่มด้วย K-Mean ว่าจะมีลักษณะเป็นอย่างไร ซึ่งจากการทดลองผู้วิจัยนำข้อมูลไปทดสอบระดับกลุ่มความเข้มข้นของปริมาณก๊าซต่าง ๆ โดยใช้พารามิเตอร์การจัดกลุ่มแบบตามลำดับชั้นหรือที่เรียกว่าการเชื่อมต่อแบบกลุ่มของข้อมูลตามการเชื่อมต่อ เป็นวิธีการวิเคราะห์กลุ่มที่พยายามสร้างลำดับชั้นของกลุ่ม การจัดกลุ่มตามลำดับชั้นจะขึ้นอยู่กับแนวคิดหลักของวัตถุที่เกี่ยวข้องกับวัตถุใกล้เคียงมากกว่าวัตถุที่อยู่ไกลออกไป ด้วยเหตุนี้อัลกอริทึมนี้จึงเชื่อมต่อวัตถุ (Objects) เพื่อสร้างกลุ่มตามระยะทางของกลุ่ม ซึ่งสามารถอธิบายได้ว่าส่วนใหญ่โดยระยะทางสูงสุดจะมีความสำคัญในการเชื่อมต่อส่วนต่าง ๆ ของกลุ่มข้อมูล ในระยะทางที่แตกต่างกันจะมีการสร้างกลุ่มข้อมูลที่แตกต่างกันซึ่งสามารถแสดงได้จากการใช้วิธีวัดความแตกต่างของต้นไม้ (Dendrogram) ซึ่งจะอธิบายว่า Hierarchical Clustering ที่มีคือกลุ่มชนิดใด การทำงานของอัลกอริทึมนี้จะไม่แบ่งพาร์ติชันของชุดข้อมูลเป็นลักษณะเดียว แต่ให้ลำดับชั้นที่กว้างเพื่อให้จัดกลุ่มข้อมูลเข้าด้วยกันในระยะทางที่กำหนด โดยที่แกน y จะหมายถึงระยะทางที่กระจุกของกลุ่มรวมกันในขณะที่วัตถุถูกวางไว้ตามแกน x เพื่อให้กลุ่มไม่สามารถผสมได้ ซึ่งจากการทดลองมีผลแสดงกลุ่มข้อมูลที่สัมพันธ์กับชนิดลักษณะการสุกของทุเรียน ดังตารางที่ 4.2

ตารางที่ 4.2 แสดงความสัมพันธ์ระหว่างกลุ่มความเข้มข้นปริมาณก๊าซกับระดับการสุกของทุเรียน

ID	ชนิด	กลุ่ม	ลูกที่	แอลกอฮอล์	ไฮโดรเจน	คาร์บอน
1	ดิบ	cluster_3	1	11538.450	5709.650	11828.975
2	ดิบ	cluster_3	2	10756.075	5456.300	11199.325
3	ดิบ	cluster_3	3	11116.700	6101.350	10519.950
.						
.						
48	ห้าม	cluster_2	48	8384.200	4123.400	9183.225
49	ห้าม	cluster_2	49	8375.500	4128.850	9199.500
50	ห้าม	cluster_2	50	8384.725	4110.600	9182.200
.						
.						
83	สุก	cluster_1	83	9545.625	4973.200	10272.150
84	สุก	cluster_1	84	9867.425	4928.050	10502.525
85	สุก	cluster_1	85	9136.275	4774.225	10015.600
.						
.						
158	งอม	cluster_0	158	9580.325	6241.725	9731.600
159	งอม	cluster_0	159	9425.975	6434.275	9819.050
160	งอม	cluster_0	160	10028.125	6209.325	9794.225

จากตารางที่ 4.2 สามารถบอกได้ถึงการจัดกลุ่มความสัมพันธ์ระหว่างกลุ่มของความเข้มข้นปริมาณก๊าซกับกลุ่มของระดับการสุกของทุเรียน โดยแบ่งได้เป็นทั้งหมด 4 กลุ่ม แต่อย่างไรก็ตามเพื่อให้การทดลองนี้เกิดความน่าเชื่อถือและมีประสิทธิภาพมากขึ้นผู้วิจัยจะทำการทดลองกำหนดค่า K ตั้งแต่ 2 ถึง 6 เพื่อดูผลการจัดกลุ่มด้วย K-Mean ว่ามีลักษณะเป็นเช่นไร ซึ่งผลการทดลองเป็นไปตามตารางที่ 4.3

ตารางที่ 4.3 แสดงค่า K ในการจัดกลุ่มด้วย K-Mean

	k=2	k=3	k=4	k=5	k=6
cluster_0	ดิบ(40) สุก (1) งอม (39)	ดิบ(2) สุก (34)งอม (37)	ดิบ(2) งอม(38)	ดิบ(25) งอม (2)	ดิบ (13) งอม (8)
cluster_1	ห้าม(40) สุก (39)งอม (1)	ดิบ (38) งอม (3)	สุก (40)	สุก (40)	ดิบ (23) งอม (1)
cluster_2		ห้าม (40) สุก (6)	ห้าม (40)	งอม (26)	สุก (38)
cluster_3			ดิบ (38) งอม (2)	ดิบ(15) งอม (12)	ดิบ (4) สุก (1) งอม (19)
cluster_4				ห้าม (40)	ห้าม (40)
cluster_5					สุก (1) งอม (12)

จากตารางที่ 4.3 สามารถบอกได้ว่าการหากกำหนดค่า  $K = 2$  จะสามารถแบ่งได้เป็นกลุ่มที่หนึ่งคือผลดิบ(40) สุก (1) และงอม (39) กลุ่มที่สองคือผลห้าม(40) สุก (39) และงอม (1) ซึ่งไม่เหมาะสมเนื่องจากยังมีการทำนายผลของงอมผิด 1 ผล และผลของสุก 1 ผล หากกำหนด  $K = 3$  จะสามารถแบ่งกลุ่มได้เป็นกลุ่มที่หนึ่งคือผลของดิบ(2) สุก (34) และงอม (37) กลุ่มที่สองดิบ (38) และงอม (3) และกลุ่มที่สามห้าม (40) และสุก (6) ซึ่งยังไม่เหมาะสมเนื่องจาก ยังมีการทำนายผลของดิบ 2 ผล ผลสุก 6 ผล และผลงอม 3 ผล หากกำหนด  $K = 4$  จะสามารถแบ่งได้เป็น กลุ่มที่หนึ่งผลดิบ(2) งอม(38) กลุ่มที่สอง สุก (40) กลุ่มที่สาม ห้าม (40) และกลุ่มที่สี่ดิบ (38) และงอม (2) ซึ่งทำนายผลดิบผิด 2 ผล และผลงอม 2 ผล หากกำหนด  $K = 5$  จะสามารถแบ่งกลุ่มได้เป็นกลุ่มที่หนึ่งดิบ(25) งอม (2) กลุ่มที่สองสุก (40) กลุ่มที่สามงอม (26) กลุ่มที่สี่ดิบ(15) งอม (12) และกลุ่มที่ห้า ห้าม (40) ซึ่งไม่เหมาะสมเนื่องจากทำนายผลดิบและผลงอมผิดเป็นจำนวนมาก และหากกำหนด  $K = 6$  จะสามารถแบ่งกลุ่มได้เป็นกลุ่มที่หนึ่งดิบ(13) และงอม (8) กลุ่มที่สอง ดิบ (23) และงอม (1) กลุ่มที่สาม สุก (38) กลุ่มที่สี่ ดิบ (4) สุก (1) และ งอม (19) กลุ่มที่ห้า ห้าม (40) และกลุ่มที่หก สุก (1) และงอม (12) ซึ่งไม่เหมาะสมเนื่องจากทำนายผลห้ามถูกเพียงชนิดเดียว ดังนั้นจึงสามารถสรุปได้ว่าการกำหนด  $K = 4$  จึงมีความเหมาะสมที่สุดเนื่องจากสามารถแบ่งกลุ่มได้เป็น ดิบ ห้าม สุก และงอม แต่ในรูปรวมโดยส่วนใหญ่จะทำนายทุเรียนที่มีลักษณะห้ามถูกทุกกรณี ผู้วิจัยยังค้นพบว่าสาเหตุที่ทำให้การทำนายผล  $K = 4$  ผิดเนื่องจากอัตราการหายไปของทุเรียนในลักษณะดิบและงอมใกล้เคียงกัน

### 4.3 ผลการจำแนกประเภทการสุกของทุเรียนจากการแพร่กระจายของก๊าซ

จากผลที่ได้จากตารางที่ 4.3 ในส่วนของการจำแนกประเภทการสุกของทุเรียนจากปริมาณก๊าซ ผู้วิจัยได้ใช้เทคนิคจำแนกประเภทหรือ Classification โดยเริ่มจากการสร้างแบบจำลองที่ใช้ในการจำแนกระดับการสุกของทุเรียนแต่ละระดับให้เกิดประสิทธิภาพและมีความถูกต้องมากที่สุดนั้น สิ่งที่สำคัญและจะต้องอาศัยองค์ประกอบหลักคือ ปริมาณก๊าซที่ได้รับจากเซ็นเซอร์ต่าง ๆ ดังนั้นการจำแนกข้อมูลนั้น ผู้วิจัยเลือกอัลกอริทึมที่เหมาะสมมาใช้งาน โดยทำการทดลองเปรียบเทียบอัลกอริทึมที่ใช้ในการวิจัย ได้แก่ เทคนิคของต้นไม้การตัดสินใจ (Decision Tree) เทคนิคของการหาเพื่อนบ้านที่ใกล้ที่สุด (K Nearest Neighbours (KNN)) เทคนิคของการสร้างเครือข่ายประสาทเทียม (Neural Network) เทคนิคของความน่าจะเป็น (Naive Bayes) และเทคนิคของการใช้สมการเส้นตรง (Support Vector Machine (SVM)) ซึ่งผู้วิจัยได้แบ่งการทดลองเป็น 2 ขั้นตอนดังนี้

1) ผู้วิจัยจะกำหนดให้ทุกอัลกอริทึมใช้การทดสอบแบบค่าเริ่มต้น (Default Parameter Set) โดยผู้วิจัยจะสร้างแบบจำลองขึ้นและกำหนดเป็นค่าเริ่มต้นเพื่อให้ระบบนำผลของค่าเริ่มต้นเหล่านี้ไปเก็บไว้ในตารางชื่อเรื่อง (Label) จากการประมวลผลของจำนวนข้อมูลทั้งหมดที่นำมาทดสอบ วิธีการที่ใช้สำหรับการสร้างค่าเริ่มต้นนี้สามารถเลือกมาจากการข้อมูลในพารามิเตอร์ที่เป็นตัวเลข ค่ามัธยฐานหรือค่าเฉลี่ย ที่สามารถระบุผ่านพารามิเตอร์คงที่เพื่อนำแบบจำลองที่ได้ไปเปรียบเทียบเพื่อหาแบบจำลองที่มีประสิทธิภาพที่สุด ซึ่งผลการวิจัยที่ได้แสดงดังตารางที่ 4.5

2) ผู้วิจัยได้ทดลองใช้การทดสอบแบบเพิ่มประสิทธิภาพ (Optimize Parameter Set) โดยผู้วิจัยจะการกำหนดค่าต่าง ๆ โดยกำหนดพารามิเตอร์ในส่วนการตั้งค่า และปรับค่าที่สมบูรณ์ที่สุด ผู้วิจัยจะกำหนดให้เป็นแบบซิงโครไนซ์ เพื่อให้ระบบประมวลผลกระบวนการย่อยและทำการวนซ้ำมากกว่าพารามิเตอร์เดียวเพื่อเพิ่มประสิทธิภาพและความแม่นยำ โดยมีการปรับค่าพารามิเตอร์ตามตารางที่ 4.4 และนำผลที่ได้ไปทำการทดลองอีกครั้งเพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพสูงสุด ซึ่งผลการวิจัยที่ได้แสดงดังตารางที่ 4.5

ตารางที่ 4.4 แสดงพารามิเตอร์และค่าพารามิเตอร์ทำการปรับค่า

Algorithm	Parameter	Optimize parameter
Decision Tree	Decision Tree.criterion	= gain_ratio
	Decision Tree.maximal_depth	= -1
	Decision Tree.apply_pruning	= true
	Decision Tree.confidence	= 1.0E-7 - 0.5
K Nearest Neighbors (KNN)	k-NN.k	= 1
	k-NN.weighted_vote	= true
	k-NN.numerical_measure	= EuclideanDistance
	k-NN.kernel_type	= dot
	k-NN.measure_types	= MixedMeasures
Neural Network	Neural Net.training_cycles	= 100
	Neural Net.learning_rate	= 0.3
	Neural Net.momentum	= 0.2
	Neural Net.error_epsilon	= 0
Naïve Bayes	Naive Bayes	= full
	(Kernel).estimation_mode	= heuristic
	Naive Bayes	= false
	(Kernel).bandwidth_selection	
	Naive Bayes.laplace_correction	
Support vector machine (SVM)	SVM.kernel_type	= linear
	SVM.gamma	= Infinity
	SVM.C	= Infinity
	SVM.svm_type	= C-SVC

จากตารางที่ 4.4 ผู้วิจัยได้กำหนดค่าพารามิเตอร์โดยการปรับค่าต่าง ๆ แบบการทดสอบเพิ่มประสิทธิภาพไปที่ละขั้นจนได้ผลการทดลองสามารถสรุปได้คือ

Decision Tree ผู้วิจัยทำการกำหนดค่าต่าง ๆ ในรูปแบบการปรับจูนค่าพารามิเตอร์ของ Decision Tree.criterion gain\_ratio, Information\_gain, Gini\_index และ Accuracy ค่าความเหมาะสม

และแม่นยำที่สุดคือ Gain\_ratio = 0.988, Information\_gain, Gini\_index ลดลงตามลำดับ Decision Tree.maximal\_depth ค่าความเหมาะสมเท่ากับ -1 และ Decision Tree.confidence ค่า grid/range อยู่ในช่วง 1.0E-7 - 0.5 ซึ่งเป็นช่วงระยะที่เหมาะสมที่สุด

K Nearest Neighbors (KNN) ในการกำหนดค่าต่าง ๆ ในการปรับจูนค่าพารามิเตอร์ของ ปรับค่า k-NN.k, k-NN.weighted\_vote, k-NN.numerical\_measure, k-NN.kernel\_type และ k-NN.measure\_types นั้น ผู้วิจัยค้นพบว่า การกำหนดค่า K = 1 จะมีผลกับการใช้งานร่วมกับ NN.numerical\_measure = MixedMeasures, k-NN.measure\_types = EuclideanDistance และ k-NN.kernel\_type = dot เมื่อทำงานร่วมกันแล้วมีความเหมาะสมและให้ค่าความแม่นยำถึง 0.981 ซึ่งเป็นค่าสูงสุดเมื่อเปรียบเทียบกับ การกำหนดค่าพารามิเตอร์ในรูปแบบอื่น

Neural Network ในการกำหนดค่าต่าง ๆ ในการปรับจูนค่าพารามิเตอร์ของ โดยปรับค่าให้ Neural Net.training\_cycles อยู่ในระหว่าง 1-100, Neural Net.learning\_rate อยู่ในช่วงระหว่าง 4.9E-324 - 1.0 เลือก linear, Neural Net.momentum ให้ไม่เกิน 1.0 และ Neural Net.error\_epsilon ไม่เกิน 100 นั้น เป็นค่าที่เหมาะสมที่สุดทั้งในเรื่องของความแม่นยำและความเร็ว

Naïve Bayes ในการปรับ Configure Operator โดยปรับค่าการระบุความหนาแน่นของเคอร์เนลในสองลักษณะคือ Full, Greedy นั้นในส่วนของ Naïve Bayes (Kernel).estimation\_mode ผู้วิจัยเลือกใช้ Full เพราะจะสอดคล้องต่อการเลือกพารามิเตอร์ในขั้นถัดไป คือ Bandwidth\_selection ผู้วิจัยจะนำแบนด์วิธชนิด Heuristic ซึ่งมีอิทธิพลเป็นอย่างมากต่อการประมาณขนาดของการเกิดของข้อมูลที่ จะเกิดขึ้นเพราะหลักการวิเคราะห์ข้อมูลของ Naïve Bayes คือการทำนายความน่าจะเป็นของการเกิด ข้อมูล เพราะหากข้อมูลเล็กเกินหรือใหญ่เกินก็ไม่เป็นประโยชน์ ในส่วนของการปรับจูนนี้ สามารถให้ ค่าคะแนนความถูกต้องได้สูงกว่าเมื่อเทียบกับการปรับแบบ Greedy และเลือกแบนด์วิธเป็น Fix

Support vector machine (SVM) ในขั้นตอนการปรับ Configure Operator ผู้วิจัยกำหนด SVM.kernel\_type โดยเลือกชนิด Linear, Poly และ rbf มาทดลองทีละค่าซึ่งจากตารางผลการ ทดลองค่าความเหมาะสมที่สุดคือ Linear โดยผลการตรวจสอบเคอร์เนล SVM.gamma นั้น เท่ากับ 0 ประกอบกับค่าการตรวจสอบจาก SVM.C มีค่าเท่ากับ SVM.C=0 นั้นหมายถึงไม่มี ข้อผิดพลาดหรืออีกนัยยะหนึ่งผลการทดสอบนี้มีความน่าเชื่อถือนั่นเอง ซึ่งเมื่อผู้วิจัยได้ทำการ ทดลองทั้ง 2 ขั้นตอนแล้วได้ผลการเปรียบเทียบประสิทธิภาพของแต่ละขั้นตอน โดยพิจารณาจาก ความถูกต้องตามตารางที่ 4.5



ตารางที่ 4.5 แสดงการเปรียบเทียบประสิทธิภาพของอัลกอริทึม

Algorithm		ดิบ	ห้าม	สุก	งอม	accuracy
Decision Tree	Default	95%	100%	90%	80%	91.25%
	Parameter Set	ดิบ(38) งอม(2)	ห้าม(40)	สุก(36) งอม(4)	ดิบ(4) ห้าม(4) งอม(32)	
	Optimize	100%	100%	100%	97.50%	99.38%
	Parameter Set	ดิบ(40)	ห้าม(40)	สุก(40)	งอม(39) งอม(1)	
K Nearest Neighbours (KNN)	Default	95%	100%	100%	92.50%	96.88%
	Parameter Set	ดิบ(38) งอม(2)	ห้าม(40)	สุก(40)	งอม(37) ดิบ(3)	
	Optimize	100%	100%	100%	97.50%	99.38%
	Parameter Set	ดิบ(38)	ห้าม(40)	สุก(40)	งอม(39) ดิบ(1)	
Neural Network	Default	97.50%	100%	100%	90%	96.88%
	Parameter Set	ดิบ(39) งอม(1)	ห้าม(40)	สุก(40)	งอม(36) ดิบ(4)	
	Optimize	100%	100%	100%	97.50%	99.97%
	Parameter Set	ดิบ(40)	ห้าม(40)	สุก(40)	ดิบ(1) งอม(39)	

ตารางที่ 4.5 แสดงการเปรียบเทียบประสิทธิภาพของอัลกอริทึม (ต่อ)

Algorithm		ดิบ	ห้าม	สุก	งอม	accuracy
Naive Bayes	Default	80%	92.50%	85%	90%	
	Parameter Set	ดิบ(32) งอม(8)	ห้าม(37) งอม(2)	สุก(34) งอม(1)	งอม(6) ดิบ(4) งอม(36)	86.88%
	Optimize Parameter Set	90% ดิบ(36) งอม(4)	100% ห้าม(40)	97.50% สุก(39) งอม(1)	95% ดิบ(2) งอม(38)	95.62%
	Default	92.50%	100%	97.50%	92.50%	
	Parameter Set	ดิบ(37) งอม(3)	ห้าม(40)	สุก(39) งอม(1)	ดิบ(2) งอม(37)	95.62%
Support vector machine (SVM)	Optimize Parameter Set	100% ดิบ(40)	100% ห้าม(40)	100% สุก(40)	100% งอม(40)	100%

จากตารางที่ 4.5 ผู้วิจัยจะขออธิบายขยายผลการทดลองการเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมดังนี้

1) Decision Tree เมื่อกำหนดค่าเป็นค่าเริ่มต้นพบว่าสามารถจำแนกประเภทระดับการสุกของทุเรียนในลักษณะห้ามได้ถูกต้องและแม่นยำสุด ในลำดับลงลงมาคือดิบ โดยมีค่าความถูกต้องในรูปรวม 91.25% แต่เมื่อผู้วิจัยทำการปรับค่าเป็นแบบการเพิ่มประสิทธิภาพพบว่าอัลกอริทึมนี้สามารถจำแนกประเภทระดับการสุกของทุเรียนได้ถูกต้องคือลักษณะดิบ ห้าม และสุก ซึ่งมีค่าความถูกต้องในภาพรวม 99.38%

2) K Nearest Neighbours (KNN) เมื่อกำหนดค่าเป็นค่าเริ่มต้นพบว่าสามารถจำแนกประเภทระดับการสุกของทุเรียนในลักษณะห้ามและสุกได้ถูกต้อง โดยมีค่าความถูกต้องในรูปรวม 96.88% แต่เมื่อผู้วิจัยทำการปรับค่าเป็นแบบการเพิ่มประสิทธิภาพพบว่าอัลกอริทึมนี้สามารถจำแนกประเภทระดับการสุกของทุเรียนได้ถูกต้องคือลักษณะดิบ ห้าม และสุก ซึ่งมีค่าความถูกต้องในภาพรวม 99.38%

3) Neural Network เมื่อกำหนดค่าเป็นค่าเริ่มต้นพบว่าสามารถจำแนกประเภทระดับการสุกของทุเรียนในลักษณะห้ามและสุกได้ถูกต้อง โดยมีค่าความถูกต้องในรูปรวม 96.88% เนื่องจากจำแนกประเภทลักษณะดิบเป็นงอมจำนวน 1 ผล และลักษณะงอมเป็นดิบจำนวน 4 ผล แต่เมื่อผู้วิจัยทำการปรับค่าเป็นแบบการเพิ่มประสิทธิภาพพบว่าอัลกอริทึมนี้สามารถจำแนกประเภทระดับการสุกของทุเรียนได้ถูกต้องคือลักษณะดิบ ห้าม และสุก ซึ่งมีค่าความถูกต้องในภาพรวม 99.97% และจำแนกประเภทลักษณะทุเรียนงอมเป็นดิบเพียง 1 ผล

4) Naive Bayes เมื่อกำหนดค่าเป็นค่าเริ่มต้นพบว่ามีค่าความถูกต้องต่ำและความสามารถในการจำแนกประเภทระดับการสุกของทุเรียนเป็นแบบคละกระจายในรูปรวมเมื่อเทียบกับอัลกอริทึมชนิดอื่น โดยมีค่าความถูกต้องในรูปรวม 86.88% แต่เมื่อผู้วิจัยทำการปรับค่าเป็นแบบการเพิ่มประสิทธิภาพพบว่าอัลกอริทึมนี้สามารถจำแนกประเภทระดับการสุกของทุเรียนได้ถูกต้องคือลักษณะห้าม ซึ่งมีค่าความถูกต้องในภาพรวม 95.62%

5) Support Vector Machine เมื่อกำหนดค่าเป็นค่าเริ่มต้นพบว่าสามารถจำแนกประเภทระดับการสุกของทุเรียนในลักษณะห้ามได้ถูกต้องในลำดับลงมาคือสุก โดยมีค่าความถูกต้องในรูปรวม 95.63% แต่เมื่อผู้วิจัยทำการปรับค่าเป็นแบบการเพิ่มประสิทธิภาพพบว่าอัลกอริทึมนี้สามารถจำแนกประเภทระดับการสุกของทุเรียนได้ถูกต้องคือลักษณะดิบ ห้าม สุก และงอม ซึ่งมีค่าความถูกต้องในภาพรวม 100%

จากการทดลองเปรียบเทียบอัลกอริทึมใน 2 ลักษณะคือ การทดสอบแบบค่าเริ่มต้น และการทดสอบแบบเพิ่มประสิทธิภาพ นั้น เมื่อผู้วิจัยกำหนดให้แต่ละอัลกอริทึมกำหนดพารามิเตอร์แบบค่าเริ่มต้นพบว่า Neural Network, K Nearest Neighbours (KNN) และ Support Vector Machine ทำงานได้ดีโดยมีความแม่นยำในการจำแนกข้อมูลสูงที่สุดและรองลงมาตามลำดับ ส่วน Decision Tree และ Naive Bayes นั้นมีความแม่นยำในการจำแนกข้อมูลต่ำกว่ากลุ่มแรกลงมาตามลำดับ แต่หลังจากที่ผู้วิจัยได้ทำการทดลองปรับพารามิเตอร์แบบเพิ่มประสิทธิภาพซึ่งจากการทดลองดังกล่าว พบว่า Support Vector Machine ทำงานได้ดีโดยมีความถูกต้องในการจำแนกข้อมูลสูงที่สุดที่ 100% และ Neural Network ทำงานได้ดีรองลงมาโดยมีความถูกต้อง 99.97 ส่วน K Nearest Neighbours (KNN), Decision Tree และ Naive Bayes นั้นมีความถูกต้องในการจำแนก

ข้อมูลตำรองลงมาตามลำดับแต่เมื่อเทียบกับการปรับค่าเริ่มต้นในขั้นต้นแล้วอัลกอริทึมทุกตัว เมื่อปรับพารามิเตอร์แล้วมี ค่าความแม่นยำในการจำแนกข้อมูลสูงกว่า 95 เปอร์เซ็นต์



## บทที่ 5

### สรุปผลและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาเครื่องมือตรวจจับก๊าซเอทีลินจากทุเรียน เพื่อทดลองแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทีลินโดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล และเพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลิน โดยใช้เทคนิคการจำแนกข้อมูลนั้น ผู้วิจัยจะขอสรุปผลและนำเสนอข้อเสนอแนะที่ได้จากการวิจัย ตามลำดับดังนี้

#### 5.1 สรุปผลการวิจัยและอภิปรายผล

จากปัญหาการจัดเก็บทุเรียนของเกษตรกรอันนำไปถึงการส่งผลถึงการขนส่งตลอดจนการซื้อขายหรือจัดจำหน่ายทุเรียนนั้นมาจากคัดแยกระดับการสุกของทุเรียนในปัจจุบันนี้ไม่เป็นไปตามความต้องการของผู้รับหรือผู้บริโภค ซึ่งอาจสามารถสังเกตได้จากร้านค้าตัวแทนจำหน่ายทุเรียนที่ส่วนใหญ่จะไม่สามารถจำหน่ายทุเรียนได้ตรงตามลักษณะความสุกที่ลูกค้าต้องการ โดยส่วนใหญ่ผู้ซื้อจะมีพฤติกรรมการเลือกบริโภคไม่เหมือนกันเช่น กรอบ กรอบนอกนุ่มใน กรอบหวาน หวานนุ่ม และสุดท้ายคือ หวานมาก ปัญหานี้อาจเกิดมาจากผู้จำหน่ายวางแผนในการจัดเก็บและดำเนินการขนส่งได้ ไม่สมบูรณ์เพราะไม่สามารถคาดคะเนได้ ผู้วิจัยจึงพัฒนาอุปกรณ์เพื่อตรวจวัดระดับการสุกของทุเรียน ซึ่งผู้วิจัยได้ดำเนินการและมีรายละเอียดโดยสรุปดังนี้

##### 5.1.1 การพัฒนาเครื่องมือตรวจจับก๊าซเอทีลินจากทุเรียน

จากปัญหาดังกล่าวผู้วิจัยได้ศึกษาเกี่ยวกับก๊าซสำคัญที่เกี่ยวข้องกับกระบวนการสุกของผลไม้ องค์ประกอบของก๊าซต่าง ๆ ที่เกี่ยวข้องกับการวิจัย อุปกรณ์ เช่น เซอร์ ซอฟต์แวร์ที่เกี่ยวข้องและรวมไปถึงเทคนิคต่าง ๆ ในการทำเหมืองข้อมูล ผู้วิจัยค้นพบว่าก๊าซเอทีลินมีความสอดคล้องกับระดับการสุกของทุเรียน ผู้วิจัยจึงทำการพัฒนาอุปกรณ์ตรวจจับการแพร่กระจายของก๊าซเอทีลินขึ้น โดยทำการติดตั้งแก๊สเซ็นเซอร์ชนิด MQ3, MQ6 และ MQ8 และทำการเขียน โปรแกรมเพื่อสั่งงานให้อุปกรณ์ต่าง ๆ ทำงานและนำอุปกรณ์ที่พัฒนาได้ไปวัดปริมาณก๊าซของทุเรียนเป้าหมายตัวอย่างที่มีลักษณะคละกันตามสภาพ โดยผู้วิจัยได้เชิญผู้เชี่ยวชาญร่วมกันตรวจสอบนั้น ซึ่งจากการวิจัยดังกล่าว

ผู้วิจัยใช้กระบวนการของการทำเหมืองข้อมูลเข้ามาช่วยและได้ผลการทดลอง โดยผู้วิจัยจะอภิปรายต่อไป

### 5.1.2 การทดลองแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทีลิน

โดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล โดยในส่วนของ การจัดแบ่งกลุ่มของข้อมูลซึ่งได้เลือกใช้อัลกอริทึม K-Mean คือ หากกำหนด  $K=2$  จะแบ่งกลุ่มได้ คือ 1) ดิบหรืองอม และ 2) ห่ามหรือสุก ซึ่งไม่เหมาะสมเนื่องจาก ดิบกับงอมนั้นในการเลือกซื้อนั้นถือว่ามีความแตกต่างกันอย่างสิ้นเชิง เมื่อกำหนด  $K=3$  จะแบ่งกลุ่มได้ คือ 1) ดิบ 2) ห่าม และ 3) สุกหรืองอม ซึ่งมีการจัดกลุ่มผิด จาก ดิบเป็นสุกหรืองอม จำนวน 2 ลูก จาก สุกเป็นห่าม จำนวน 6 ลูก และจากสุกหรืองอมเป็น ดิบ จำนวน 3 ลูก เมื่อกำหนด  $K=4$  จะแบ่งกลุ่มได้ คือ 1) ดิบ 2) ห่าม 3) สุก และ 4) งอม ซึ่งมีการจัดกลุ่มผิด จาก ดิบเป็นงอม จำนวน 2 ลูก และจาก งอมเป็นดิบ 2 ลูก เมื่อกำหนด  $K=5$  และ  $K=6$  จะเริ่มมีกลุ่มซ้ำ คือ ดิบหรืองอม ซึ่งมีความแตกต่างกันดังที่กล่าวไว้ข้างต้น และมีการจัดกลุ่มผิดจำนวนมาก การกำหนดให้  $K=4$  จึงแบ่งกลุ่มระดับการสุกจากค่าก๊าซทั้ง 3 ชนิด ได้เหมาะสมที่สุด ซึ่งผลการวิจัยพบว่าสามารถแบ่งการสุกของทุเรียนจากปริมาณก๊าซที่ได้จากเซ็นเซอร์ต่าง ๆ ได้เป็น 4 กลุ่ม คือ ดิบ ห่าม สุก และงอม และเป็นการจัดแบ่งกลุ่มที่เหมาะสมกับการเลือกซื้อทุเรียนตามที่ผู้วิจัยคาดหวังเอาไว้ นอกจากนี้ผู้วิจัยยังค้นพบสาเหตุที่ทำให้แบบจำลองทำการแบ่งกลุ่มผิดพลาดคือ ทุเรียนจะมีระดับการหายใจสูง เมื่ออยู่ในอาการดิบ และลดลงมาเมื่ออยู่ในอาการห่ามและลดลงอีกเมื่ออยู่ในอาการสุก และจะกลับมาหายใจสูงอีกครั้งเมื่ออยู่ในอาการงอม ด้วยเหตุผลนี้เองจึงทำให้อัลกอริทึมมีการจำแนกผิดพลาดเพราะระดับอาการดิบและอาการงอม มีลักษณะการหายใจและปริมาณก๊าซที่ใกล้เคียงกัน ซึ่งหากใช้กระบวนการทางวิทยาศาสตร์พบว่าปริมาณก๊าซแตกต่างกันอย่างสิ้นเชิง

### 5.1.3 การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลิน

โดยใช้เทคนิคการจำแนกข้อมูล ในส่วนของการจำแนกระดับการสุกของทุเรียนจากปริมาณก๊าซเอทีลิน ผู้วิจัยได้ใช้เทคนิคการจำแนกข้อมูล โดยเริ่มจากการสร้างโมเดลที่ใช้ในการประเมินระดับการสุก โดยผู้วิจัยเลือกอัลกอริทึมที่นำมาใช้ทั้งหมด 5 อัลกอริทึม ได้แก่ Decision Tree K Nearest Neighbors (KNN) Neural Network Naive Bayes และ Support Vector Machine (SVM) ซึ่งผลการวิจัยในกรณีที่ใช้พารามิเตอร์เริ่มต้นของ Rapid Miner พบว่า Neural Network, Support Vector Machine และ K Nearest Neighbours (KNN) ทำงานได้ดีโดยมีความแม่นยำในการจำแนกข้อมูลสูงที่สุดและรองลงมาตามลำดับ ในส่วน Naive Bayes และ Decision Tree นั้นมีค่าความ



แม่นยำในการจำแนกข้อมูลต่ำกว่าตามลำดับ หลังจากนั้นผู้วิจัยได้ทำการทดลองอีกครั้งแบบปรับจูนพารามิเตอร์เอง โดยใช้การทดสอบแบบเพิ่มประสิทธิภาพใน Rapid Miner ซึ่งจากการทดลองดังกล่าวพบว่าโมเดลที่มีค่าความแม่นยำในการจำแนกข้อมูลที่สูงที่สุดคือ Support Vector Machines (SVM) ซึ่งสอดคล้องกับงานวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคมรื้อน (ภรณ์ยา อามฤครัตน์ และ พยุง มีสัจ, 2553) ในงานวิจัยการพัฒนาประสิทธิภาพการจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ (นิเวศ จิระวิจิตชัย, ปริญญา สงวนศักดิ์ และ พยุง มีสัจ, 2554) และระบบแจ้งเตือนโซเชียลมีเดียไทยสำหรับธุรกิจด้วยซอฟต์แวร์เวกเตอร์แมชชีน (สุพัตรา วิริยะวิสุทธิสกุล, ปริญญา สงวนศักดิ์, พิสิษฐ์ ชาญเกียรติกิจ และ ชูชาติ หลุไชยะศักดิ์, 2559) เพราะเทคนิค SVM นั้นมีหลักการของโมเดลสมการเส้นตรงที่แบ่งข้อมูลเป็นสอง Class จากนั้นจะใช้ Kernel Function เพื่อทำการแปลงข้อมูลเพื่อให้โมเดลสมการเส้นตรงไปอยู่ในมิติที่สูงขึ้น (Dimension) ซึ่งสมการเชิงเส้นนี้มีความสามารถในการแยกแยะข้อมูลทำให้เกิดการเพิ่มระยะของเส้นแบ่งเขตที่มีความกว้างเท่ากับ  $2 / \|w\|^2$  ซึ่งสอดคล้องกับงานวิจัยการคัดแยกประเภทของมะเร็งเม็ดเลือดขาวโดยใช้วิธีการจัดอันดับร่วมกับเทคนิคซอฟต์แวร์เวกเตอร์แมชชีน (ภัทราวุฒิ แสงศิริ, วิเชียร ณ ศจีมาจ และ พยุง มีสัจ, 2553) เพราะฉะนั้นสำหรับงานวิจัยนี้ SVM จึงเหมาะที่จะนำมาใช้ในการประเมินระดับการสุกมากที่สุด ในแง่ของค่าความแม่นยำในการจำแนกข้อมูลและที่จะนำมาใช้ในการประเมินระดับการสุกมากที่สุด แต่อย่างไรก็ตามในรูปรวมโดยส่วนใหญ่การทดลองพบว่ามีความถูกต้องมากกว่า 95% ในทุกอัลกอริทึมยกเว้น Naive Bayes ที่ผลการทดลองอยู่ในระดับต่ำเมื่อเทียบกับอัลกอริทึมแบบอื่น โดยมีค่าความถูกต้องแบบเริ่มต้น 86.88% โดยไม่สามารถจำแนกประเภทของทุเรียนแต่ละชนิดได้ถูกต้อง และในส่วนของแบบเพิ่มประสิทธิภาพมีค่าความถูกต้อง 95.62% และสามารถจำแนกประเภทของทุเรียนชนิดห้ามได้ถูกต้องเพียงอย่างเดียวจากข้อมูลทั้งหมด

## 5.2 ข้อเสนอแนะ

ในการวิจัยและพัฒนาอุปกรณ์ตามข้างต้นนั้น ตามรูปการณ์วัสดุอุปกรณ์สามารถจัดหาได้ในราคาถูกและนำมาพัฒนาทดแทนกับอุปกรณ์หลักที่มีราคาแพงได้ รวมไปถึงความง่ายในการจัดหานั้นเป็นไปตามเป้าประสงค์ของงานวิจัย และอุปกรณ์สามารถใช้วัดระดับการสุกของทุเรียนได้ แต่อย่างไรก็ตามปัญหา อุปสรรค ข้อขัดข้อง ในการวิจัยนี้พบว่ายังมีความล่าช้าในการตรวจวัดความเข้มข้นของก๊าซดังกล่าว ในอนาคตผู้วิจัยจะทำการพัฒนาอุปกรณ์สำหรับตรวจวัดความสุกของทุเรียน โดยใช้ไอคิวโนและการเลือกอัลกอริทึมจึงอาจต้องมีการพิจารณาเพิ่มเติมนอกจากประเด็นของความถูกต้อง เช่น ความยากง่ายในการนำโมเดลไปพัฒนา ความเร็วในการประมวลผลต่อไป

## บรรณานุกรม

- กิตติ วงศ์พิเชษฐ. (2559). *วิทยาเอ็มบริโอของพืชดอก*. อุบลราชธานี: คณะเกษตรศาสตร์ มหาวิทยาลัยอุบลราชธานี.
- ดวงแก้ว สวามิภักดิ์. (2531). *การเขียนโปรแกรมภาษา C*. กรุงเทพฯ: สถาบันวิจัยและให้คำปรึกษา แห่ง มหาวิทยาลัยธรรมศาสตร์.
- เดช ธรรมศิริ และ พยุง มีสัจ. (2554). การจำแนกข้อมูลด้วยวิธีแบบร่วมกันตัดสินใจจากพื้นฐานของ เทคนิคต้นไม้ตัดสินใจ. *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 21, 293 - 330.
- นิเวศ จิระวิจิตรชัย, ปริญญา สงวนสัจย์, และ พยุง มีสัจ. (2554). การพัฒนาประสิทธิภาพการจด หมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ. *วารสารพัฒนบริหารศาสตร์*, 51(3), 188-205.
- ภรณ์ยา อามฤรัตน์ และ พยุง มีสัจ. (2553). การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูล ของโรค ลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม. *National Conference on Computing and Information Technology*, 12(1), 54 - 63.
- ภัทราวุฒิ แสงศิริ, วิเชียร ณศจีมาจ และ พยุง มีสัจ. (2553). การคัดแยกประเภทของมะเร็งเม็ดเลือด ขาวโดยใช้วิธีการจัดอันดับร่วมกับเทคนิคซอฟต์แวร์เวกเตอร์แมชชีน. *วารสาร มช.(บศ.)* 10(2), 10 - 17.
- มหาวิทยาลัยราชภัฏรำไพพรรณีและเครือข่ายนักวิจัยสาขาพืชในจังหวัดจันทบุรี. (2556). *ทุเรียน ผลไม้วัฒนธรรมอาเซียน*. กรุงเทพฯ : คณะเทคโนโลยีการเกษตร.
- เมธี ขำพวง, พันทิวา คำแก้ว, สมทรง แจ็งอักษร และจิตวัฒนา บุญเลิศ. (2561). *อุปกรณ์วัดความสุข ของทุเรียน*. สืบค้นจาก <https://workpointnews.com/2018/05/15/%E0%B8%AD%E0%B8%B8%E0B8%9B%E0%8%81%E0%B8%A3%E0%B8%93%E0%B9%8C%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B8%84%E0%B8%A7%E0%B8%B2%E0%B8%A1%E0%B8%AA%E0%B8%B8%E0%B8%81%E0%B8%82%E0%B8%AD%E0%B8%87%E0%B8%97%E0%B8%B8/>
- วีระยุทธ พิมพาภรณ์ และ พยุง มีสัจ. (2557). การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดย วิธีการเลือกลักษณะสำคัญแบบพลวัตเพื่อเพิ่มประสิทธิภาพของอัลกอริธึมการจัดกลุ่มบน ปริภูมิย่อย. *วารสารเทคโนโลยีสารสนเทศ*, 10(2), 43-51.
- สายชล สตินสมบูรณ์ทอง. (2558). *การทำเหมืองข้อมูล*. กรุงเทพฯ: จามจุรีโปรดักท์.

## บรรณานุกรม (ต่อ)

- สุรวัชร ศรีเปารยะ และ สายชล สตินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการ  
จำแนกกลุ่มการเป็นโรคไตเรื้อรัง. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(5), 840-853.
- สุพัตรา วิริยะวิสุทธิสกุล, ปริญญา สงวนศักดิ์, พิสิษฐ์ ชาญเกียรติก้อง, และชูชาติ หลงไชยะศักดิ์.  
(2559). ระบบแจ้งเตือนโซเซียลมีเดียไทยสำหรับธุรกิจด้วยซอฟต์แวร์เวกเตอร์แมชชีน.  
*สารสารปัญญาทัศน์*, 8(ฉบับพิเศษ) (สิงหาคม 2559), 223 - 233.
- สำนักวิจัยเศรษฐกิจการเกษตร สำนักงานเศรษฐกิจการเกษตร. (2559). *สถานการณ์สินค้าเกษตรที่สำคัญและแนวโน้มปี 2560*. สืบค้นจาก <http://www.oae.go.th/assets/portals/1/files/journal/2561/indicator2560.pdf>
- หิรัญ หิรัญประดิษฐ์. (2554). *ประวัติของทุเรียนในประเทศไทยในสารานุกรมไทยสำหรับเยาวชน โดยพระราชประสงค์ในพระบาทสมเด็จพระเจ้าอยู่หัว เล่ม 28 เรื่องที่ 4, 99-125*. กรุงเทพฯ: โครงการสารานุกรมไทยสำหรับเยาวชน.
- อชิพงษ์ สุริยา และ มงคล ปุษยตานนท์. (2549). *ระบบประเมินความสุกของผลไม้โดยใช้จุ่มกอิเล็คทรอนิกส์*. กรุงเทพฯ: ภาควิชาวิศวกรรมไฟฟ้าและอิเล็คทรอนิกส์ มหาวิทยาลัยอุบลราชธานี.
- เอกสิทธิ์ คุณานันทกุล. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิค คาด้า ไม่นิ่ง เบื้องต้น*. กรุงเทพฯ: สำนักพิมพ์ คาด้าคิวบ์.
- Arduino. (2018 a). *Arduino Products*. Retrieved from <https://www.arduino.cc/en/Main/Products>
- Arduino. (2018 b). *Arduino playground MQ gas sensor*. Retrieved from <http://playground.arduino.cc/Main/MQGasSensor>.
- Adams, D. & Yang, S. F. (1979). Ethylene Biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic Acid as an Intermediate in the Conversion of Methionine to Ethylene. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1), 170-174.
- Aha, D. W. (1997). *Lazy learning*. Kluwer Academic. USA: Norwell, MA.
- Avila, P. A., Toledo, F., Park, Y. S., Jung, S. T., Kang, S. G., Heo, B. G. & Gorinstein, S. (2008). Antioxidant properties of durian fruit as influenced by ripening. *LWT - Food Science and Technology*, 41(10), 2118-2125.

### บรรณานุกรม (ต่อ)

- Bhargava, N., Sharma, G., Bhargava, R. & Mathuria, M. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research*, 3(6) (June 2013), 1114 - 1119.
- Blanke, M. M. (2014). Reducing ethylene levels along the food supply chain: a key to reducing food waste. *Journal of the science of food and agriculture*, 94(12), 2357–2361.
- Eberhardt, A., Schmitt, K. & Wöllenstein, J. (2017). Rotating interference filter spectrometer for the detection of ethylene in the ripening process of climacteric fruit. *19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, Kaohsiung, Taiwan
- Hashim, N. (2008). Study on Combustion Characteristics of Different Compositions of LPG, *UMP*, 1 - 56.
- Hsu, C. W., Chang, C. C. & Lin, C. J. (2003). A practical guide to support vector classification *National Taiwan University Taipei Taiwan*, 106(1), 1-15.
- Kaushlendra, T., Saurabh, P., Murtaza, M., & Tanushri, K., (2015). Fruit ripening of climacteric and non climacteric fruit. *Journal of Environmental and Applied Bioresearch*, 4(1), 27-34.
- Krongyut, W., Srilaong, V., Uthairatanakij, A., Wongs-Aree, C., Esguerra, E. B. & Kanlayanarat, S. (2011). Physiological changes and cell wall degradation in papaya fruits cv. ‘Kaek Dum’ and ‘Red Maradol’ treated with 1- methylcyclopropene. *International Food Research Journal*, 4(18), 1251-1259.
- Liu, M., Pirrello, J., Chervin, C., Roustan, J. P. & Bouzayen, M. (2015). Ethylene Control of Fruit Ripening: Revisiting the Complex Network of Transcriptional Regulation. *Plant Physiology*, 169, 2380-2390.
- Nath, P., Bouzayen, M., Mattoo, A. K. & Pech, J. C. (2014). *Fruit Ripening: Physiology, Signalling and Genomics*. United Kingdom: CPI Group (UK) Ltd.
- Noiwan, D., Kumpoun, W., Whangcha, K., Saengnil, K., & Uthaibutra, J. (2016). Effect of Coating Materials on Storage Life of Mango cv. Mahajanaka. *Agricultural Science Journal*, 5(37), 132-135.

## บรรณานุกรม (ต่อ)

- Pacharawongsakda, E. (2014). *An Introduction to Data Mining Techniques*. Bangkok: Data Cube.
- Protasio Pereira, L. F., Galvao, R. M., Kobayashi, A. K., B.Cacao, S. M. & Esteves Vieira, L. G. (2005). Ethylene production and acc oxidase gene expression during fruit ripening of *Coffea arabica* L. *Braz. J. Plant Physiol.*, 17(3) Londrina July/Sept, 7-10.
- Quinlan, J. R.(1993). *C4.5, Programs for Machine Learning*. CA: San Mateo.
- RapidMiner. (2017). *Operator Reference Guide*. Retrieved from [https://docs.rapidminer.com/latest/studio/operators/validation/cross\\_validation.html](https://docs.rapidminer.com/latest/studio/operators/validation/cross_validation.html)
- Rowley, H. A., Baluja, S. & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 20(1) (January 1998), 23 - 38.
- Subhadrabandhu, S. & Ketsa, S. (2001). *Durian: King of Tropical Fruit*. Bangkok: CABI Publishing.
- Wills, R. B. H., Harris, D. R., Spohr, L. J., & Golding, J. B. (2014). Reduction of energy usage during storage and transport of bananas by management of exogenous ethylene levels. *Postharvest Biology and Technology*, 89(1), 7-10.





ผนวก ก

มหาวิทยาลัยรังสิต Rangsit University

## บทความที่ได้รับการตีพิมพ์ในการประชุมวิชาการระดับชาติมหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9

การประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9 “วิจัยและนวัตกรรมเพื่อสังคม” | 517

### การวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลินโดยใช้เทคนิคการทำเหมืองข้อมูล

อุดมศักดิ์ เพียรไพโรจน์<sup>1\*</sup> วุฒิพงษ์ ชินศรี<sup>2</sup>

#### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์และจำแนกประเภทระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลินโดยใช้เทคนิคการทำเหมืองข้อมูล ซึ่งก๊าซเอทีลินนั้นมีความสัมพันธ์กับระดับการสุกของผลไม้ ผู้วิจัยจึงพัฒนาอุปกรณ์วัดระดับการสุกของทุเรียนด้วย Arduino และเชื่อมต่อกับเซ็นเซอร์ MQ3, MQ6 และ MQ8 เพื่อตรวจสอบการแพร่กระจายของก๊าซเอทีลิน แทนเซ็นเซอร์ก๊าซเอทีลินซึ่งมีราคาค่อนข้างแพง เมื่อได้ข้อมูลจาก sensor จำนวน 160 รายการ ผู้วิจัยได้ใช้เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering) ในการแบ่งกลุ่มทุเรียนตามระดับการสุก โดยเลือกอัลกอริทึม K-Mean มาใช้ในการทดลอง ซึ่งผลการทดลองสามารถแบ่งกลุ่มระดับการสุกได้เป็น 4 กลุ่ม คือ 1) ดิบ 2) ห้าม 3) สุก และ 4) งอม ตามลักษณะการสุกของทุเรียน หรือกล้วยที่คนไทยส่วนใหญ่เข้าใจกัน สำหรับการวิเคราะห์ระดับการสุกของทุเรียนนั้น ผู้วิจัยใช้เทคนิคการจำแนกข้อมูล (Classification) ในการสร้างโมเดล โดยทำการเปรียบเทียบอัลกอริทึมที่นำมาใช้ในการสร้างโมเดล ได้แก่ Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naive Bayes และ Support Vector Machine (SVM) โดยเปรียบเทียบค่าความแม่นยำในการจำแนกข้อมูล (Accuracy) โดยแบ่งกลุ่มของข้อมูลเพื่อทำการทดสอบแบบ Cross validation ซึ่งพบว่า Support Vector Machine และ Neural Network มีค่าความแม่นยำในการจำแนกข้อมูลสูงที่สุดใกล้เคียงกัน นอกจากนี้โมเดลที่ได้จากทุกอัลกอริทึมมีค่าความแม่นยำในการจำแนกข้อมูลสูง เกือบ 100% จึงสามารถนำโมเดลเหล่านั้นไปพัฒนาต่อเพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทีลินได้

**คำสำคัญ :** ก๊าซเอทีลิน, การสุกของผลไม้, เซ็นเซอร์, การทำเหมืองข้อมูล, การแบ่งกลุ่มของข้อมูล, การจำแนกข้อมูล

<sup>1</sup> หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมการผลิตและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต  
e-mail: UdomsakPaeanpairoj@gmail.com

<sup>2</sup> สาขาวิชาการจัดการความมั่นคงปลอดภัยไซเบอร์ วิทยาลัยนวัตกรรมการผลิตและเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต  
e-mail: wutthipong.c@su.ac.th

\*ผู้นิพนธ์หลัก e-mail: UdomsakPaeanpairoj@gmail.com

THE ANALYSIS OF DURIAN RIPENING LEVELS FROM DIFFUSION OF ETHYLENE GASES  
BY USING DATA MINING TECHNIQUES

Udomsak Paeanpairoj<sup>1</sup> Wutthipong Chinnasri<sup>2</sup>

**Abstract**

The purpose of this research aims to analyze Durian ripening level from diffusion of Ethylene gas by data mining techniques. Due to the fact that Ethylene is associated with fruit ripening level. A tool, Arduino board connected with MQ3, MQ6 and MQ8 detection sensors, has been developed for the assessment. These sensors are substituted for Ethylene one which is quite expensive. After collecting 160 data by this tool, data mining procedures are mainly used for data clustering. The experimental result demonstrates that K-Means clustering algorithm classify durian ripening level into 4 types; 1) unripe, 2) pre-ripe, 3) ripe, 4) overripe follow the ripening level of Durian or Banana as Thai people know. The ripening analysis has been use the classification technique, When it comes to classification, various types of algorithm, for instance, Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naive Bayes and Support Vector Machine (SVM), have been compared by accuracy rate of their performance through cross validation procedure. It has been found that Support Vector Machine and Neural Network algorithm are similarly the most accurate. However, the accuracy of all algorithm in this experimental closely to 100%, these algorithms can be further developed to analyze fruit ripening level from diffusion of gas analysis

**Keywords:** Ethylene, Ripening, Gas Sensor, Data Mining, Clustering, Classification

---

<sup>1</sup> Master of Science Program in Digital Innovation and Information Technology, Rangsit University  
e-mail : [udomsakpaeanpairoj@gmail.com](mailto:udomsakpaeanpairoj@gmail.com)

<sup>2</sup> Cybersecurity Management College of Digital Innovation and Information Technology, Rangsit University  
e-mail: [wutthipong.c@rsu.ac.th](mailto:wutthipong.c@rsu.ac.th)

\*Corresponding author e-mail : [udomsakpaeanpairoj@gmail.com](mailto:udomsakpaeanpairoj@gmail.com)

## บทนำ

ประเทศไทยเป็นประเทศอยู่ในเขตร้อนชื้นจึงมีผลไม้หลากหลายชนิด ซึ่งผลไม้สามารถจัดแบ่งประเภทได้หลายรูปแบบขึ้นอยู่กับวิธีการเลือกปัจจัยที่จะนำมาใช้แบ่งประเภท ซึ่งในงานวิจัยนี้สนใจเกี่ยวกับการวิเคราะห์ระดับการสุกของผลไม้ จึงขอแบ่งประเภทของผลไม้เป็น 2 ประเภท (Kaushlendra, Saurabh, Murtaza, & Tanushri, 2015) ตามอัตราการหายใจ ได้แก่ ผลไม้ที่มีอัตราการหายใจเปลี่ยนแปลงตามอายุหลังเก็บเกี่ยวผลไม้ที่แก่แล้ว (Climacteric fruit) หรือสามารถบ่มให้สุกได้ เช่น กล้วย มะม่วง มะละกอ ทูเรียน และผลไม้ที่มีอัตราการหายใจไม่เปลี่ยนแปลงตามอายุหรือเปลี่ยนแปลงตามอายุในอัตราที่น้อยมากหลังเก็บเกี่ยวผลไม้ที่แก่แล้ว (Non-Climacteric Fruit) หรือไม่สามารถบ่มให้สุกได้ เช่น ส้ม สับปะรด มะนาว และเงาะ

ปัญหาในการจัดเก็บ การขนส่ง และการซื้อขายผลไม้ นั้น มักเกิดกับผลไม้ที่สามารถบ่มสุกได้เพราะถ้าเกษตรกร ผู้ประกอบการ และผู้จำหน่ายวางแผนในการจัดเก็บและดำเนินการขนส่งได้ไม่สมบูรณ์จะทำให้ผลไม้ที่นำไปจำหน่ายให้กับผู้บริโภคไม่ตรงตามความต้องการ ทำให้ต้องสูญเสียเวลาและค่าใช้จ่ายในการจัดเก็บสินค้าเพิ่มเติม ซึ่งมีงานวิจัยหลายชิ้นที่ศึกษาเกี่ยวกับการควบคุมระดับการสุกของผลไม้ในขั้นตอนการจัดเก็บและการขนส่งสินค้า (R.B.H. Wills, D.R. Harris, L.J. Spohr, & J.B. Golding, 2014), (Michael M Blanke, 2014), (Noiwan, Kumpoun, Whangcha, Saengnil, & Uthaibutra, 2006) อย่างไรก็ตามในมุมมองของผู้วิจัยนั้นผลไม้ที่สามารถบ่มสุกได้และสามารถมองเห็นการเปลี่ยนแปลงสีผิวของผลไม้ได้ด้วยเอกลักษณ์ (Non-Scientific) เกษตรกรหรือผู้จำหน่ายผลไม้ยังพอที่จะระบุได้ว่าผลไม้ที่มีระดับการสุกมากน้อยเพียงใด โดยอาศัยประสบการณ์และการสังเกตสีผิวของผลไม้ แต่ผลไม้ที่สามารถบ่มสุกได้และสีผิวไม่เปลี่ยนแปลงหรือเปลี่ยนแปลงเล็กน้อย เช่น ทูเรียน ขนุน เกษตรกรหรือผู้จำหน่ายต้องอาศัยประสบการณ์และการสังเกตเป็นอย่างมาก ผู้วิจัยจึงมีแนวคิดที่จะทำการวิเคราะห์ระดับการสุกของผลไม้จากการแพร่กระจายของก๊าซโดยใช้เทคนิคการทำเหมืองข้อมูล โดยเลือก ทูเรียน มาเป็นตัวแทนของผลไม้ที่สามารถบ่มสุกได้และไม่สามารถสังเกตจากภายนอกได้ว่ามีระดับการสุกมากน้อยเพียงใด มาใช้ในการทดลอง ประกอบกับทูเรียนเป็นผลไม้ชนิดหนึ่งที่ทำให้ประเทศไทยมีชื่อเสียงโด่งดังไปทั่วโลก และได้รับการขนานนามว่าเป็นราชาแห่งผลไม้ (Subhadrabandhu & Ketsa, 2001) โดยผู้บริโภคจะเลือกทานทูเรียนในระดับการสุกที่แตกต่างกัน ซึ่งจากที่ผู้วิจัยได้ศึกษางานวิจัยที่นำเสนอเกี่ยวกับระดับการสุกของทูเรียนจะแบ่งออกเป็น 3 ระดับ (Arancibia-Avila, et al., 2008) คือ ดิบ (Unripe) สุก (Ripe) และงอม (Overripe) แต่ในสภาพความเป็นจริงแล้วคนไทยบางส่วนชอบทานทูเรียนห่าม (Pre Ripe) และในงานวิจัยที่ศึกษาเกี่ยวกับการเปลี่ยนแปลงทางกายภาพของมะละกอ ได้กล่าวถึงระดับการสุกที่เรียกว่า ห่าม ไว้ว่าสีผิวของมะละกอ จะเป็นสีเหลืองประมาณ 10-20% (Krongyut, W, et al., 2011) ประกอบกับยังมีผลไม้อื่น ๆ เช่น กล้วย มะม่วง ที่มีระดับการสุกที่เรียกว่าห่าม ซึ่งผลไม้เหล่านี้จะมีการเปลี่ยนสีผิวจากสีเขียวเป็นสีเหลืองในบางส่วน ในงานวิจัยนี้ผู้วิจัยจึงขอแบ่งระดับการสุกของทุเรียนเป็น 4 ระดับ คือ ดิบ ห่าม สุก และงอม เนื่องจากทุเรียนดิบสามารถนำไปทำเป็นทุเรียนทอดได้ ส่วนทุเรียนงอมสามารถนำไปเป็นทุเรียนกวนได้ แต่สำหรับทุเรียน ห่าม และสุก มักนำไปจำหน่ายให้กับผู้บริโภคโดยไม่ต้องแปรรูป ซึ่งผู้จำหน่ายใช้อุปกรณ์ไม้เคาะร่วมกับการสังเกตสีของหนามเพื่อระบุว่าทุเรียนสุกในระดับใด แต่ก็ไม่สามารถมั่นใจได้จึงต้องมีการผ่าออกบางส่วนเพื่อให้ผู้บริโภคลองใช้นิ้วกดลงไปเนื้อทุเรียน เพื่อตรวจสอบอีกครั้งว่ามีระดับความสุกตามที่ต้องการหรือไม่ ซึ่งวิธีการดังกล่าวยังไม่มีความแม่นยำและไม่สามารถสร้างความน่าเชื่อถือให้กับผู้บริโภคได้ อีกทั้งยังจำเป็นต้องผ่าทุเรียนออกบางส่วนจึงจะสามารถตรวจสอบระดับความสุกของทุเรียนได้ ซึ่งเมื่อผู้วิจัยได้ศึกษาเกี่ยวกับกระบวนการสุกของผลไม้ (Fruit Ripening Process) ที่สามารถบ่มสุกได้ เมื่อนำมาบ่มต่อหรือปล่อยให้ทิ้งไว้ในสภาพที่เหมาะสมจะ

เกิดการเปลี่ยนแปลงหลายประการ อาทิเช่น รสชาติ สี กลิ่น ซึ่งหากพิจารณาทางด้านเคมีแล้ว (Scientific) กระบวนการสุกของผลไม้จะมีความเกี่ยวข้องกับก๊าซเอทิลีน (Ethylene) ซึ่งเอทิลีนจะมีค่าต่ำเมื่อผลไม้ยังดิบและจะมีค่าสูงเมื่อผลไม้สุก

ก๊าซเอทิลีนจึงเป็นปัจจัยสำคัญในการสุกของผลไม้ (Liu, Pirrello, Chervin, Roustan, & Bouzayen, 2015), (Nath, P., Bouzayen, M., Mattoo, A. K., & Pech, J. C., 2014) ดังนั้น หากสามารถวัดระดับของก๊าซเอทิลีนได้จะสามารถวัดระดับการสุกผลไม้ได้เช่นกัน ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์ เพื่อวิเคราะห์ระดับการสุกของผลไม้จากการแพร่กระจายของก๊าซโดยใช้เทคนิคการทำเหมืองข้อมูล โดยผู้วิจัยจะใช้เทคนิคการทำเหมืองข้อมูลมาช่วยในการวิเคราะห์และแบ่งประเภทข้อมูล ซึ่งน่าจะช่วยให้เกษตรกร ผู้ประกอบการ หรือผู้จำหน่ายสามารถวางแผนเพื่อลดเวลาในการขนส่งและค่าใช้จ่ายในการจัดเก็บ รวมไปถึงการเลือกทุเรียนจำหน่ายให้กับผู้บริโภคได้อย่างมีประสิทธิภาพ เพราะสามารถวัดระดับการสุกของทุเรียนได้

### วัตถุประสงค์การวิจัย

1. เพื่อทดลองการแบ่งกลุ่มทุเรียนจากปริมาณก๊าซเอทิลีนโดยใช้เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering)
2. เพื่อวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีนโดยใช้เทคนิคการจำแนกข้อมูล (Classification)
3. เพื่อเปรียบเทียบอัลกอริทึมที่ใช้ในการสร้างโมเดลสำหรับวิเคราะห์ระดับการสุกของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีน

### วิธีการดำเนินการวิจัย

การดักจับการแพร่กระจายของก๊าซเพื่อนำค่าต่าง ๆ มาวิเคราะห์นั้น สามารถใช้ Arduino ซึ่งเป็นบอร์ดไมโครคอนโทรลเลอร์ที่สามารถเชื่อมต่อกับเซ็นเซอร์ (Sensor) ต่าง ๆ ซึ่งรวมไปถึงก๊าซเซ็นเซอร์ (Gas Sensor) ดังนั้นหากนำเอทิลีนก๊าซเซ็นเซอร์ (Ethylene Gas Sensor) มาเชื่อมต่อกับ Arduino จะสามารถดักจับก๊าซเอทิลีนและสามารถนำข้อมูลที่ตรวจจับได้มาทำการวิเคราะห์ แต่อย่างไรก็ตามเอทิลีนก๊าซเซ็นเซอร์ นั้นไม่มีจำหน่ายโดยทั่วไปในประเทศไทย อีกทั้งยังมีราคาสูงเมื่อเทียบกับก๊าซเซ็นเซอร์อื่น ๆ เช่น ไฮโดรเจนก๊าซเซ็นเซอร์ (Hydrogen Gas Sensor) แอลกอฮอล์ก๊าซเซ็นเซอร์ (Alcohol Gas Sensor) ผู้วิจัยจึงทำการวิเคราะห์สูตรทางเคมีของก๊าซเอทิลีน คือ  $C_2H_4$  (A. Eberhardt, K. Schmitt, & J. Wöllenstein, 2017) ซึ่งหมายถึง มีคาร์บอน 2 อะตอม และมีไฮโดรเจน 4 อะตอม ผู้วิจัยจึงใช้ MQ3 gas sensor จับแอลกอฮอล์; เอทานอล (Ethanol;  $C_2H_5OH$ ) ซึ่งเป็นองค์ประกอบระหว่างเอทิลีน ( $C_2H_4$ ) และน้ำ ( $H_2O$ ) (D.O. ADAMS AND S. F. YANG, 1978) จากนั้นใช้ MQ6 gas sensor จับก๊าซไฮโดรเจน (Hydrogen) และในส่วนของก๊าซคาร์บอน (C) นั้นยังไม่มี sensor ใดที่ใช่ตรวจจับได้ ผู้วิจัยจึงใช้ MQ8 Gas Sensors ที่เป็น sensor ตรวจจับ LPG ตรวจจับแทนได้เนื่องจากก๊าซ LPG มีส่วนประกอบด้วยของโพรเพน, บิวเทน ซึ่งมีคาร์บอนเป็นธาตุประกอบหลัก (Norazlan Hashim, UMP, 2008) ส่วนการทำงานของ Arduino นั้นเมื่อได้รับค่าปริมาณก๊าซที่ถูกส่งมาจาก MQ Gas Sensors อุปกรณ์จะนำค่าต่าง ๆ ไปจัดเก็บแสดงผล และนำค่านั้นไปประมวลผลต่อโดยการใช้เทคนิคการทำเหมืองข้อมูล โดยมีวิธีการดำเนินการวิจัยดังนี้



1. ผู้วิจัยได้พัฒนากล่องอุปกรณ์สำหรับวัดระดับการสุกของผลไม้โดยจำลองมาจากระบบการหายใจของมนุษย์โดยนำ Arduino และ Gas Sensors ต่างๆ เข้าไปติดตั้งพร้อมกับพัดลมเพื่อดูดเอาก๊าซต่างๆ ที่ห่อหุ้มรอบทุเรียนเข้ามาประมวลผล ตามภาพที่ 1



ภาพที่ 1 แสดงกล่องสำหรับวัดระดับการสุกของผลไม้

เมื่อผู้วิจัยนำเอาอุปกรณ์ตามภาพที่ 1 ทำการวัดระดับก๊าซแอลกอฮอล์ ไฮโดรเจน และคาร์บอน โดยการนำทุเรียนใส่ในอุปกรณ์ระบบจะทำงานโดยแสดงผลค่าของปริมาณก๊าซออกมาเป็นลักษณะกราฟ และตัวเลข ผู้วิจัยจะเริ่มเก็บข้อมูลเมื่อกราฟสูงที่สุดและเริ่มเป็นเส้นตรง (Saturation point) จากนั้นจึงเริ่มเก็บข้อมูลที่เป็นตัวเลขระดับปริมาณก๊าซของทุเรียน

2. หลังจากเก็บข้อมูลตามที่ต้องการได้แล้ว ผู้วิจัยจะใช้เทคนิคการทำเหมืองข้อมูล (Data mining) มาช่วยในการวิเคราะห์ โดยการใช้เทคนิคการแบ่งกลุ่มของข้อมูล (Clustering) ในการตรวจสอบเบื้องต้นว่าข้อมูลที่จัดเก็บได้นั้นสามารถแบ่งออกได้กี่กลุ่ม และใช้เทคนิคการจำแนกข้อมูล (Classification) เพื่อสร้างโมเดลในการทำนายระดับการสุกของผลไม้ ซึ่งจะเริ่มจากนำข้อมูลที่จัดเก็บเอาไว้จำนวน 160 ชุด (Attribute) แล้วใช้วิธีการ Cross Validation (RapidMiner, 2017) โดยผู้วิจัยจะทำการทดลองใช้อัลกอริทึมที่ต่างกันใน การสร้างโมเดล ได้แก่ Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naive Bayes และ Support vector machine (SVM) จากนั้นจะนำโมเดลที่ได้มาเปรียบเทียบกันเพื่อดูว่าโมเดลใดมีความแม่นยำในการจำแนกกลุ่มข้อมูล (accuracy) สูงที่สุด ผู้วิจัยได้ออกแบบการทดลองไว้ 3 ขั้นตอนดังนี้

- 2.1 สร้างเครื่องมือที่ใช้สำหรับการเก็บค่าก๊าซต่าง ๆ
- 2.2 ประเมินระดับการสุกโดยมนุษย์และจัดเก็บค่าก๊าซต่าง ๆ ที่ได้จากเซ็นเซอร์แต่ละชนิด
- 2.3 นำข้อมูลที่จัดเก็บมาทำการทดสอบด้วยเทคนิคของการทำเหมืองข้อมูล

3. ในการทดลองนี้ผู้วิจัยกำหนดให้ทุกอัลกอริทึมใช้ Default Parameter Set เพื่อสร้างโมเดลและนำโมเดลที่ได้มาเปรียบเทียบกันว่าโมเดลไหนมีประสิทธิภาพที่ดีที่สุด จากนั้นผู้วิจัยได้ทดลองใช้ Optimize Parameter Set ทำการทดลองอีกครั้งโดยมีการออกแบบการปรับพารามิเตอร์ไว้ดังนี้

3.1 Decision Tree (R. Quinlan., 1993) ผู้วิจัยจะใช้ข้อมูลแต่ละคุณลักษณะเพื่อให้ระบบพิจารณา ค่าความแตกต่างของ Entropy และตัดสินใจแบ่งเป็นกลุ่มย่อยและพิจารณาเลือกกลุ่มตามคุณลักษณะต่างๆ จากผลลัพธ์ของ Normalized information gain ที่สูงสุด ผู้วิจัยจะใช้พารามิเตอร์ Decision Tree.criterion, Decision Tree.maximal\_depth, Decision Tree.confidence และ Decision Tree.apply\_pruning ในการทดสอบ



3.2 K Nearest Neighbors (KNN) ผู้วิจัยใช้พารามิเตอร์ `k-NN.kernel_type`, `k-NN.k`, `k-NN.weighted_vote`, `k-NN.numerical_measure` และ `k-NN.measure_types` เนื่องจากเป็นฟังก์ชันที่สามารถกำหนดน้ำหนัก และระยะห่าง เพื่อให้กำหนดค่า `k` โดยทดสอบบนพื้นฐานของ `cross-validation` ในการทดลองนี้ผู้วิจัยจะทดสอบกำหนดค่า `k` ในลักษณะต่างๆ เพื่อให้เกิดความแม่นยำมากที่สุด

3.3 Neural Network ผู้วิจัยจะนำค่า Input ไปกำหนดค่าน้ำหนัก (Weight) ระหว่างโหนดและปรับเปลี่ยนค่าน้ำหนักด้วยการนำค่าน้ำหนักคูณกับ Input ของแต่ละขา ผลที่ได้จากทุกๆ ขานั้นจะนำมาเปรียบเทียบกับ Threshold ที่กำหนดไว้โดยผู้วิจัยจะทำการทดลองกำหนดพารามิเตอร์ด้วย `Neural Net.learning_rate` เพื่อคูณกับ Error Rate และ Node ที่เข้ามา แล้วบวกเข้ากับค่าน้ำหนัก เพื่อทำให้เกิดค่าน้ำหนักใหม่ `Neural Net.training_cycles` เพื่อระบุจำนวนรอบการทดสอบใน `back-propagation` ค่า Output จะถูกเปรียบเทียบกับคำตอบที่ถูกต้องเพื่อคำนวณค่าของฟังก์ชัน `error-defined` ที่กำหนดไว้ล่วงหน้า ข้อผิดพลาด จะถูกส่งกลับผ่านเครือข่าย การใช้ข้อมูลนี้อัลกอริทึมจะปรับน้ำหนักของแต่ละการเชื่อมต่อเพื่อลดค่าของฟังก์ชันข้อผิดพลาดด้วยจำนวนเล็กน้อย `Neural Net.momentum` เพื่อใช้ในการปรับค่าน้ำหนัก และใช้เพื่อเหนี่ยวนำค่า Output ที่ได้ในแต่ละรอบ เพื่อป้องกันการเกิด Error จากการ Train และใช้ `Neural Net.error_epsilon` ใช้เพื่อการเพิ่มประสิทธิภาพ และจะหยุดทำงานหากค่าการ Training ต่ำกว่าค่า `epsilon`

3.4 Naïve Bayes ผู้วิจัยจะทำการทดลองจำแนกข้อมูลโดยใช้หลักความน่าจะเป็นซึ่งเป็นพื้นฐานและสมมติฐานของ Bayes' Theorem โดยการใช้ความสัมพันธ์ของแต่ละโหนดและกำหนดให้มีความสัมพันธ์กันตามแบบที่แบบจำลอง โดยจะนำค่าตัวแปรที่เกี่ยวข้องเก็บไว้ในแต่ละโหนดโดยใช้เงื่อนไขความน่าจะเป็น หากภายในข้อมูลการ Training มีค่าแอตทริบิวต์ที่ระบุไม่เคยเกิดขึ้นในบริบทของคลาสที่ระบุความเป็นไปได้เชิงเงื่อนไขจะถูกตั้งค่าเป็นศูนย์ เมื่อค่าศูนย์นี้คูณพร้อมกับความน่าจะเป็นอื่นๆ ค่าเหล่านี้จะถูกตั้งค่าเป็นศูนย์และผลลัพธ์จะทำให้เข้าใจผิด ผู้วิจัยจึงทดลองปรับพารามิเตอร์ `Naive Bayes.laplace_correction` เพื่อหลีกเลี่ยงปัญหานี้ให้เพิ่มหนึ่งรายการในแต่ละจำนวนเพื่อหลีกเลี่ยงการเกิดค่าเป็นศูนย์ ในการเพิ่มจำนวนนั้นนั้นจะมีผลเพียงน้อยนิดต่อความน่าจะเป็น ในส่วนของ `Naive Bayes (Kernel).estimation_mode` ใช้เพื่อระบุความหนาแน่นของเคอร์เนลโดยใช้ตัวเลือกสองแบบ คือ `Full`, `greedy` และใช้ `Naive Bayes (Kernel).bandwidth_selection` เพื่อระบุความหนาแน่น ความเร็วของ `Kernel Approximation`

3.5 Support Vector Machine (SVM) ผู้วิจัยจะทำการแปลงข้อมูลให้อยู่ในสมการเส้นตรงด้วยการใช้วิธีการ `Kernel function` เพื่อแปลงข้อมูลให้โมเดล `Linear` ไปอยู่ใน `Dimension` ที่สูงขึ้นด้วยการเรียงตัวของข้อมูลนี้จะได้พื้นที่มิติที่สูงขึ้น (`Higher Dimensional Space`) ผู้วิจัยจะทดลองปรับพารามิเตอร์ `SVM.kernel_type` โดยจะทดสอบแบบ `linear`, `polynomial`, `radial basis function - RBF` และ `sigmoid` จากนั้นจะใช้ `SVM.gamma` เพื่อใช้ตรวจสอบเคอร์เนล `SVM.kernel_type` และใช้ `SVM.C` เพื่อป้องกันข้อผิดพลาดของ `SVM.svm_type` ที่เป็น `c-SVC`, `epsilon-SVR`, `nu-SVR`

## ผลการวิจัย

หลังจากที่ผู้วิจัยได้พัฒนาเครื่องมือที่ใช้สำหรับการเก็บค่าก๊าซต่าง ๆ ของทุเรียนได้แล้วนั้นผู้วิจัยพบว่าทุเรียนดิบ ที่ถูกตัดมาจากต้นนั้นจะมีอัตราการหายใจสูง (`Ethylene`) ใกล้เคียงกับทุเรียนที่มีอาการอ้อมที่

มีอัตราการหายใจสูงสุด (Peak) ซึ่งสามารถพิสูจน์ได้จากปริมาณก๊าซที่ทุเรียนคายออกมา จากนั้นผู้วิจัยได้เชิญผู้เชี่ยวชาญทำการประเมินระดับการสุกและทำการจัดเก็บค่าต่าง ๆ เรียบร้อยแล้วนั้น ผู้วิจัยจะนำข้อมูลที่จัดเก็บมาได้ นำไปทำเหมืองข้อมูล จึงได้ผลการทดลองดังนี้

1. ทำการแบ่งกลุ่มโดย K-Mean เพื่อกำหนดกลุ่มของระดับการสุก ซึ่งผู้วิจัยได้ศึกษาจากงานวิจัยที่เกี่ยวข้องพบว่า K-Mean สามารถใช้ในการแบ่งกลุ่มเมื่อทราบจำนวนกลุ่มที่แน่นอน (สุรวัชร ศรีเปารยะ และ สายชล สิมบูรณ์ทอง, 2560) และ (วีระยุทธ พิมพารณ์ และ พยุง มีสังข์, 2557) โดยการกำหนดศูนย์กลางของแต่ละคลัสเตอร์ เพื่อหาระยะห่างระหว่างข้อมูลแต่ละตัวกับจุดศูนย์กลาง (Mean) ของแต่ละคลัสเตอร์โดยกำหนดให้ข้อมูลที่

อยู่คลัสเตอร์ที่ใกล้สุด จากนั้นคำนวณหาจุดศูนย์กลางของแต่ละคลัสเตอร์ใหม่ และทำซ้ำจนข้อมูลอยู่ในคลัสเตอร์เดิม ไม่มีการเปลี่ยนแปลง ในการทดลองนี้ผู้วิจัยจะไม่กำหนดค่า Label เพื่อให้ K-Mean แบ่งกลุ่มของข้อมูลแก่สทั้ง 3 ชนิด โดยเบื้องต้นผู้วิจัยตั้งใจกำหนดกลุ่มของระดับการสุกเป็น 4 ระดับ คือ ดิบ ห้าม สุก และงอม ตามลักษณะการสุกของทุเรียนหรือกล้วย ที่คนไทยส่วนใหญ่เข้าใจกัน แต่อย่างไรก็ตามผู้วิจัยได้ทดลองกำหนดค่า K ตั้งแต่ 2 ถึง 6 เพื่อดูผลการแบ่งกลุ่มด้วย K-Mean ว่าจะมีลักษณะเป็นอย่างไร ซึ่งผลการทดลองแสดงดังตารางที่ 1

ตารางที่ 1 แสดงค่า K ในการแบ่งกลุ่มด้วย K-Mean

	k=2	k=3	k=4	k=5	k=6
cluster_0	ดิบ(40) สุก (1) งอม (39)	ดิบ(2) สุก (34) งอม (37)	ดิบ(2) งอม(38)	ดิบ(25) งอม (2)	ดิบ (13) งอม (8)
cluster_1	ห้าม(40) สุก (39) งอม (1)	ดิบ (38) งอม (3)	สุก (40)	สุก (40)	ดิบ (23) งอม (1)
cluster_2		ห้าม (40) สุก (6)	ห้าม (40)	งอม (26)	สุก (38)
cluster_3			ดิบ (38) งอม (2)	ดิบ(15) งอม (12)	ดิบ (4) สุก (1) งอม (19)
cluster_4				ห้าม (40)	ห้าม (40)
cluster_5					สุก (1) งอม (12)

จากตารางที่ 1 หากกำหนด K = 2 จะแบ่งกลุ่มได้ คือ 1) ดิบหรืองอม และ 2) ห้ามหรือสุก ซึ่งไม่เหมาะสม เนื่องจาก ดิบกับงอมนั้นในการเลือกซื้อนั้นถือว่ามี ความแตกต่างกันอย่างสิ้นเชิง เมื่อกำหนด K = 3 จะแบ่งกลุ่มได้ คือ 1) ดิบ 2) ห้าม และ 3) สุกหรืองอม ซึ่งมีการแบ่งกลุ่มผิด จาก ดิบเป็นสุกหรืองอม จำนวน 2 ลูก จาก สุกเป็นห้าม จำนวน 6 ลูก และจากสุกหรืองอม เป็นดิบ จำนวน 3 ลูก เมื่อกำหนด K = 4 จะแบ่งกลุ่มได้ คือ 1) ดิบ 2) ห้าม 3) สุก และ 4) งอม ซึ่งมีการแบ่งกลุ่มผิด จาก ดิบเป็นงอม จำนวน 2 ลูก และจาก งอมเป็นดิบ 2 ลูก เมื่อกำหนด K = 5 และ K = 6 จะเริ่มมีกลุ่มซ้ำ คือ ดิบหรืองอม ซึ่งมีความแตกต่างกันดังที่กล่าวไว้ข้างต้น และมีการแบ่งกลุ่มผิดจำนวนมาก การกำหนดให้ K = 4 จึงแบ่งกลุ่มระดับการสุกจากค่าก๊าซทั้ง 3 ชนิด ได้เหมาะสมที่สุด และเป็นการจัดแบ่งกลุ่มที่เหมาะสมกับการเลือกซื้อทุเรียนตามที่คุณวิจัยคาดหวังเอาไว้

2. ในการจำแนกระดับการสุกของทุเรียนนั้นจะต้องอาศัยองค์ประกอบที่สำคัญคือปริมาณก๊าซที่ได้รับจากเซ็นเซอร์ต่าง ๆ ดังนั้นการจำแนกข้อมูล นั้น ผู้ใช้จะต้องเลือกอัลกอริทึมที่เหมาะสมใช้งาน ผู้วิจัยจึงทำการทดลอง

เปรียบเทียบอัลกอริทึมต่าง ๆ มาใช้ในการวิจัย ได้แก่ Decision Tree, K Nearest Neighbours (KNN), Neural Network, Naive Bayes และ Support vector machine (SVM) จากการที่ผู้วิจัยกำหนดให้ทุกอัลกอริทึมใช้ Default Parameter Set เพื่อนำโมเดลที่ได้มาเปรียบเทียบเพื่อหาโมเดลที่มีประสิทธิภาพที่สุด และผู้วิจัยได้ทดลองใช้ Optimize Parameter Set ใน Rapid Miner โดยการปรับค่าพารามิเตอร์ ตามตารางที่ 2 และนำผลที่ได้ไปทำการทดลองอีกครั้ง ซึ่งผลการวิจัยที่ได้แสดงดังตารางที่ 3

ตารางที่ 2 แสดงพารามิเตอร์และค่าพารามิเตอร์ทำการปรับค่า

Algorithm	Parameter	Optimize parameter
Decision Tree	Decision Tree.criterion	= gain_ratio
	Decision Tree.maximal_depth	= -1
	Decision Tree.apply_pruning	= true
	Decision Tree.confidence	= 1.0E-7
K Nearest Neighbors (KNN)	k-NN.k	= 1
	k-NN.weighted_vote	= true
	k-NN.numerical_measure	= EuclideanDistance
	k-NN.kernel_type	= dot
Neural Network	k-NN.measure_types	= MixedMeasures
	Neural Net.training_cycles	= 100
	Neural Net.learning_rate	= 0.3
	Neural Net.momentum	= 0.2
Naïve Bayes	Neural Net.error_epsilon	= 0
	Naive Bayes (Kernel).estimation_mode	= full
	Naive Bayes (Kernel).bandwidth_selection	= heuristic
Support vector machine (SVM)	Naive Bayes.laplace_correction	= false
	SVM.kernel_type	= linear
	SVM.gamma	= Infinity
	SVM.C	= Infinity
	SVM.svm_type	= C-SVC

ตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพของ Algorithm

Algorithm	ดีบ	ห้าม	สูง	งอม	accuracy	
Decision Tree	Default	95%	100%	90%	80%	91.25%
	Parameter Set	ดีบ(38) งอม(2)	ห้าม(40)	สูง(36) งอม(4)	ดีบ(4) ห้าม(4) งอม(32)	
	Optimize	100%	100%	100%	97.50%	
	Parameter Set	ดีบ(40)	ห้าม(40)	สูง(40)	งอม(39) งอม(1)	
K Nearest Neighbours (KNN)	Default	95%	100%	100%	92.50%	96.88%
	Parameter Set	ดีบ(38) งอม(2)	ห้าม(40)	สูง(40)	งอม(37) ดีบ(3)	
	Optimize	100%	100%	100%	97.50%	
	Parameter Set	ดีบ(38)	ห้าม(40)	สูง(40)	งอม(39) ดีบ(1)	
Neural Network	Default	97.50%	100%	100%	90%	96.88%
	Parameter Set	ดีบ(39) งอม(1)	ห้าม(40)	สูง(40)	งอม(36) ดีบ(4)	
	Optimize	100%	100%	100%	97.50%	
	Parameter Set	ดีบ(40)	ห้าม(40)	สูง(40)	ดีบ(1) งอม(39)	
Naive Bayes	Default	80%	ห้าม(37)	85%	90%	86.88%
	Parameter Set	ดีบ(32) งอม(8)	สูง(1) งอม(2)	สูง(34) งอม(6)	ดีบ(4) งอม(36)	
	Optimize	90%	100%	97.50%	95%	
	Parameter Set	ดีบ(36) งอม(4)	ห้าม(40)	สูง(39) งอม(1)	ดีบ(2) งอม(38)	
Support vector machine (SVM)	Default	92.50%	100%	97.50%	92.50%	95.62 %
	Parameter Set	ดีบ(37) งอม(3)	ห้าม(40)	สูง(39) งอม(1)	ดีบ(2) งอม(37)	
	Optimize	100%	100%	100%	100%	
	Parameter Set	ดีบ(40)	ห้าม(40)	สูง(40)	งอม(40)	

จากตารางที่ 3 เมื่อกำหนดให้แต่ละอัลกอริทึมกำหนดพารามิเตอร์แบบ Default Parameter Set พบว่า Neural Network, K Nearest Neighbours (KNN) และ Support Vector Machine ทำงานได้ดีโดยมีค่าความแม่นยำในการจำแนกข้อมูลสูงที่สุดและรองลงมาตามลำดับ ส่วน Decision Tree และ Naive Bayes นั้นมีค่าความแม่นยำในการจำแนกข้อมูลต่ำกว่ากลุ่มแรกลงมาตามลำดับ แต่หลังจากที่ผู้วิจัยได้ทำการทดลองปรับพารามิเตอร์แบบ Optimize Parameter ซึ่งจากการทดลองดังกล่าว พบว่า Support Vector Machine, Neural Network และทำงานได้ดีโดยมีค่าความแม่นยำในการจำแนกข้อมูลสูงที่สุดเท่ากันที่ 100% ส่วน K Nearest Neighbours (KNN), Decision Tree และ Naive Bayes นั้นมีค่าความแม่นยำในการจำแนกข้อมูลต่ำรองลงมาตามลำดับแต่เมื่อเทียบกับการปรับค่า Default Parameter Set ในขั้นต้นแล้วอัลกอริทึมทุกตัว เมื่อปรับพารามิเตอร์แล้วมี ค่าความแม่นยำในการจำแนกข้อมูลสูงกว่า 95 เปอร์เซ็นต์

## สรุปผลและอภิปรายผลการวิจัย

หลังจากที่ผู้วิจัยได้พัฒนาอุปกรณ์ที่ทำการติดตั้งเซ็นเซอร์ต่างๆ ได้แก่ MQ3, MQ6, และ MQ8 และนำไปทดลองวัดปริมาณก๊าซของทุเรียนที่นำมาเป็นตัวอย่างเป็นตัวอย่างจำนวน 160 ลูก ซึ่งจากข้อมูลดังกล่าว ผู้วิจัยได้ใช้กระบวนการของการทำเหมืองข้อมูล โดยผู้วิจัยจะขอสรุปตามวัตถุประสงค์การวิจัยดังนี้

1. ในส่วนของการจัดแบ่งกลุ่มของข้อมูลซึ่งได้เลือกใช้อัลกอริทึม K-Mean ซึ่งผลการวิจัยพบว่าสามารถแบ่งการสุกของทุเรียนจากปริมาณก๊าซที่ได้จากเซ็นเซอร์ต่าง ๆ ได้เป็น 4 กลุ่ม คือ ดิบ ห่าม สุก และงอม ซึ่งเป็นไปตามแนวทางที่ผู้วิจัยได้ศึกษาไว้ และเป็นไปตามแนวทางในการเลือกซื้อทุเรียนของผู้บริโภค นอกจากนี้ผู้วิจัยยังค้นพบสาเหตุที่ทำให้โมเดลแบ่งกลุ่มผิดพลาดคือ ทุเรียนจะมีระดับการหายใจสูง เมื่ออยู่ในอาการดิบ และลดลงมาเมื่ออยู่ในอาการห่ามและลดลงอีกเมื่ออยู่ในอาการสุกและจะกลับมาหายใจสูงอีกครั้งเมื่ออยู่ในอาการงอม ด้วยเหตุผลนี้เองจึงทำให้อัลกอริทึมมีการจำแนกผิดพลาดเพราะระดับอาการดิบและอาการงอม มีลักษณะการหายใจและปริมาณก๊าซที่ใกล้เคียงกัน

2. ในส่วนของการจำแนกระดับการสุกของทุเรียนจากปริมาณก๊าซเอทีเอ็น ผู้วิจัยได้ใช้เทคนิคการจำแนกข้อมูล โดยเริ่มจากการสร้างโมเดลที่ใช้ในการประเมินระดับการสุก โดยผู้วิจัยเลือกอัลกอริทึมที่นำมาใช้ทั้งหมด 5 อัลกอริทึม ได้แก่ Decision Tree, K Nearest Neighbors (KNN), Neural Network, Naïve Bayes และ Support Vector Machine (SVM) ซึ่งผลการวิจัย ในกรณีที่ใช้พารามิเตอร์เริ่มต้นของ Rapid Miner พบว่า Neural Network, Support Vector Machine และ K Nearest Neighbours (KNN) ทำงานได้ดีโดยมีค่าความแม่นยำในการจำแนกข้อมูลสูงที่สุดและรองลงมาตามลำดับ ส่วน Naive Bayes และ Decision Tree นั้นมีค่าความแม่นยำในการจำแนกข้อมูลต่ำกว่าตามลำดับ

3. หลังจากนั้นผู้วิจัยได้เปรียบเทียบประสิทธิภาพของอัลกอริทึมด้วยการทดลองปรับพารามิเตอร์โดยใช้ Optimize Parameter ใน Rapid Miner ซึ่งจากการทดลองดังกล่าว พบว่า โมเดลที่มีค่าความแม่นยำในการจำแนกข้อมูลที่สูงที่สุดคือ Support Vector Machines (SVM) ซึ่งสอดคล้องกับงานวิจัยการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคมรอน (ภรณ์ยา อามฤตรัตน์ และ พยุง มีสัจ, 2553) งานวิจัยการพัฒนาประสิทธิภาพการจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ (นิเวศ จิระวิชาติชัย ,ปริญญา สงวนสัตย์ และ พยุง มีสัจ, 2554) และระบบแจ้งเตือนโซเซียลมีเดียไทยสำหรับธุรกิจด้วยซอฟต์แวร์เวกเตอร์แมชชีน (สุพัตรา วิริยะวิสุทธิสกุล, ปริญญา สงวนสัตย์, พิสิษฐ์ ชาญเกียรติกิจ และชูชาติ หล่อไชยศักดิ์, 2559) เพราะเทคนิค SVM นั้นมีหลักการของโมเดล Linear ที่แบ่งข้อมูลเป็นสอง Class จากนั้นจะใช้ kernel function เพื่อทำการแปลงข้อมูลเพื่อให้โมเดล Linear ไปอยู่ในมิติที่สูงขึ้น (Dimension) ซึ่งสมการเชิงเส้นนี้มีความสามารถในการแยกแยะข้อมูลทำให้เกิดการเพิ่มระยะของเส้นแบ่งเขตที่มีความกว้างเท่ากับ  $2 / ||w||^2$  ซึ่งสอดคล้องกับงานวิจัยการคัดแยกประเภทของมะเร็งเม็ดเลือดขาว โดยใช้วิธีการจัดอันดับร่วมกับเทคนิคซอฟต์แวร์เวกเตอร์แมชชีน (ภัทรารุณี แสงศิริ, ศจีมาจ ณ วิเชียร และ พยุง มีสัจ, 2553) ดังนั้นสำหรับงานวิจัยนี้เป็นไปตามวัตถุประสงค์โดยอัลกอริทึม SVM เหมาะที่จะนำมาใช้ในการประเมินระดับการสุกมากที่สุดในการจำแนกค่าความแม่นยำในการจำแนกข้อมูล



### ข้อเสนอแนะ

แต่อย่างไรก็ตามในอนาคตผู้วิจัยจะทำการพัฒนาอุปกรณ์สำหรับตรวจวัดความสุขของทุเรียน โดยใช้ Arduino นั้น การเลือก Algorithm จึงอาจต้องมีการพิจารณาเพิ่มเติมนอกจากประเด็นของความแม่นยำในการจำแนกข้อมูล เช่น ความยากง่ายในการนำโมเดลไปพัฒนา ความเร็วในการประมวลผล ต่อไป

### กิตติกรรมประกาศ

ขอขอบคุณ International Journal of Pure and Applied Mathematics ที่อนุญาตให้นำบทความของผู้วิจัยที่ได้รับการตีพิมพ์เป็นภาษาอังกฤษมาแปลเป็นภาษาไทยเพื่อเผยแพร่ให้กับประชาชนคนไทยอีกครั้งหนึ่งในการประชุมวิชาการระดับชาติ ขอขอบคุณคณะผู้จัดการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9 “วิจัยและนวัตกรรมเพื่อสังคม” ที่อนุญาตให้เผยแพร่บทความวิจัยนี้ ขอขอบคุณบิดา มารดา ผู้ให้กำเนิดและครอบครัวเพื่อโรฟโรจน์ ที่ให้การสนับสนุนด้านการศึกษาด้วยดีตลอดมา ขอขอบคุณ ผศ.ดร.วุฒิพงษ์ ชินศรี ในฐานะอาจารย์ที่ปรึกษาที่มอบวิชาความรู้ทั้งในด้านการศึกษาและด้านการพัฒนาคุณภาพชีวิต ขอขอบคุณสถาบันวิชาการป้องกันประเทศ กองบัญชาการกองทัพไทย ที่เป็นสื่อกลางในการมอบทุนการศึกษาในระดับปริญญาโท และท้ายสุดขอขอบคุณท่าน ดร.อาทิตย์ อุไรรัตน์ อธิการบดีมหาวิทยาลัยรังสิต ที่มอบทุนการศึกษาเรียนดีให้กับข้าพเจ้าจนสำเร็จการศึกษาระดับปริญญาโทนี้ ซึ่งข้าพเจ้าจะนำความรู้นี้ไปใช้พัฒนากองทัพและประเทศชาติสืบต่อไป

### เอกสารอ้างอิง

- นิเวศ จิระวิชาติชัย,ปริญญา สงวนสัตย์, และ พยุง มีสัจ. (2554). การพัฒนาประสิทธิภาพการจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติ. *วารสารพัฒนบริหารศาสตร์ ปีที่ 51 ฉบับที่ 3*, 188-205.
- ภรณ์ยา อามฤครัตน์ และ พยุง มีสัจ. (2553). การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม. *National Conference on Computing and Information Technology, v12, No.1*, 54 - 63.
- ภัทรารุณี แสงศิริ, วิเชียร ณศจีมาจ และ พยุง มีสัจ. (2553). การคัดแยกประเภทของมะเร็งเม็ดเลือดขาวโดยใช้วิธีการจัดอันดับร่วมกับเทคนิคซัพพอร์ตเวกเตอร์แมชชีน. *วารสาร มช.(บศ.) 10.(2)*, 10 - 17.
- วีระยุทธ พิมพ์ภรณ์ และ พยุง มีสัจ. (2557). การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยวิธีการเลือกลักษณะสำคัญแบบพลวัตเพื่อเพิ่มประสิทธิภาพของอัลกอริทึมการจัดกลุ่มบนปริภูมิย่อย. *วารสารเทคโนโลยีสารสนเทศ ปีที่ 10 ฉบับที่ 2*, 43-51.
- สุพัตรา วิริยะวิสุทธิสกุล,ปริญญา สงวนสัตย์, พิสิษฐ์ ชาญเกียรติก้อง, และชูชาติ หลุยยะศักดิ์. (2559). ระบบแจ้งเตือนโซเชียลมีเดียไทยสำหรับธุรกิจด้วยซัพพอร์ตเวกเตอร์แมชชีน. *สารสารปัญญาภิวัตน์ ปีที่ 8 ฉบับพิเศษ, ปีที่ 8 ฉบับพิเศษ ประจำเดือนสิงหาคม*, 223 - 233.
- สุรวีชร ศรีเปารยะ และ สายชล สิ้นสมบุญทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง. *วารสารวิทยาศาสตร์และเทคโนโลยี ปีที่ 25 ฉบับที่ 5*, 840-853.



- Adams, D. O., & Yang, S. F. (1979). Ethylene Biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic Acid as an Intermediate in the Conversion of Methionine to Ethylene. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 76, No. 1, Jan., 170-174.
- Arancibia-Avila, P., Toledo, F., Park, Y.-S., Jung, S.-T., Kang, S.-G., Heo, B.-G., & Gorinstein, S. (2008). Antioxidant properties of durian fruit as influenced by ripening. *LWT - Food Science and Technology*, 41(10), 2118-2125.
- Blanke, M. M. (2014). Reducing ethylene levels along the food supply chain: a key to reducing food waste? *Journal of the science of food and agriculture*, 94(12), 2357-2361.
- Eberhardt, A., Schmitt, K., & Wöllenstein, J. (2017). Rotating interference filter spectrometer for the detection of ethylene in the ripening process of climacteric fruit. *19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, (pp. 1445-1448). Kaohsiung, Taiwan.
- Hashim, N. (2008). Study on Combustion Characteristics of Different Compositions of LPG. *UMP*, 1 - 56.
- Kaushlendra, T., Saurabh, P., Murtaza, M., & Tanushri, K. (2015). Fruit ripening of climacteric and non climacteric fruit. *Journal of Environmental and Applied Bioresearch*, 4(1), 27-34.
- Krongyut, W., Srilaong, V., Uthairatanakij, A., Wongs-Aree, C., Esguerra, E. B., & Kanlayanarat, S. (2011). Physiological changes and cell wall degradation in papaya fruits cv. 'Kaek Dum' and 'Red Maradol' treated with 1- methylcyclopropene. *International Food Research Journal*, 4(18), 1251-1259.
- Liu, M., Pirrello, J., Chervin, C., Roustan, J.-P., & Bouzayen, M. (2015). Ethylene Control of Fruit Ripening: Revisiting the Complex Network of Transcriptional Regulation. *Plant Physiology*, 169, 2380-2390.
- Nath, P., Bouzayen, M., Mattoo, A. K., & Pech, J. C. (2014). *Fruit Ripening: Physiology, Signalling and Genomics*. United Kingdom: CPI Group (UK) Ltd.
- Noiwan, D., Kumpoun, W., Whangcha, K., Saengnil, K., & Uthaibutra, J. (2006). Effect of Coating Materials on Storage Life of Mango cv. Mahajanaka. *Agricultural Science Journal*, 5(37), 132-135.
- R. Quinlan, M. K. (1993). C4.5, Programs for Machine Learning. *San Mateo, CA*.

RapidMiner. (2017). *Operator Reference Guide*. Retrieved December 14, 2017, from RapidMiner Documentation:

[https://docs.rapidminer.com/latest/studio/operators/validation/cross\\_validation.html](https://docs.rapidminer.com/latest/studio/operators/validation/cross_validation.html)

Subhadrabandhu, S., & Ketsa, S. (2001). *Durian: King of Tropical Fruit*. CABI Publishing.

Wills, R., Harris, D., Spohr, L., & Golding, J. (2014). Reduction of energy usage during storage and transport of bananas by management of exogenous ethylene levels. *Postharvest Biology and Technology*, 89, 7-10.





เกียรติบัตรในการแสดงการเข้าร่วมนำเสนอผลงานภาคบรรยายในการประชุมวิชาการระดับชาติ  
มหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ 9



เกียรติบัตรฉบับนี้ ให้ไว้เพื่อแสดงว่า

อุดมศักดิ์ เพียรไพโรจน์ วุฒิพงษ์ ชินศรี

ได้เข้าร่วมนำเสนอผลงาน ภาคบรรยาย ในสาขาการเกษตร อาหาร และสิ่งแวดล้อม  
เรื่อง “การวิเคราะห์ระดับการสุขของทุเรียนจากการแพร่กระจายของก๊าซเอทิลีนโดยใช้เทคนิคการทำเหมืองข้อมูล”

ในการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏกลุ่มศรีอยุธยา ครั้งที่ ๙  
วันที่ ๑๘ - ๑๙ ตุลาคม ๒๕๖๑ ณ มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์  
ให้ไว้ ณ วันที่ ๑๙ ตุลาคม ๒๕๖๑

(อาจารย์ ดร.สุพจน์ ทรายแก้ว)

รักษาราชการแทน

อธิการบดีมหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์

มหาวิทยาลัยราชภัฏรังสิต Rangsit University



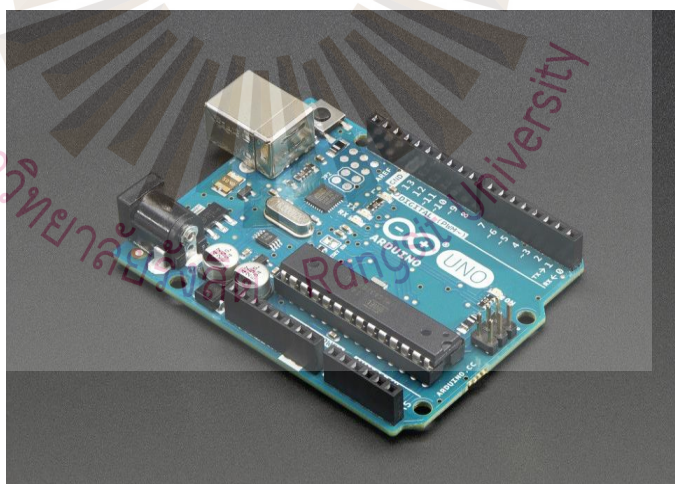
## อุปกรณ์ที่ใช้ในการพัฒนาเครื่องมือสำหรับทดสอบระดับความสูงของทุเรียน

1. เซนเซอร์ที่ใช้ในการตรวจจับก๊าซต่างๆ ในตระกูล MQ Gas Sensor



ภาพแสดง MQ Gas Sensor

2. ไมโครคอนโทรลเลอร์บอร์ดที่ใช้สำหรับประมวลผลต่างๆ ของปริมาณก๊าซที่ได้รับแต่ละชนิด



ภาพแสดงบอร์ดอาดูโน้



3. พัฒนเป็นตัวดึงเอาอากาศรอบวัตถุเป้าหมายไปส่งต่อที่จุดรับกลิ่นหรือจุดตรวจวัดก๊าซเอทีลิน



ภาพแสดงพัดลมดูดอากาศ

4. เมื่อนำเอาอุปกรณ์ต่างๆ มาติดตั้งเข้าระบบแล้ว ผู้วิจัยจึงเขียนโปรแกรมเพื่อทำการบังคับให้อุปกรณ์ต่างๆ ทำงานตามรูปแบบที่ต้องการ และนำทุเรียนมาใส่เพื่อทำการวัดปริมาณของก๊าซ



ภาพแสดงกล่องอุปกรณ์วัดระดับความสุกของทุเรียน

5. เมื่อเครื่องมือทำงานจะส่งผลปริมาณก๊าซที่ได้รับมาส่งที่บอร์ดเพื่อโชว์ระดับความเข้มข้นของก๊าซ



ภาพแสดงปริมาณความเข้มข้นของก๊าซเอทีลิน

## ประวัติผู้วิจัย

ชื่อ	อุดมศักดิ์ เพียรไพโรจน์
วัน เดือน ปีเกิด	21 มิถุนายน 2523
สถานที่เกิด	จังหวัดราชบุรี ประเทศไทย
ประวัติการศึกษา	มหาวิทยาลัยราชภัฏธนบุรี ปริญญาบริหารธุรกิจบัณฑิต สาขาวิชาการจัดการ, 2554 มหาวิทยาลัยราชภัฏสวนสุนันทา ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์, 2558 มหาวิทยาลัยรังสิต ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ, 2561
ทุนการศึกษา	ทุนการศึกษาเรียนดี ดร.อาทิตย์ อุไรรัตน์, 2559
รับรางวัล	STARTUP THAILAND ปี 2560
ที่อยู่ปัจจุบัน	1/217 ตำบลมหาสวัสดิ์ อำเภอบางกรวย จังหวัดนนทบุรี 11130
สถานที่ทำงาน	สำนักงานปลัดบัญชีทหารเรือ กองบัญชาการกองทัพเรือ
ตำแหน่งปัจจุบัน	ประจำแผนกควบคุมภายใน สำนักงานประมาณ

