การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล

วิทยานิพนธ์

ของ

ภัครพล อาจอาษา

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ตุลาคม 2564

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล

วิทยานิพนธ์

ของ

ภัครพล อาจอาษา

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ตุลาคม 2564

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

The analysis of water quality using clustering techniques

Phukkaraphon Ardarsa

A Thesis Submitted in Partial Fulfillment of Requirements

for Master of Science (Information Technology)

October 2021

Copyright of Mahasarakham University

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายภัครพล อาจอาษา แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

....ประธานกรรมการ
(รศ. ดร. สิทธิชัย บุษหมั่น)

....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผศ. ดร. โอฬาริก สุรินต๊ะ)

....กรรมการ
(ผศ. ดร. แกมกาญจน์ สมประเสริฐศรี)

....กรรมการ
(ผศ. ดร. สาธิต แสงประดิษฐ์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

....
(ผศ. ศศิธร แก้วมั่น)
คณบดีคณะวิทยาการสารสนเทศ

....
(รศ. ดร. กริสน์ ชัยมูล)
คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล
ผู้วิจัย ภัครพล อาจอาษา
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. โอฬาริก สุรินต๊ะ
ปริญญา วิทยาศาสตรมหาบัณฑิต **สาขาวิชา** เทคโนโลยีสารสนเทศ
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2564

บทคัดย่อ

การจัดการทรัพยากรน้ำเป็นหนึ่งในปัญหาที่ทั่วโลกกำลังเผชิญอยู่ในปัจจุบัน เช่น สภาวะโลกร้อน พื้นที่แห้งแล้ง และสารเคมีที่เป็นพิษในน้ำ จำเป็นต้องจัดการกับการจัดการทรัพยากรน้ำอย่างเร่งด่วน งานวิจัยฉบับนี้มุ่งเน้นในการศึกษาการวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล ของเขื่อนลำปะทาว จังหวัดชัยภูมิ โดยทำการเก็บข้อมูลด้วยทุ่นลอยน้ำที่ติดตั้งเซนเซอร์ตรวจวัดคุณภาพน้ำ จำนวน 5 พารามิเตอร์ประกอบด้วยเซนเซอร์วัดค่าปริมาณออกซิเจนที่ละลายในน้ำ ค่าอุณภูมิของน้ำ ค่าความเป็นกรดเป็นเบสของน้ำ ค่าความขุ่นของน้ำ และค่าการนำไฟฟ้าของน้ำ เก็บรวบรวมข้อมูลคุณภาพน้ำในช่วงเดือน มกราคม-มีนาคม พ.ศ. 2564 จำนวนทั้งสิ้น 13,608 ชุดข้อมูล รอบบริเวณเขื่อนจำนวน 5 จุด ผลจากการพิจารณาหาจำนวนกลุ่มที่ดีที่สุดได้จำนวน 5 กลุ่ม โดยวิธีการ PCA+K-Means ซึ่งพิจารณาจากวิธีการ Elbow Method และ Silhouette Score ที่ได้ค่าอยู่ที่ 0.7015 ทั้งนี้ยังได้ทำการหาจำนวนกลุ่มของคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data และ Fitzroy River Data ที่เป็นชุดข้อมูล Open Data โดยผลการหาจำนวนกลุ่มที่ดีที่สุดได้จำนวนกลุ่มเท่ากับ 7 กลุ่ม และ 3 กลุ่ม ตามลำดับ

คำสำคัญ : คุณภาพน้ำ, การจัดกลุ่มข้อมูล

TITLE	The analysis of water quality using clustering techniques		
AUTHOR	Phukkaraphon Ardarsa		
ADVISORS	Assistant Professor Olarik Surinta , Ph.D.		
DEGREE	Master of Science	**MAJOR**	Information Technology
UNIVERSITY	Mahasarakham University	**YEAR**	2021

ABSTRACT

Water resource management is one of the biggest challenges that are being faced, such as a warming climate, arid land, and toxic chemicals in the water. It is essential to deal with water resource management urgently In this research mainly focus on monitoring the water quality in the Lam Pa Thao dam Chaiyaphum Province. Data were collected by using a floating buoy with 5 parameter water quality sensor consisting of dissolved oxygen, temperature, pH, total dissolved solids, and electric conductivity. collected the water quality data from January - March 2021, including 13,608 instances around the dam area of 5 points. The best five cluster were determined by PCA+K-Means based on the Elbow Method and Silhouette Score of 0.7015. Furthermore,The water quality clusters were quantified from the Charles river and Fitzroy river datasets.The results of finding the best number of cluster were equal to 7 cluster and 3 cluster respectively.

Keyword : Water Quality, Clustering

กิตติกรรมประกาศ

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.โอฬาริก สุรินต๊ะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่คอยเสียสละเวลาอันมีค่าให้คำปรึกษา คำแนะนำ ชี้แนะข้อบกพร่อง ช่วยแก้ไขปัญหาและติดตามความความก้าวหน้าของงานวิจัยด้วยความเอาใจใส่เป็นอย่างดี ทั้งยังฝึกฝนให้ผู้วิจัยมีทักษะทางด้านความคิดด้านการอ่าน การเขียนและการนำเสนอผลงานทางวิชาการ ขอขอบพระคุณอาจารย์ทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้วิจัยด้วยจิตเมตตา ขอขอบพระคุณ รองศาสตราจารย์ ดร.สิทธิชัย บุษหมั่น ประธานกรรมการสอบ ผู้ช่วยศาสตราจารย์ ดร.แกมกาญจน์ สมประเสริฐศรี และผู้ช่วยศาสตราจารย์. ดร.สาธิต แสงประดิษฐ์ กรรมการสอบที่ให้ความรู้และคำแนะนำทำให้วิทยานิพนธ์เล่มนี้มีความสมบูรณ์และสำเร็จลุล่วงไปด้วยดี ขอขอบพระคุณ บิดามารดา ภรรยา เพื่อน ๆ ทุกท่าน ที่คอยสนับสนุนและเป็นกำลังใจให้ด้วยดีเสมอมา จนทำให้วิทยานิพนธ์นี้ได้สำเร็จไปด้วยดี คุณค่าและประโยชน์อันพึงได้มาจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแด่ผู้มีพระคุณทุกท่าน

ภัครพล อาจอาษา

พหุนํ ปณฺ ฑิโต ชีเว

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....ง
บทคัดย่อภาษาอังกฤษ.....จ
กิตติกรรมประกาศ.....ฉ
สารบัญ.....ช
สารบัญตาราง.....ญ
สารบัญภาพ.....ฏ
บทที่ 1 บทนำ.....1
1.1 หลักการและเหตุผล.....1
1.2 วัตถุประสงค์ของการวิจัย.....3
1.3 ความสำคัญของการวิจัย.....3
1.4 ขอบเขตงานวิจัย.....4
1.5 นิยามศัพท์เฉพาะ.....4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....6
2.1 ทฤษฎีที่เกี่ยวข้อง.....6
2.1.1 เขื่อนลำปะทาว.....6
2.1.2 คุณภาพน้ำที่เหมาะสมกับการเลี้ยงปลา.....8
2.1.3 อินเทอร์เน็ตของสรรพสิ่ง.....8
2.1.4 การเรียนรู้แบบไม่มีผู้สอน.....9
2.1.4.1 การจัดกลุ่ม (Clustering).....9
2.1.4.1.1 การจัดกลุ่มแบบเคมีน (k-mean).....11
2.1.4.1.2 การจัดกลุ่มแบบ Mean Shift.....11

พหุนาม ศิโต ชีเว

2.1.4.1.3 การจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian Mixture Model) ... 12
2.1.5 วิธีการวัดประสิทธิภาพของการจัดกลุ่ม (Evaluation of Clustering) ... 14
2.1.5.1 การวัดประสิทธิภาพของการจัดกลุ่ม K-Means ... 14
2.1.5.1.1 Elbow Method ... 14
2.1.5.1.2 Silhouette Score ... 15
2.1.5.2 การวัดประสิทธิภาพของการจัดกลุ่ม Gaussian mixture model ... 16
2.1.5.2.1 AIC-Akaike information criterion ... 16
2.1.5.2.1 BIC- Bayesian information criterion ... 16
2.1.6 Principal Component Analysis (PCA) ... 17
2.2 งานวิจัยที่เกี่ยวข้อง ... 17
บทที่ 3 วิธีการดำเนินการวิจัย ... 26
3.1 ข้อมูลที่ใช้ในการทดลอง ... 26
3.2 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล ... 29
3.3 การเก็บรวบรวมข้อมูล ... 32
3.3.1 ระยะเวลาในการเก็บรวบรวมข้อมูล ... 32
3.3.2 พื้นที่ในการเก็บข้อมูลคุณภาพน้ำ ... 32
3.4 การสร้างแบบจำลอง ... 33
3.5 การวิเคราะห์ข้อมูลและการประเมินประสิทธิภาพของแบบจำลอง ... 33
บทที่ 4 ผลการดำเนินการวิจัย ... 35
4.1 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว ... 35
4.1.1 พารามิเตอร์ค่าออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen concentration) ... 35
4.1.2 พารามิเตอร์ค่าอุณหภูมิ (Temperature) ... 35
4.1.3 พารามิเตอร์ค่า PH ... 36

4.1.4 พารามิเตอร์ค่าความขุ่นของน้ำ (Turbidity)... 36

4.1.5 พารามิเตอร์ค่าการนำไฟฟ้าของน้ำ(Electric Conductivity).................................... 37

4.2 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำ Charles River Buoy Data 42

4.3 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำ Fitzroy River Data 47

บทที่ 5 สรุปผลและอภิปรายผล... 57

5.1 สรุปผล.. 57

5.2 อภิปรายผล 58

5.3 ข้อเสนอแนะ 59

บรรณานุกรม.. 60

ภาคผนวก.. 68

ประวัติผู้เขียน.. 77

สารบัญตาราง

ตารางที่ 3.1 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว 27

ตารางที่ 3.2 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดของ Charles River Buoy Data 28

ตารางที่ 3.3 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดของ Fitzron river data 28

ตารางที่ 4.1 ผลการหาจำนวนกลุ่มจากชุดข้อมูลทั้ง 3 ชุดข้อมูล 52

ตารางที่ 4.2 ตารางเกณฑ์คุณภาพของน้ำ เพื่อวิเคราะห์กลุ่มคุณภาพน้ำ 53

ตารางที่ 4.3 ผลการเปรียบเทียบค่า Sum of Squared Error และ Silhouette Score สำหรับวิธีการ K-Mean และ PCA+ K-Mean 55

ตารางที่ ข1 ชุดข้อมูลที่ใช้ในการทดลอง 76

สารบัญภาพ

ภาพประกอบที่ 2.1 พื้นที่เขื่อนลำปะทาว 7

ภาพประกอบที่ 2.2 การจัดกลุ่มข้อมูล 10

ภาพประกอบที่ 2.3 จุดหักศอกของ Elbow Method สำหรับการหาจำนวนกลุ่มที่เหมาะสม 14

ภาพประกอบที่ 2.4 การหาจำนวนกลุ่มโดยพิจารณาจากค่า Silhouette Score 16

ภาพประกอบที่ 3.1 ผังการดำเนินงาน 26

ภาพประกอบที่ 3.2 เซนเซอร์วัดปริมาณออกซิเจนที่ละลายในน้ำ 29

ภาพประกอบที่ 3.3 เซนเซอร์วัดค่าอุณภูมิของน้ำ 30

ภาพประกอบที่ 3.4 เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ 30

ภาพประกอบที่ 3.5 เซนเซอร์วัดค่าความขุ่นของน้ำ 30

ภาพประกอบที่ 3.6 เซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ 31

ภาพประกอบที่ 3.7 ไดอะแกรมของอุปกรณ์ที่ใช้ในการจัดเก็บข้อมูล 31

ภาพประกอบที่ 3.8 ทุ่นอุปกรณ์ที่ใช้สำหรับเก็บข้อมูลจากเขื่อนลำปะทาวและส่วนประกอบอุปกรณ์เซนเซอร์สำหรับจัดเก็บข้อมูลคุณภาพน้ำ 32

ภาพประกอบที่ 3.9 จุดในการจัดเก็บคุณภาพน้ำบริเวณพื้นที่เขื่อนลำปะทาว 32

ภาพประกอบที่ 4.1 กราฟแสดงข้อมูลค่าออกซิเจนที่ละลายในน้ำ 35

ภาพประกอบที่ 4.2 กราฟแสดงข้อมูลค่าอุณหภูมิ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว 36

ภาพประกอบที่ 4.3 กราฟแสดงข้อมูลค่า PH จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว 36

ภาพประกอบที่ 4.4 กราฟแสดงข้อมูลค่าความขุ่นของน้ำ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว 37

ภาพประกอบที่ 4.5 กราฟแสดงข้อมูลค่าการนำไฟฟ้าของน้ำ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว 37

ภาพประกอบที่ 4.6 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-Means กับชุดข้อมูลของเขื่อนลำปะทาว . 38

ภาพประกอบที่ 4.7 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูลเขื่อนลำปะทาว 39

ภาพประกอบที่ 4.8 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-Means กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score 40

ภาพประกอบที่ 4.9 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า AIC b) ค่า BIC 41

ภาพประกอบที่ 4.10 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า AIC b) ค่า BIC..... 42

ภาพประกอบที่ 4.11 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-means กับชุดข้อมูล Charles River Buoy Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score..... 43

ภาพประกอบที่ 4.12 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูล Charles River Buoy Data 44

ภาพประกอบที่ 4.13 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-means กับชุดข้อมูล Charles River Buoy Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score..... 45

ภาพประกอบที่ 4.14 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูล Charles River Buoy Data a) ค่า AIC b) ค่า BIC 46

ภาพประกอบที่ 4.15 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูล Charles River Buoy Data a) ค่า AIC b) ค่า BIC 46

ภาพประกอบที่ 4.16 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-Means กับชุดข้อมูล Fitzroy River Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score 48

ภาพประกอบที่ 4.17 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูล Fitzroy River Data 48

ภาพประกอบที่ 4.18 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-means กับชุดข้อมูล Fitzroy River Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score..... 49

ภาพประกอบที่ 4.19 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูล Fitzroy River Data a) ค่า AIC b) ค่า BIC 50

ภาพประกอบที่ 4.20 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูล Fitzroy River Data Data a) ค่า AIC b) ค่า BIC 51

ภาพประกอบที่ 4.21 ผลการจัดกลุ่มคุณภาพน้ำของเขื่อนลำปะทาว 5 กลุ่ม..... 53

ภาพประกอบที่ ก1 บทความที่ได้รับการตีพิมพ์..... 70

ภาพประกอบที่ ก2 บทความที่ได้รับการตีพิมพ์..... 71

ภาพประกอบที่ ก3 บทความที่ได้รับการตีพิมพ์..... 72

ภาพประกอบที่ ก4 บทความที่ได้รับการตีพิมพ์ 73

ภาพประกอบที่ ก5 บทความที่ได้รับการตีพิมพ์ 74

พหูน ปณ ฑิโต ชีเว

บทที่ 1
บทนำ

1.1 หลักการและเหตุผล

การพัฒนาด้านวิทยาศาสตร์ เทคโนโลยีประเทศไทย จำเป็นอย่างยิ่งที่จะต้องปรับรูปแบบการดำเนินงานโดยมุ่งเน้นความร่วมมือระหว่างภาครัฐและเอกชนในการพัฒนา เพื่อยกระดับศักยภาพการผลิตภาคเกษตร รวมถึงการพัฒนานวัตกรรมเพื่อยกระดับคุณภาพสังคมและการดำรงชีวิตของประชาชน โดยอาศัยความรู้ทางการเกษตร ผนวกกับเทคโนโลยีและนวัตกรรมด้านคอมพิวเตอร์ อิเล็กทรอนิกส์และเซ็นเซอร์เข้าช่วยเพื่อเพิ่มความแม่นยำของระบบจัดการผลิต ปัจจุบันการเกษตรของไทยเข้ามามีบทบาทสำคัญต่อระบบเศรษฐกิจของประเทศ ประชากรส่วนใหญ่ของประเทศประกอบอาชีพการเกษตร[1] หนึ่งในอาชีพการเกษตรของประเทศคือการเพาะเลี้ยงปลานิลซึ่งเป็นปลาที่มีความสำคัญทางเศรษฐกิจและเป็นปลาน้ำจืดที่มีผลผลิตเป็นอันดับหนึ่งของประเทศไทย[2] กลุ่มเกษตรกรเลี้ยงปลาในกระชัง เขื่อนลำปะทาว จังหวัดชัยภูมิ จัดตั้งเมื่อปี 2542 ปัจจุบันมีสมาชิกมากกว่า 80 คน ในรอบหลายปีที่ผ่านมากลุ่มได้ประสบปัญหาปลาตายเนื่องจากน้ำเน่าเสีย และสภาพอากาศแปรปรวนส่งผลต่อคุณภาพน้ำ ทำให้ปลาเกิดอาการปลาน๊อคน้ำ เกิดความเสียหายกับเกษตรกรเป็นจำนวนมาก Sryasak และคณะ[3] กล่าวไว้ว่า ฤดูกาล อากาศ และสภาพภูมิอากาศมีผลต่อการเพาะเลี้ยงสัตว์น้ำ โดยที่อิทธิพลของสภาพอากาศและฤดูกาลจะส่งผลกระทบต่อคุณภาพของน้ำ เช่น อุณภูมิของน้ำ ปริมาณออกซิเจนละลายน้ำ เป็นต้น ซึ่งปัจจัยเหล่านี้จะมีผลต่อการเจริญเติบโต อัตราการรอด และความเสี่ยงในการเกิดโรคของสัตว์น้ำ

อินเทอร์เน็ตทุกสรรพสิ่ง (IoT) เป็นการเชื่อมต่ออุปกรณ์ทางกายภาพ ผ่านทางอินเทอร์เน็ต ถูกเสนอครั้งแรกโดย Kevin Ashton ในปี 1999 ภายใต้โครงการที่ชื่อ “Auto-ID” ที่มหาวิทยาลัย Massachusetts Institute of Technology จากเทคโนโลยี RFID ที่จะทำให้เป็นมาตรฐาน กลายมาเป็นแนวคิดของการทำให้อุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ สื่อสารกันได้โดยอาศัยตัว Sensor ในการสื่อสารถึงกันและสามารถเชื่อมต่อกันได้ ซึ่งเป็นพื้นฐานการขยายตัวของการบริการด้วยอินเทอร์เน็ต [4] IoT ถือว่าเป็นแนวทางใหม่ของการวิจัยเชิงนวัตกรรมด้านการเกษตร เพื่อนำไปใช้สำหรับการเกษตรด้านต่าง ๆ มุ่งเน้นไปที่คุณลักษณะเฉพาะ ของอุปกรณ์และเทคโนโลยีการสื่อสารไร้สาย สามารถสั่งการหรือควบคุมผ่านระบบเครือข่ายอินเทอร์เน็ต ในการใช้งานด้านการเกษตรหรือฟาร์มอัจฉริยะโดยมีการวิเคราะห์ การตรวจสอบอย่างละเอียดและครอบคลุม ใช้ในแก้ปัญหาที่เกิดขึ้นเพื่อวางแผนในอนาคต ปัจจุบันมีจำนวนแอพพลิเคชั่นที่ใช้ประโยชน์จาก IoT เพิ่มขึ้นเป็นจำนวนมาก [5] [6] เห็นได้ว่าในปัจจุบันได้มีการนำเอาเทคโนโลยี IoT มาใช้ในงานด้านการเกษตรเพื่อให้เป็น

การเกษตรแบบอัจฉริยะ ในขณะที่ Muangprathub และคณะ[7] ได้นำเสนอการเกษตรแบบอัตโนมัติบนพื้นฐานของเครือข่ายเซ็นเซอร์ไร้สาย โดยนำข้อมูลจาก อุปกรณ์ IoT คือ อุณภูมิ ความชื้น และความชื้นในดินเพื่อทำนายผลผลิต และนำข้อมูลที่ได้จากเซนเซอร์มาวิเคราะห์ข้อมูลด้วยการทำ Data Mining ให้มีประสิทธิภาพมากขึ้น Mehra และคณะ[8] ได้นำเสนอการพัฒนาเทคนิคการทำฟาร์มสมัยใหม่โดยใช้อัลกอริทึม machine learning และใช้ IoT ช่วยในการสื่อสารระหว่างเครื่องสู่เครื่อง โดยใช้ Arduino ที่เชื่อมต่อกับเซ็นเซอร์ต่าง ๆส่งไปยังคลาวด์ เพื่อการจัดเก็บและการแสดงผลข้อมูล จากวิวัฒนาการของ IoT และการเพิ่มจำนวนของเซ็นเซอร์อย่างต่อเนื่อง ทำให้เกิดความท้าทายเกี่ยวกับข้อมูลที่กำลังขยายตัว การวิเคราะห์ข้อมูลเป็นพื้นฐานของการใช้งาน IoT และเป็นเครื่องมือที่สำคัญ เพื่อตอบสนองวัตถุประสงค์ และสนับสนุนการตัดสินใจให้ตรงตามความต้องการของผู้ใช้งานหรือองค์กรในการวิเคราะห์หรือประมวลผลต่าง ๆ[9] การส่งเสริมและพัฒนาการเกษตรอย่างมีประสิทธิภาพโดยมีการนำเทคโนโลยี IoT มาใช้กับกระบวนการผลิตทางการเกษตรให้มีความทันสมัย โดยมีการวิเคราะห์ข้อมูลเพื่อค้นหาผลลัพธ์ที่ต้องการในการพัฒนาให้เป็นเกษตรอัจฉริยะผ่านฟังก์ชั่นการตรวจจับ การประมวลผล การสื่อสารข้อมูล และการแจ้งเตือนของ IoT ได้อย่างถูกต้องและแม่นยำ[10] นอกจากการนำ IoT ไปใช้ในการเพาะปลูกพืชแล้วยังสามารถนำไปประยุกต์ใช้ในการเกษตรสำหรับการเพาะเลี้ยงสัตว์น้ำได้อีกด้วย

การเจริญเติบโตและการอยู่รอดของการเพาะเลี้ยงสัตว์น้ำ ขึ้นอยู่กับปัจจัยหลักอย่างหนึ่งคือคุณภาพน้ำ ดังนั้นจึงจำเป็นอย่างยิ่งที่จะต้องทราบถึงการเปลี่ยนแปลงคุณภาพน้ำเพื่อให้เกษตรกรได้เตรียมพร้อมในการหาวิธีป้องกันได้ทันท่วงที ระบบตรวจสอบคุณภาพน้ำด้วยเซ็นเซอร์ ที่มีส่วนประกอบหลักของ Wireless Sensor Network ประกอบด้วยไมโครคอนโทรลเลอร์สำหรับการประมวลผล ทำหน้าที่สื่อสารระหว่างเซ็นเซอร์ ทำให้สามารถเข้าถึงข้อมูลแบบเรียลไทม์ได้ โดยใช้เทคโนโลยี IoT [11] Ahmed และคณะ[12] ได้ทำการวิจัยเรื่อง การเรียนรู้เครื่องสำหรับการทำนายคุณภาพน้ำ โดยมีวัตถุประสงค์เพื่อประเมินความสัมพันธ์ระหว่างพารามิเตอร์ของคุณภาพน้ำโดยใช้โครงข่ายประสาทเทียม(ANN) พยากรณ์คุณภาพน้ำ และเพื่อตรวจสอบประสิทธิภาพของแบบจำลองสำหรับตำแหน่งเชิงพื้นที่ของแม่น้ำยะโฮร์ในประเทศมาเลเซีย ในขณะที่ Hajigholizadeh และคณะ [13] ได้ศึกษาการวิเคราะห์กลุ่ม(CA) และการจำแนก (DA) เพื่อประเมินคุณภาพน้ำและการวิเคราะห์การเปลี่ยนแปลง เพื่อควบคุมและจัดการมลพิษรวมถึงการป้องกันคุณภาพของน้ำในแม่น้ำ Parra และคณะ [14] ได้ทำการออกแบบและพัฒนาเซ็นเซอร์ตรวจจับความขุ่นของน้ำราคาประหยัดสำหรับการตรวจสอบคุณภาพน้ำในฟาร์มเลี้ยงปลาโดยใช้เซนเซอร์ราคาประหยัดที่สามารถเชื่อมต่อกับโหนดที่แตกต่างกันได้ เพื่อใช้แยกแยะความขุ่นของน้ำในฟาร์มเลี้ยงปลา Carrasco และคณะ[15]ทำการประเมินคุณภาพน้ำด้วยสถิติแบบหลายตัวแปรโดยใช้วิธี HJ-Biplot โดยใช้ตัวแปรในการวิจัย อาทิเช่น อุณภูมิ ความเป็นกรดเป็นเบส ความขุ่นของน้ำ ไนเตรตและค่าออกซิเจนที่ละลายในน้ำ ที่ได้จากกลุ่ม

น้ำปานามา และทำการวิเคราะห์กลุ่มด้วยอัลกอลิทึมเคมีน Gao,Xiao และChen [16] ได้นำเสนองานวิจัยและกล่าวไว้ว่า คุณภาพน้ำและความปลอดภัยของผลผลิตจากสัตว์น้ำที่เลี้ยงมีความสำคัญอย่างมาก และได้ทำการพัฒนา ระบบเพาะเลี้ยงปลาอัจฉริยะ ด้วย IoT สามารถช่วยเพิ่มรายได้ให้กับเกษตรกร อีกทั้งยังนำเสนอชุดพยากรณ์คุณภาพน้ำแบบอัจฉริยะด้วยการทำนายค่าพารามิเตอร์ที่ได้จากเซนเซอร์วัดค่าปริมาณออกซิเจนที่ละลายน้ำ เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ เซนเซอร์วัดค่าอุณภูมิของน้ำ เซนเซอร์วัดค่าความขุ่น และเซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ ซึ่งในงานวิจัยนี้ผู้วิจัยจะอ้างอิงค่าพารามิเตอร์ในการเก็บข้อมูลคุณภาพน้ำ จำนวน 5 พารามิเตอร์ในการเก็บข้อมูลคุณภาพเพื่อเป็นชุดข้อมูลในการทดลอง โดยอ้างอิงจากพารามิเตอร์ชุดพยากรณ์คุณภาพน้ำแบบอัจฉริยะของ Gao และคณะ

ในการวิจัยครั้งนี้ ผู้วิจัยทำการศึกษาวิจัยการเก็บข้อมูลคุณภาพน้ำ ในเขื่อนลำปะทาว จังหวัดชัยภูมิ ด้วย IoT โดยใช้ Nodemcu เชื่อมต่อกับเซนเซอร์วัดค่าจำนวน 5 ตัว คือเซนเซอร์วัดค่าปริมาณออกซิเจนที่ละลายน้ำ เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ เซนเซอร์วัดค่าอุณภูมิของน้ำ เซนเซอร์วัดค่าความขุ่นของน้ำ เซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ ส่งข้อมูลจากอุปกรณ์ IoT ไปยังอุปกรณ์บันทึกข้อมูล ซึ่งก่อนหน้านี้ยังไม่มีการหาจำนวนกลุ่มคุณภาพน้ำของเขื่อนลำปะทาว ผู้วิจัยจึงมีแนวคิดที่จะหาจำนวนกลุ่มคุณภาพน้ำของเขื่อนลำปะทาวเพื่อนำไปประยุกต์ใช้ในแก้ปัญหาคุณภาพน้ำสำหรับการเพาะเลี้ยงสัตว์น้ำ และปัญหาปลาน็อคน้ำในอนาคต โดยทำการสร้างโมเดลเพื่อวิเคราะห์คุณภาพน้ำโดยใช้เทคนิคการจัดกลุ่ม (Clustering) ในการเปรียบเทียบ 3 เทคนิค คือ เทคนิคการจัดกลุ่มแบบเคมีน (K-Mean) เทคนิคการจัดกลุ่มแบบมีนชิฟ (Mean Shift) เทคนิคการจัดกลุ่มแบบจำลองผสมเกาส์เชียน (Gaussian Mixture Model) และนำวิธี การวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธี

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อจัดกลุ่มคุณภาพน้ำในเขื่อนลำปะทาว
2. เพื่อเปรียบเทียบผลการจัดกลุ่มคุณภาพน้ำด้วยเทคนิคการจัดกลุ่มร่วมกับการวิเคราะห์องค์ประกอบหลัก

1.3 ความสำคัญของการวิจัย

1. เพื่อจัดกลุ่มคุณภาพน้ำในเขื่อนลำปะทาว
2. เพื่อเปรียบเทียบผลการจัดกลุ่มคุณภาพน้ำด้วยเทคนิคการจัดกลุ่มร่วมกับการวิเคราะห์องค์ประกอบหลัก

1.4 ขอบเขตงานวิจัย

1. พื้นที่ในการเก็บข้อมูลคุณภาพน้ำ

พื้นที่ในการเก็บข้อมูลคุณภาพน้ำในการวิจัย เขื่อนลำปะทาว อำเภอแก้งคร้อ จังหวัดชัยภูมิ

2. คุณภาพน้ำที่ใช้ในการเก็บรวบรวม

ค่าของคุณภาพน้ำได้จากการเก็บข้อมูลด้วย IoT จำนวน 5 ค่าประกอบด้วยค่าปริมาณออกซิเจนที่ละลายในน้ำ ค่าอุณภูมิของน้ำ ค่าความเป็นกรดเป็นเบสของน้ำ ค่าความขุ่นของน้ำ และค่าการนำไฟฟ้าของน้ำ

3. ระยะเวลาในการเก็บข้อมูลคุณภาพน้ำ

การเก็บรวบรวมข้อมูลในการศึกษาวิจัยครั้งนี้โดยวิธีการเก็บข้อมูลจากอุปกรณ์เซนเซอร์ที่ใช้ในการตรวจวัดค่า จำนวน 5 ค่า บริเวณที่เกษตรกรเลี้ยงปลานิลในกระชัง และบริเวณริมฝั่งของเขื่อนลำปะทาว ซึ่งได้ทำการเก็บข้อมูลจำนวนทั้งสิ้น 13,608 ชุดข้อมูล ที่จัดเก็บในช่วงเดือนมกราคม-มีนาคม พ.ศ. 2564

4. จำนวนจุดในการเก็บข้อมูลคุณภาพน้ำ

บริเวณพื้นที่ เขื่อนลำปะทาว อำเภอแก้งคร้อ จังหวัดชัยภูมิ จำนวน 5 จุด

5. การวิเคราะห์คุณภาพน้ำ

ในการวิเคราะห์คุณภาพน้ำของการศึกษาวิจัยครั้งนี้จะมีการวิเคราะห์การแบ่งกลุ่มคุณภาพน้ำว่าควรแบ่งเป็นกี่กลุ่ม โดยใช้เทคนิคการจัดกลุ่มแบบเคมีน (K-Mean) เทคนิคการจัดกลุ่มแบบมีนชิฟ (Mean Shift) เทคนิคการจัดกลุ่มแบบจำลองผสมเกาส์เซียน (Gaussian Mixture Model) และนำวิธีการวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธี

1.5 นิยามศัพท์เฉพาะ

1. เขื่อน หมายถึง เขื่อนลำปะทาว อำเภอแก้งคร้อ จังหวัดชัยภูมิ

2. ชุดข้อมูลเขื่อนลำปะทาว หมายถึง ค่าของคุณภาพน้ำได้จากการเก็บข้อมูลด้วย IoT จำนวน 5 ค่าประกอบด้วยค่าปริมาณออกซิเจนที่ละลายในน้ำ ค่าอุณภูมิของน้ำ ค่าความเป็นกรดเป็นเบสของน้ำ ค่าความขุ่นของน้ำ และค่าการนำไฟฟ้าของน้ำ

3. ชุดข้อมูลจาก Charles River Buoy Data หมายถึง ชุดข้อมูลที่ได้มาจากเว็บไซต์ของสำนักงานคุ้มครองสิ่งแวดล้อมแห่งสหรัฐอเมริกา (United States Environmental Protection Agency) ประกอบไปด้วย 8 พารามิเตอร์ ดังนี้ Water Temperature, Specific Conductance,

PH, Dissolved Oxygen Concentration, Dissolved Oxygen % Saturation, Turbidity, Chlorophyll และ Phycoerythrin

4. ชุดข้อมูลจาก Fitzroy River Data หมายถึง ชุดข้อมูลที่ได้มาจากการเผยแพร่โดย Queensland Government ประเทศออสเตรเลีย ที่จัดเก็บในช่วงปี ค.ศ. 1993-2003 ซึ่งประกอบไปด้วย 7 พารามิเตอร์ ประกอบด้วย Dissolved Oxygen concentration, Dissolved Oxygen % saturation, pH, Salinity, Specific Conductance และ Temperature

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยเรื่อง การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล ผู้วิจัยได้ดำเนินการศึกษาค้นคว้าเอกสารและทบทวนทฤษฎีและงานวิจัยที่เกี่ยวข้อง ดังต่อไปนี้

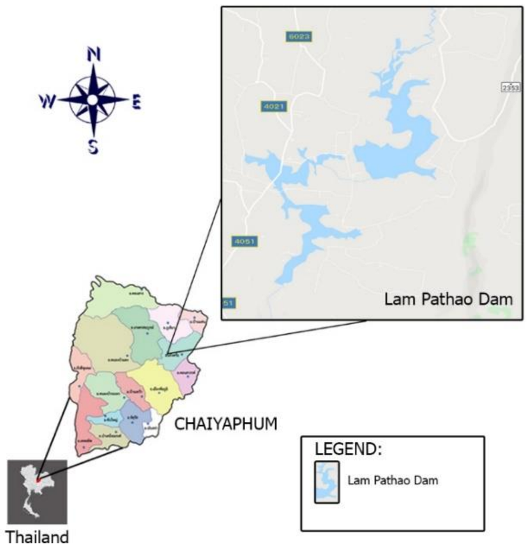
2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 เขื่อนลำปะทาว

เขื่อนลำปะทาวเป็นหนึ่งในเขื่อนที่สำคัญของจังหวัดชัยภูมิ เป็นลำน้ำสาขาของลำน้ำชี โดยมีลักษณะเป็นเขื่อนดิน แบ่งออกเป็น 2 แห่ง คือ เขื่อนลำปะทาวบนและเขื่อนลำปะทาวล่าง ฉัตรชัย โรจน์เรืองไร และคณะ[17] ได้อธิบายลักษณะเขื่อนลำปะทาวไว้ว่า เขื่อนลำปะทาวบนเป็นเขื่อนขนาดใหญ่ ตั้งอยู่บนพื้นที่บ้านท่ากอก ตำบลเก่าย่าดี อำเภอแก้งคร้อ จังหวัดชัยภูมิ ปิดกั้นลำห้วยปะทาว โดยลักษณะเขื่อนมีความสูง 24 เมตร มีความยาว 875 เมตร ตัวอ่างเก็บน้ำมีความจุอยู่ที่ 44 ล้านลูกบาศก์เมตร มีความจุในการใช้งาน 41 ล้านลูกบาศก์เมตร ติดตั้งเครื่องกำเนิดไฟฟ้าพลังงานน้ำ ขนาดกำลังผลิตรวม 4,500 กิโลวัตต์ โดยสามารถผลิตพลังงานไฟฟ้าได้ปริมาณ 18.41 ล้านหน่วย จ่ายไฟฟ้าให้กับพื้นที่อำเภอแก้งคร้อ และอำเภอคอนสวรรค์ อีกทั้งส่งน้ำให้พื้นที่ชลประทาน ประมาณ 15,000 ไร่ ส่วนเขื่อนลำปะทาวล่างเป็นเขื่อนขนาดกลาง ตั้งอยู่ในพื้นที่บ้านท่าหินโงม ตำบลท่าหินโงม อำเภอเมือง จังหวัดชัยภูมิ โดยพื้นที่อ่างเก็บน้ำอยู่ระหว่างรอยต่อระหว่างอำเภอแก้งคร้อ และอำเภอเมืองชัยภูมิ มีความสูง 37.5 เมตร มีความยาว 237 เมตร อ่างเก็บน้ำมีความจุ 14.8 ล้านลูกบาศก์เมตร ส่งน้ำให้พื้นที่ชลประทานประมาณ 12,000 ไร่ และปล่อยน้ำสำหรับการท่องเที่ยวในเขตอุทยานแห่งชาติน้ำตกตาดโตน อีกทั้งยังเก็บกักน้ำสำหรับการผลิตประปาสำหรับอำเภอเมือง จังหวัดชัยภูมิ เกษตรกรในพื้นที่บริเวณรอบเขื่อนจำนวนหนึ่งใช้พื้นที่อ่างเก็บน้ำทั้งตัวเขื่อนบนและตัวเขื่อนล่าง สำหรับการเลี้ยงปลาในกระชัง อาชีพการเลี้ยงปลาในกระชังเกิดจากการรวมกลุ่มของเกษตรกร โดยก่อนหน้านี้สภาพทั่วไปบริเวณรอบเขื่อนเกษตรกรส่วนใหญ่ประกอบอาชีพทำไร่เป็นอาชีพหลัก แต่เนื่องจากสภาพพื้นที่อยู่ติดกับเขื่อน ปัจจุบันจึงหันมาประกอบอาชีพเลี้ยงปลาในกระชัง โดยได้รับการสนับสนุนจากหน่วยงานภาครัฐ ให้มีการใช้ประโยชน์จากแหล่งน้ำภายใต้โครงการ ส่งเสริมการเลี้ยงปลานิลในกระชังของหน่วยงานภาครัฐ ที่ก่อตั้งในปี 2543 โดยภาครัฐให้การส่งเสริมและสนับสนุนปัจจัยการผลิต เช่น กระชังปลา อาหารปลา และให้ความรู้ทางวิชาการ ในระยะแรกมีเกษตรกรเข้าร่วมโครงการจำนวน 20 ราย มีกระชัง 52 กระชัง หลังจากนั้นเมื่อเกษตรกรที่เข้าร่วมโครงการชุดแรกประสบความสำเร็จ จึงมีเกษตรกรรายอื่น ๆ หันมาให้ความสนใจและเข้าร่วมกลุ่ม

เลี้ยงปลาในกระชัง จนกลายเป็นแหล่งน้ำที่มีการเพาะเลี้ยงปลากระชังที่มากที่สุดในจังหวัดชัยภูมิ จนกระทั่งในปี 2548 มีผู้เพาะเลี้ยงปลากระชังจำนวน 228 ราย รวม 699 กระชัง [18] และดำเนินการต่อเนื่องมาจนถึงปัจจุบัน

เขื่อนลำปะทาวอยู่ในการดูแลของหน่วยจัดการต้นน้ำลำปะทาว ภายใต้การกำกับดูแลของส่วนราชการหลัก 4 กระทรวงคือ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม กระทรวงเกษตรและสหกรณ์ กระทรวงพลังงาน และกระทรวงมหาดไทย อยู่ในเขตพื้นที่ของอุทยานแห่งชาติภูแลนคา ลักษณะภูมิประเทศของเขื่อนลำปะทาวเป็นภูเขาสลับซับซ้อน สภาพป่ามีทั้งป่าทึบและป่าโปร่ง ลักษณะโดยทั่วไปเป็นที่ราบสูงมีเทือกเขาล้อมรอบที่ราบเอาไว้ ตอนกลางพื้นที่เป็นหุบเขากว้าง พื้นที่ด้านทิศตะวันตกเฉียงเหนือมีเทือกเขาภูเขียว ภูกลางล้อมรอบ ที่มีระดับความสูงประมาณ 905 เมตร จากระดับน้ำทะเลปาน และทางด้านตะวันออกเฉียงเหนือมีเทือกเขาภูแลนคาล้อมรอบ ที่มีระดับความสูงประมาณ 945 เมตร จากระดับน้ำทะเลปานกลาง บริเวณเขื่อนบนตั้งอยู่ระหว่างละติจูดที่ 16.112244 ลองติจูดที่ 102.075509 และเขื่อนล่างตั้งอยู่ละติจูดที่ 16.033439 ลองติจูดที่ 102.041330 ดังภาพประกอบที่ 2.1 เขื่อนลำปะทาวเริ่มดำเนินการก่อสร้างภายใต้โครงการไฟฟ้าพลังงานน้ำเขื่อนลำปะทาว โดยกรมพัฒนาและส่งเสริมพลังงาน เมื่อเดือนมีนาคม พ.ศ. 2531 และสามารถกักแก็บน้ำได้ในเดือนตุลาคม พ.ศ. 2536 เป็นต้นมา สภาพต้นน้ำของเขื่อนมีความอุดมสมบูรณ์ โดยชาวบ้านบริเวณรอบเขื่อนได้ใช้ประโยชน์ด้านสาธารณูปโภค ด้านการเกษตร ด้านการท่องเที่ยว ประเพณีวัฒนธรรม อุตสาหกรรมในครัวเรือน การประมงพื้นบ้านและประมงอาชีพ (การเลี้ยงปลากระชัง)[19] [20] ซึ่งในการวิจัยนี้ผู้วิจัยมีความสนใจในการวิเคราะห์กลุ่มคุณภาพน้ำบริเวณรอบเขื่อนลำปะทาวน้ำด้วยวิธีการจัดกลุ่มข้อมูล



ภาพประกอบที่ 2.1 พื้นที่เขื่อนลำปะทาว

2.1.2 คุณภาพน้ำที่เหมาะสมกับการเลี้ยงปลา

งานประมง ศูนย์ศึกษาการพัฒนาเขาหินซ้อนอันเนื่องมาจากพระราชดำริ [21] กล่าวไว้ว่าแหล่งน้ำและคุณภาพน้ำเป็นอีกปัจจัยหนึ่ง ที่มีความสำคัญต่อการเจริญเติบโต สุขภาพ การดำรงชีพ การสืบพันธุ์และแพร่พันธุ์ของสัตว์น้ำ เนื่องจากสัตว์น้ำต้องอาศัยน้ำเป็นสื่อกลางในการหายใจ การหาอาหาร การรักษาสมดุล กิจกรรมทางชีวเคมี การใช้อาหาร และการขับถ่ายของเสีย ฉะนั้นน้ำจึงเปรียบเหมือนบ้านของสัตว์น้ำ หากคุณภาพน้ำที่เหมาะสมและดีแล้วสัตว์น้ำก็จะเจริญเติบโต มีสุขภาพและมีคุณภาพที่ดี จึงสามารถจำหน่ายได้ในราคาที่สูง เมื่อมีการเพาะเลี้ยงสัตว์น้ำที่สามารถจัดการควบคุมคุณภาพน้ำได้ ก็จะช่วยให้การเพาะเลี้ยงสัตว์น้ำประสบความสำเร็จได้ อย่างไรก็ตาม คุณภาพน้ำที่ดี จะมีความสัมพันธ์กับชนิดและคุณภาพดิน แหล่งน้ำและคุณภาพน้ำ โดยดัชนีคุณภาพน้ำเพื่อใช้ในการเพาะเลี้ยงสัตวน้ำ แบ่งเป็น 3 ลักษณะดังนี้

1) ลักษณะทางกายภาพ หมายถึงดัชนีคุณภาพน้ำที่ผันแปรอันเกิดจากลักษณะที่สามารถตรวจวัดได้และมีผลกระทบตอสิ่งมีชีวิตในทางตรงหรือทางออม เชน สี ความขุ่น อุณหภูมิ ความนำไฟฟา ปริมาณสารแขวนแขวนลอย เป็นต้น

2) ลักษณะทางเคมีภาพ หมายถึงดัชนีคุณภาพน้ำที่ผันแปรอันเนื่องจากปฏิกิริยาทางเคมีที่สามารถตรวจวัดได้และมีผลกระทบตอสิ่งมีชีวิตในน้ำ ทั้งทางตรงและทางอ้อม เชน ความเป็นกรดเป็นด่าง ความกระด้าง ปริมาณออกซิเจนละลาย ปริมาณคารบอนไดออกไซดอิสระ ไนโตรเจน ฟอสฟอรัส ไฮโดรเจนซัลไฟด์ ความเค็ม โลหะหนัก สารพิษ เป็นต้น

3) ลักษณะทางชีวภาพ หมายถึง คุณภาพน้ำที่ผันแปรเนื่องจากสิ่งมีชีวิตในน้ำอันมีผลกระทบต่อสิ่งมีชีวิตในน้ำ ทั้งทางตรงและออม เชน แพลงกตอนพืชและสัตว แบคทีเรีย พืชน้ำ เชื้อโรค เป็นต้น

ดังนั้น คุณภาพน้ำจึงเป็นปัจจัยสำคัญของการอยู่รอดและการเจริญเติบโตของปลาที่เพาะเลี้ยง หากคุณภาพน้ำเหมาะสมจะทำให้ปลามีสุขภาพดีและมีการเจริญเติบโตอย่างต่อเนื่อง การจัดการคุณภาพน้ำเป็นเรื่องที่ต้องให้ความสำคัญและใส่ใจอย่างสม่ำเสมอ เนื่องจากในการเพาะเลี้ยงปลาให้มีการเติบโตและมีอัตราการรอดสูงนั้นขึ้นอยู่กับคุณภาพของน้ำ[22] [23] [24] ในงานนี้จะใช้การจัดเก็บข้อมูลคุณภาพน้ำจากบริเวณรอบเขื่อนลำปะทาวด้วยเทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง

2.1.3 อินเทอร์เน็ตของสรรพสิ่ง

อินเทอร์เน็ตของสรรพสิ่ง หรือ Internet of Thing (IoT) เป็นการเชื่อมต่ออุปกรณ์ทางกายภาพ ผ่านทางอินเทอร์เน็ต ถูกเสนอครั้งแรกโดย Kevin Ashton ในปี 1999 ภายใต้โครงการที่ชื่อ “Auto-ID” ที่มหาวิทยาลัย Massachusetts Institute of Technology จากเทคโนโลยี RFID ที่จะทำให้เป็นมาตรฐาน กลายมาเป็นแนวคิดของการทำให้อุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ สื่อสารกัน

ได้โดยอาศัยตัว Sensor ในการสื่อสารถึงกันและสามารถเชื่อมต่อกันได้ ซึ่งเป็นพื้นฐานการขยายตัวของการบริการด้วยอินเทอร์เน็ต[25] เชื่อมต่ออุปกรณ์ โดยใช้เทคโนโลยีไร้สาย ช่วยในการถ่ายโอนข้อมูล การสื่อสาร และแบ่งปันข้อมูลได้ทุกที่ทุกเวลาผ่านทางอินเทอร์เน็ต โดยอาศัยการทำงานร่วมกันระหว่างฮาร์ดแวร์และซอฟต์แวร์ โดยมีความมุ่งหมายในการทำงานของอุปกรณ์ได้ทุกที่ทุกเวลา [26] ในงานวิจัยนี้จะเป็นการเก็บข้อมูลคุณภาพน้ำโดยใช้เทคโนโลยี IoT เข้ามาช่วยในการเก็บข้อมูลคุณภาพน้ำ

2.1.4 การเรียนรู้แบบไม่มีผู้สอน

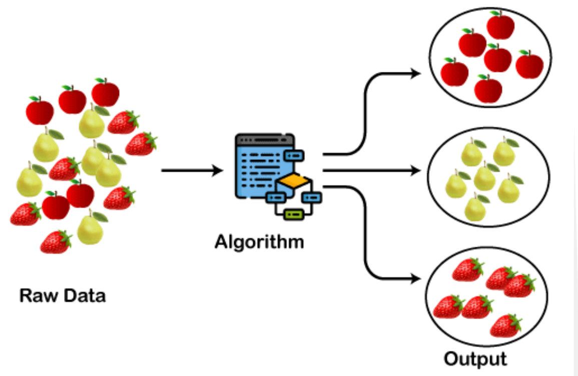
การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นหนึ่งในเทคนิคการเรียนรู้ของเครื่อง โดยเป็นการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบไม่มีผู้สอนจะแตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุ ลาเบลหรือประเภทเอาไว้ก่อน การเรียนรู้แบบไม่มีผู้สอนจะพิจารณาวัตถุเป็นเซตของตัวแปรที่สุ่ม แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูล เป็นการเรียนรู้ที่ไม่มีข้อมูลตัวอย่างที่ใช้บอกว่าข้อมูลนั้นคืออะไร แต่จะเรียนรู้จากการหาความสัมพันธ์จากข้อมูลนำเข้า (Input) โดยพิจารณาจากรูปแบบ (Patterns) หรือโครงสร้างข้อมูล (Data Structure) แล้วนำมาทำการจัดกลุ่ม (Cluster) บนพื้นฐานของความเหมือน (Similarities) และความแตกต่าง (Differences) ระหว่างรูปแบบของข้อมูล [27] โดยจะได้กล่าวถึงในหัวข้อที่ 2.1.4.1

2.1.4.1 การจัดกลุ่ม (Clustering)

การจัดกลุ่ม จัดอยู่ในกระบวนการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) โดยมีกระบวนการจัดการวัตถุหรือค่าต่าง ๆ ให้อยู่ในกลุ่มที่เหมาะสม วัตถุหรือค่าที่อยู่ในกลุ่มเดียวกันจะมีความคล้ายคลึงกัน และมีความแตกต่างกันกับวัตถุในกลุ่มอื่น โดยความเหมือนหรือต่างกันสามารถเปรียบเทียบได้ด้วยความใกล้ชิดกันของวัตถุใด ๆ โดยใช้ระยะทางเป็นตัวชี้วัด [28]

การจัดกลุ่มข้อมูล (Data Clustering) หนึ่งในกระบวนการวิธีการวิเคราะห์ข้อมูล ที่ใช้ในกระบวนการเรียนรู้ของเครื่อง เป็นการเรียนรู้ข้อมูลจากกลุ่มตัวอย่างที่ไม่มีการกำหนดค่าเป้าหมาย (Target) หรือฉลาก (Label) โดยกระบวนการจะทำการแบ่งชุดข้อมูลออกเป็นกลุ่ม (Cluster) นำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน[29] ซึ่งวิธีการของการจัดกลุ่มจะอาศัยความเหมือน (Similarity) หรือ ความใกล้ชิด (Proximity) ดังแสดงในภาพประกอบที่ 2.2 ซึ่งแสดงถึงอัลกอลิทึมของการจัดกลุ่ม โดยจากภาพประกอบที่ 2.2 แสดงให้เห็นถึงการนำเข้าข้อมูลจากผลไม้ ที่เป็นข้อมูลดิบ (Raw Data) ผ่านการจัดกลุ่มด้วยอัลกอลิทึมจัดกลุ่มข้อมูล จากนั้นจะแสดงผลลัพธ์การจัดกลุ่มออกเป็น 3 กลุ่มตามคุณสมบัติที่ใกล้เคียงกันของกลุ่มข้อมูลนำเข้า โดยการจัดกลุ่มข้อมูลสามารถคำนวณได้จากการวัดระยะระหว่างเวกเตอร์ของข้อมูลโดยใช้วิธีการวัดระยะห่าง สามารถวัดได้หลายวิธี เช่น ระยะห่างเชิงยูคลิดยกกำลังสอง (Squared

Euclidean Distance) การวัดระยะแบบแมนฮัตตัน (Manhattan Distance) เป็นต้น กระบวนการแบ่งกลุ่มข้อมูลอาจใช้เป็นขั้นตอนเบื้องต้นของการวิเคราะห์ข้อมูล เพื่อช่วยในการลดขนาดข้อมูล เพื่อแยกเป็นหลาย ๆ กลุ่มและคัดเฉพาะบางกลุ่มเพื่อทำการวิเคราะห์ต่อไป หรือแยกการวิเคราะห์ออกเป็นสำหรับแต่ละกลุ่ม ก่อนที่จะนำไปวิเคราะห์ด้วยวิธีการอื่นต่อไป



ภาพประกอบที่ 2.2 การจัดกลุ่มข้อมูล [30]

ประเภทของการจัดกลุ่ม วิธีการจัดกลุ่มแบ่งออกเป็นแบบกว้างๆ คือแบบ Hard Clustering โดยการจัดกลุ่มแบบนี้ข้อมูลจะอยู่ในกลุ่มเดียวเท่านั้น และแบบ Soft Clustering โดยที่ข้อมูลสามารถอยู่ในอีกกลุ่มหนึ่งได้[31] และยังมีวิธีการอื่น ๆ ซึ่งในการวิจัยนี้จะขอกล่าวถึงวิธีการจัดกลุ่มข้อมูลตามการเรียนรู้ของเครื่อง ซึ่งแบ่งเป็นดังต่อไปนี้

การจัดกลุ่มแบบ Partitioning Clustering เป็นการแบ่งกลุ่มของข้อมูลที่เกิดจากจุดข้อมูล (Data Point) ที่กระจุกตัวใกล้กับจุดกึ่งกลางของกลุ่ม (Centroid) มากที่สุด

การจัดกลุ่มแบบ Density-Based Clustering เป็นการแบ่งกลุ่มของข้อมูลเกิดจากการกระจุกตัวของ Data Point ที่เกาะกันอย่างหนาแน่นและไม่เป็นรูปลักษณ์ที่ตายตัว โดยอัลกอริทึมนี้ทำได้โดยการระบุกลุ่มต่าง ๆ ในชุดข้อมูลและเชื่อมต่อพื้นที่ที่มีความหนาแน่นสูงเข้ากลุ่มโดยที่พื้นที่ข้อมูลจะถูกแบ่งออกจากกันโดยพื้นที่ว่างของการกระจายตัว

การจัดกลุ่มแบบ Distribution Model-Based Clustering เป็นการแบ่งกลุ่มที่ใช้การสันนิษฐานว่าข้อมูลมีรูปแบบการแจกแจงแบบใดแบบหนึ่ง เช่น การแจกแจงปกติ (Normal Distributions) เมื่อระยะห่างระหว่างจุดศูนย์กลางของการแจกแจง กับ Data Point เพิ่มมากขึ้น ความน่าจะเป็นที่ Data Point ที่เป็นส่วนหนึ่งของการแจกแจงนั้นจะลดลง

การจัดกลุ่มแบบ Hierarchical Clustering การแบ่งกลุ่มประเภทนี้จะสร้างให้เกิดต้นไม้ของกลุ่มข้อมูลขึ้น เหมาะสำหรับข้อมูลที่มีลำดับชั้น เช่น อนุกรมวิธาน (Taxonomy) การแบ่งกลุ่มลักษณะนี้มี 2 ประเภท คือ ล่างขึ้นบน (Agglomerative) และ บนลงล่าง (Divisive)

การจัดกลุ่มแบบ Fuzzy Clustering เป็นประเภทของวิธีการที่วัตถุข้อมูลอาจอยู่ในกลุ่มหรือคลัสเตอร์มากกว่าหนึ่งกลุ่ม ชุดข้อมูลแต่ละชุดมีชุดของสัมประสิทธิ์การเป็นสมาชิกซึ่งขึ้นอยู่กับระดับความเป็นสมาชิกที่จะอยู่ในคลัสเตอร์นั้น ๆ

ในงานวิจัยนี้ผู้วิจัยใช้จะใช้อัลกอลิทึมการจัดกลุ่มข้อมูล 3 วิธีการได้แก่ การจัดกลุ่มแบบเคมีน (k-mean) การจัดกลุ่มแบบ Mean Shift และการจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian Mixture Model) ทดลองกับชุดข้อมูล ซึ่งจะได้กล่าวถึงดังหัวข้อต่อไปนี้

2.1.4.1.1 การจัดกลุ่มแบบเคมีน (k-mean)

การจัดกลุ่มแบบ K-means หรือเรียกอีกอย่างหนึ่งว่า การวิเคราะห์กลุ่มแบบไม่เป็น ขั้นตอน (Nonhierarchical Cluster Analysis) หรือ การแบ่งส่วน (Partioning) โดยใช้ทฤษฎี Euclidean Distance เป็นมาตรวัดระยะพื้นฐานที่ใช้สำหรับหาระยะทางระหว่างจุดสองจุด เป็นที่นิยมใช้ในงานประเภทต่าง ๆ เป็นอย่างมาก เพราะง่ายต่อความเข้าใจ และลักษณะการคำนวณที่คล้ายกับทฤษฎีบทพีทาโกรัส [32]โดยอาศัยการคำนวณระยะทางระหว่าง Centroid กับข้อมูลที่จะนำมาจัดกลุ่ม โดยสามารถคำนวณได้ดังสมการที่ 2.1

$$J = \sum_{j=i}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \quad (2.1)$$

โดยที่ k คือจำนวนกลุ่ม (Class), n คือจำนวน Instance, c คือจุดศูนย์กลาง (Centroid) ของกลุ่ม j

2.1.4.1.2 การจัดกลุ่มแบบ Mean Shift

Mean Shift อัลกอลิทึมเป็นหนึ่งอัลกอลิทึมในการจัดกลุ่มข้อมูล โดย Mean Shift พัฒนาและนำเสนอ ในปี 1975 โดย Fukunaga และ Hostetler[33] ถูกนำมาใช้ในงานด้าน Cluster Analysis และ Image Processing หลักสำคัญคือการใช้ Density Function เป็นตัวสำคัญในการทำ Feature Space Analysis เพื่อค้นหาจำนวนกลุ่มของข้อมูล [16] เป็นการจัดกลุ่มที่ไม่ต้องมีการกำหนดจำนวนกลุ่มไว้ก่อน ซึ่งอัลกอลิทึมนี้จะทำการกำหนดจำนวนกลุ่มให้เองโดยอัตโนมัติ โดยใช้การประมาณการความหนาแน่นเป็นพื้นฐานในการในการเปลี่ยนแปลงของค่าเฉลี่ย [34] โดยสามารถคำนวณได้ดังสมการที่ 2.2 และ 2.3 [35]

$$M_h(y) = \left[\frac{1}{n_x} \sum_{i=1}^{n_x} x_i\right] - y_0 \quad (2.2)$$

โดยที่ n_x คือจุดรอบบริเวณพื้นที่ที่สนใจ และ x_iคือจุดข้อมูล (Data Point) , y_0คือค่าประมาณเริ่มต้นของค่าเฉลี่ย และ$m_h(y)$ คือการเปลี่ยนแปลงของเวกเตอร์ไปยัง y_0ในทิศทางของความหนาแน่นของความน่าจะเป็นที่สูงขึ้น

Mean Shift Vector ได้มาจากการไล่ระดับของการประมาณการความหนาแน่นคือ

$$M_h(y) = \left[\frac{\sum_{i=1}^{n} x_i g_i}{\sum_{j=1}^{n} g_i}\right] - x \tag{2.3}$$

โดยg_i คือการไล่ระดับของฟังก์ชันเคอร์เนลสำหรับจุดข้อมูล x_i

2.1.4.1.3 การจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian Mixture Model)

เทคนิคการจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian Mixture Model) ในการวิจัยนี้จะใช้เป็นหนึ่งในวิธีการของการจัดกลุ่มคุณภาพน้ำ เป็นการสร้างแบบจำลองจากหนึ่งในหลายๆ กลุ่ม ซึ่งกลุ่มอาจมีความแตกต่างกัน แต่ข้อมูลภายในกลุ่มเดียวกันสามารถนำมาเป็นโมเดลต้นแบบได้ แบบจำลองการผสมแบบเกาส์เป็นแบบจำลองที่สร้างขึ้นด้วยฟังก์ชันของความน่าจะเป็นแบบเกาส์เซียน (เป็นการแจกแจงแบบปกติ) เพื่อให้ได้ความถูกต้องแล้วแยกออกมาให้เป็นหลายๆ ส่วนตามฟังก์ชั่นของการแจกแจงแบบปกติ [36] He และคณะ[37]กล่าวไว้ว่า แบบจำลองการผสมแบบเกาส์ (GMM) เป็นวิธีการทางสถิติที่ครบวงจรสำหรับการจัดกลุ่ม แต่ละคลัสเตอร์ถูกแทนด้วยการแจกแจงแบบเกาส์ โดยที่ กระบวนการจัดกลุ่มจะเปลี่ยนเป็นการประมาณค่าพารามิเตอร์ของส่วนผสมแบบเกาส์เซียน โดยใช้ Expectation-Maximization Algorithm (EM)

Expectation-maximization (EM) algorithm เป็นการคำนวณโดยใช้พื้นฐาน Maximum Likelihood Estimation โดยใช้วิธีประมาณค่าพารามิเตอร์ประกอบด้วย 2 ขั้นตอน ได้แก่ ขั้นตอนประมาณค่าคาดหวัง (Expectation : E step)ได้แก่ การประมาณค่า Log-Likelihood ของฟังก์ชันพารามิเตอร์ และขั้นตอนการใช้ค่าสูงสุด (Maximization : M Step) เป็นขั้นตอนการแทนค่าขาดหายด้วยค่าที่ได้จากขั้นตอนประมาณค่าคาดหวัง และทำการประมาณค่าคาดหวังซ้ำเพื่อเปรียบเทียบจนได้ค่าที่เปลี่ยนแปลงน้อยที่สุด และใช้ค่านั้นแทนข้อมูลที่ขาดหาย ข้อดีของวิธีนี้คือ การใช้ข้อมูลได้ครบทุกตัวอย่าง ทั้งมีข้อมูลขาดหายและไม่มีข้อมูลขาดหาย และไม่เกิดความลำเอียงหากข้อมูลขาดหาย มีการกระจายตัวทั้งแบบการขาดหายแบบสุ่มและมีการกระจายตัวแบบสุ่มอย่างสมบูรณ์ ขณะที่ วรางคณา เหนือคูเมืองและเกษศักดิ์ดา ศรีโคตร [38] ได้สรุปขั้นตอนวิธีการทำ EM สำหรับ Gaussian Mixture Model ไว้ว่า จาก Gaussian Mixture Model เป้าหมาย

ก็เพื่อทำให้ค่า Likelihood Function มีค่าสูงสุดเมื่อเทียบกับตัวแปรที่ประกอบด้วย ค่าเฉลี่ยμ ค่าความเบี่ยงเบน Σ และ Mixing Coefficient π โดยสามารถสรุปเป็นขั้นตอนได้ดังนี้

ขั้นตอนที่ 1 ขั้นตอนการกำหนดค่าเริ่มต้น ของค่าเฉลี่ย μ_k ค่าของความเบี่ยงเบน Σ_k และ Mixing Coefficient π_k จากนั้นแทนในสมการเพื่อทำการหาค่า Log Likelihood

ขั้นตอนที่ 2 ขั้นตอน E-Step โดยการหาค่า Responsibility โดยแทนค่าเริ่มต้น ดังสมการที่ 2.4

$$\gamma(z_{mk}) = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{k} \pi_j N(x|\mu_j, \Sigma_j} \tag{2.4}$$

ขั้นตอนที่ 3 ขั้นตอน M-Step โดยการหาค่าตัวแปรใหม่โดยการใช้ค่า Responsibility ที่ได้จากขั้นตอนที่2 ขั้นตอน E-Step แสดงดังสมการที่ 2.5-2.9

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma\,(z_{nk}) X_n \tag{2.5}$$

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma\,(z_{nk})(X_n - \mu_k^{new})^T \tag{2.6}$$

$$\pi_k^{new} = \frac{N_k}{N} \tag{2.7}$$

$$N_k = \Sigma_{n=1}^{N} \gamma(Z_{nk}) \tag{2.8}$$

ขั้นตอนที่ 4 หาค่า Log Likelihood ดังสมการที่ 2.15

$$\text{In } p(X|\, \pi, \mu, \Sigma) = \sum_{n=1}^{N} \text{In} \left\{ \sum_{k=1}^{K} \pi_k N(X_n | \mu_k, \Sigma_k) \right\} \tag{2.9}$$

จากนั้นตรวจสอบการลู่เข้าโดยการทำซ้ำตั้งแต่ขั้นตอนที่ 2

2.1.5 วิธีการวัดประสิทธิภาพของการจัดกลุ่ม (Evaluation of Clustering)

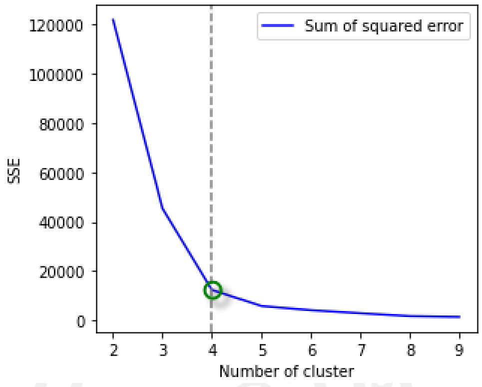
2.1.5.1 การวัดประสิทธิภาพของการจัดกลุ่ม K-Means

2.1.5.1.1 Elbow Method

Elbow Method เป็นหนึ่งในอัลกอลิทึมที่เอาไว้วัด Performance ในการทำ Clustering ของ K-Mean คือวิธีการวัดหาค่า Error ผลรวมของระยะห่างระหว่าง Object กับ Centroid เพื่อการเลือกค่า K ที่เหมาะสมหรือการหาค่า Optimal Cluster Number จุดที่เหมาะสมของจำนวนกลุ่ม (Clusters) คือจุดที่กราฟมีลักษณะ หักศอก ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่ม ที่ดีที่สุด[39] [40]การหาจำนวนกลุ่มที่ดีที่สุดจะถูกนำมาจากค่าผลรวมความคลาดเคลื่อนกำลังสอง Sum of Square Error ซึ่งลดลงอย่างมีนัยสำคัญ ในการคำนวณค่า Sum of Square Error เมื่อค่าที่ได้จากการคำนวณมีความผิดพลาดน้อยลงความชันของเส้นโค้งจะเริ่มเรียบ (Smooth) และจะเกิดเป็นมุมที่มีลักษณะคล้ายข้อศอก (Elbow) ซึ่ง ณ จุด Elbow นี้ จะเป็นจุดที่แสดงค่าของจำนวนกลุ่ม (k) ที่เหมาะสมที่สุด สามารถแสดงตามรอยเส้นประที่แสดงดังภาพประกอบที่ 2.3 เมื่อวิเคราะห์เส้นกราฟแล้ว จะทำให้ได้จำนวนกลุ่ม (k) ที่มีความเหมาะสมโดยใช้สมการในการคำนวนหาค่า SSE [41] ดังสมการที่ 2.10 ซึ่งในงานนี้การหาจำนวนกลุ่มของคุณภาพน้ำจากกลุ่มตัวอย่างจะใช้วิธีการ Elbow Method เข้ามาช่วยในการหาจำนวนกลุ่มที่เหมาะสม

$$SSE = \sum_{i=1}^{n}(y_i - \acute{y})^2 \quad (2.10)$$

โดยที่ n คือจำนวนอินสแตนซ์ , y_i คือ คือค่าของอินสแตนซ์ที่ i^{th} และ $\acute{y}$ คือค่าเฉลี่ยของอินสแตนซ์



ภาพประกอบที่ 2.3 จุดหักศอกของ Elbow Method สำหรับการหาจำนวนกลุ่มที่เหมาะสม

2.1.5.1.2 Silhouette Score

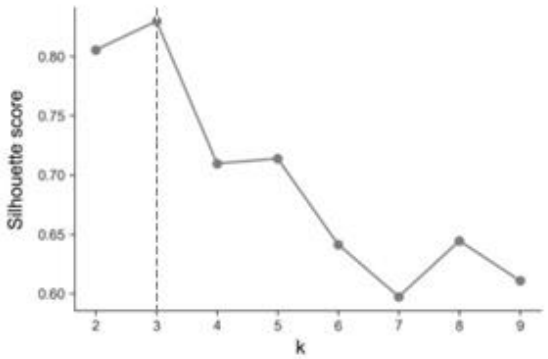
Silhouette Score เป็นอัลกอลิทึมหนึ่งที่เอาไว้วัดประสิทธิภาพในการทำ Clustering ของ K-Mean ซึ่งในงานนี้จะใช้ในการหาจำนวนกลุ่มของคุณภาพน้ำ พิจารณาร่วมกับวิธีการ Elbow Method ที่ได้กล่าวไปข้างต้นนั้น Silhouette Score เป็นเทคนิคที่ใช้วัดว่า Instance นั้นว่ามีความเหมือนกับกลุ่มที่อยู่มากเพียงใด เมื่อเทียบกับกลุ่มอื่น ๆ ค่าของ Silhouette อยู่ในช่วง -1 ถึง 1 ยิ่งมีค่ามากแสดงว่า Instance มีความคล้ายกับกลุ่มที่อยู่นั้นมากและมีความคล้ายกับกลุ่มอื่นน้อย [42] ถ้าค่า Silhouette Score มีค่ามาก หรือเข้าใกล้ +1 หมายความว่าเป็นการหาจำนวนกลุ่มที่ดี ตามรอยเส้นประที่ปรากฏดังภาพประกอบที่ 2.6 Silhouette เป็นมาตรวัดที่อาศัยการยึดเหนี่ยวภายในกลุ่ม และมีความสามารถในการแยกกันระหว่างกลุ่ม โดยใช้ค่าเฉลี่ยของระยะห่างระหว่างจุดกับกลุ่มที่อยู่ใกล้ชิด เทียบกับจุดที่อยู่ภายในกลุ่มเดียวกัน เป็นเครื่องมือที่ใช้สำหรับการจัดกลุ่มข้อมูลใช้เพื่อเลือกจำนวนกลุ่มที่เหมาะสม โดยหาค่า Silhouette จะใช้กับข้อมูลที่เป็น Ratio Scale และเหมาะกับการจัดกลุ่มข้อมูลที่ข้อมูลแต่ละกลุ่มเกาะกลุ่มกัน เมื่อมีการจัดกลุ่มจะมีการพิจารณาสมาชิกที่อยู่ใกล้เคียงถึงคุณลักษณะสองประการคือ ความต่างกัน (Dissimilarities) และความคล้ายคลึงกัน (Similarities) ซึ่งมักนำมาใช้หาจำนวนกลุ่มที่เหมาะสมสำหรับการจัดกลุ่ม โดยเฉพาะกรณีที่เราไม่ทราบกลุ่มที่แท้จริงของข้อมูลมาก่อน [43] Silhouette สามารถใช้ Distance Metric อย่างเช่น Euclidean หรือ Manhattan distance มาช่วยในการคำนวณการวัดระยะ โดยมีสมการในการคำนวณหา Silhouette Score [44] ดังสมการที่ 2.11-2.13

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \tag{2.11}$$

$$b(i) = \underset{k \neq i}{min} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j) \tag{2.12}$$

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} d(i,j) \tag{2.13}$$

โดยที่ $d(i,j)$ คือ Distance ระหว่าง i and j ใน (C), k คือจำนวนของ Clusters, $a(i)$ คือค่าเฉลี่ย Intra-Cluster Distance, และ $b(i)$ คือค่าเฉลี่ยของ Inter-Cluster Distance.



ภาพประกอบที่ 2.4 การหาจำนวนกลุ่มโดยพิจารณาจากค่า Silhouette Score [45]

2.1.5.2 การวัดประสิทธิภาพของการจัดกลุ่ม Gaussian mixture model

2.1.5.2.1 AIC-Akaike information criterion

Akaike Information Criterion (AIC) เป็นหนึ่งในเครื่องมือที่นิยมใช้ในการสร้างแบบจำลองทางสถิติ ถูกนำเสนอในปี 1973 โดย Hirotugu Akaike [46] เป็นส่วนขยายของหลักการความน่าจะเป็นสูงสุด ถูกนำไปใช้เพื่อประเมินพารามิเตอร์ของแบบจำลองตามโครงสร้างและมิติของแบบจำลองได้รับการกำหนด ความคิดเชิงสัมบูรณ์ของ Akaike คือการรวมกันเป็นขั้นตอนเดียวในกระบวนการประเมินด้วยการกำหนดโครงสร้างและมิติ[47] ซึ่งเป็นเกณฑ์ที่พิจารณาจากการความคลาดเคลื่อน เพื่อนำมาใช้ปรับค่าการพยากรณ์ให้มีความแม่นยำมากขึ้น ในงานนี้จะใช้ค่า AIC และ BIC ซึ่งค่า BIC จะได้กล่าวในหัวข้อที่ 2.1.12 มาพิจารณาการหาจำนวนกลุ่มของ Gaussian mixture model ซึ่งจะพิจารณาจากค่าที่ต่ำสุดของทั้งสองค่าในการหาจำนวนกลุ่ม ขณะที่ Kim-Hung และคณะ [48] กล่าวว่า AIC เป็นวิธีการเพื่อเปรียบเทียบโมเดลที่แตกต่างกับผลลัพธ์ที่ได้รับ ดังสมการที่ 2.14

$$AIC \coloneqq -2l(\hat{\theta}|y) + 2K, \tag{2.14}$$

โดยที่ $l(\hat{\theta}|y)$คือ ความน่าจะเป็นสูงสุดของแบบจำลอง

และ K คือจำนวนพารามิเตอร์

2.1.5.2.1 BIC- Bayesian information criterion

Bayesian Information Criterion (BIC) คือ เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ ถูกนำเสนอครั้งแรกโดย Schwarz ในปี 1978 โดยจะเลือกตัวแบบที่ให้ค่าบีไอซีต่ำสุด ซึ่งในงานนี้จะใช้ BIC ร่วมกับ AIC ในการพิจารณา [48]   เป็นตัวแบบที่ถูกต้องโดยมีสมการดังสมการที่ 2.15

$$BIC \coloneqq -2l(\hat{\theta}|y) + Kln(n), \tag{2.15}$$

โดยที่ $l(\theta|y)$คือ ความน่าจะเป็นสูงสุดของแบบจำลอง K คือจำนวนพารามิเตอร์ และ n คือขนาดของตัวอย่างข้อมูล

2.1.6 Principal Component Analysis (PCA)

PCA คือการหารูปแบบความสัมพันธ์ระหว่าง ตัวแปรของข้อมูล เพื่อนำไปทำกระบวนการลดมิติของข้อมูล โดยที่ยังคงไม่ทำให้สูญเสีย information หรือข้อมูลสำคัญไป PCA จัดเป็น Unsupervised Machine Learning Algorithm เพราะไม่ได้มีการนำเอา Label หรือ Target หรือ Class มาพิจารณาร่วมด้วย ถือว่าทั้ง Dataset คือข้อมูล Class เดียวกันหมด [49] เป็นเทคนิคหนึ่งที่ใช้กระบวนการทางสถิติและเมทริกซ์ (Matrix) เพื่อเข้ามาช่วยอธิบายข้อมูลให้เป็นที่เข้าใจได้ง่ายยิ่งขึ้น เพื่อลดความซับซ้อนลง

หลักการเบื้องต้นของ PCA เริ่มจากการทำ Normalize ข้อมูล ด้วย Standard Deviation ได้ จากนั้นนำค่าที่ได้มาคำนวณหา Covariance Matrix จากนั้นนำไปคำนวณ Eigen Vector และ Eigen Value เพื่อใช้เป็นค่าในสกัดคุณสมบัติเด่นออกมา [50] การหา Component จะหาได้จากสมการที่ 2.16

$$C = W\Lambda W^T \qquad (2.16)$$

โดยที่ Λ คือเมทริกซ์แนวทแยงซึ่งประกอบด้วยค่าลักษณะเฉพาะที่เรียงลำดับทั้งหมดของ C

W คือ เมทริกซ์ ทั้งหมดของ C

การหา Component ที่เหมาะสม

ในการหาจำนวน Component ที่เหมาะสมของการลดมิติข้อมูลด้วย PCA เพื่อแก้ปัญญา High-Dimensional Data เพื่อลดจำนวนของตัวแปรลง [51]โดยสามารถหาได้จากสมการที่ 2.17

$$y = W_y^T p \qquad (2.17)$$

2.2 งานวิจัยที่เกี่ยวข้อง

Carrasco และคณะ[52] กล่าวไว้ว่า คุณภาพน้ำเป็นสิ่งที่ละเอียดอ่อนและเป็นสิ่งที่ต้องพึงตระหนักถึงลักษณะทางกายภาพ เคมี และชีวภาพ ความซับซ้อนของธรรมชาติสำหรับการศึกษาคุณภาพน้ำนั้น มีความจำเป็นที่ต้องค้นหาตัวแบบเพื่อทำการระบุตัวแปรที่มีผลต่อคุณภาพน้ำที่มากที่สุด ดังนั้นการใช้เทคนิคการวิเคราะห์แบบหลายตัวแปรจะช่วยให้การหาความสัมพันธ์ และข้อสรุปที่

จะช่วยให้กำหนดสถานะของคุณภาพน้ำทางชีวภาพ กายภาพและทางเคมีได้ ในการประเมินคุณภาพน้ำมีการประเมินตัวแปร ดังต่อไปนี้ อุณหภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) ความใสของน้ำ (Transparency) ความขุ่นของน้ำ (Turbidity) ไนเตรด (Nitrates) ออโธฟอสเฟต (Orthophosphates) ฟอสฟอรัส (Phosphorus) ไนโตรเจน (Total Nitrogen) คลอโรฟิลล์(Chlorophyll) รังสีจากดวงอาทิตย์ (Solar Radiation) ออกซิเจนละลายน้ำ (Dissolved Oxygen) ไมโครซิสติน (Microcystins) ทั้งหมดนี้จะเป็นตัวแปรที่กำหนดคุณภาพน้ำจากแหล่งน้ำตัวอย่าง คือ แม่น้ำ Gamboa และ Paraiso จากลุ่มน้ำปานามา โดยมีกระบวนการวิจัยอยู่ทั้งหมด 3 ขั้นตอน โดยขั้นตอนที 1 เป็นการวิเคราะห์ทางสถิติขั้นพื้นฐาน โดยการหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ขั้นตอนที่ 2 เป็นการวิเคราะห์โดยใช้ HJ-Biplot เพื่อการวิเคราะห์ทางสถิติสำหรับการค้นหาการเชื่อมโยงระหว่างตัวแปรทางเคมี กายภาพและชีวภาพ ด้วยเทคนิค HJ-Biplot ขั้นตอนที่ 3 เป็นการวิเคราะห์กลุ่ม เพื่อระบุการจัดกลุ่มของการสุ่มตัวอย่างและตัวแปรที่ได้รับผลกระทบ โดยกลุ่มจะถูกคำนวณผ่านพิกัด Biplot ที่ได้จากขั้นตอนที่ 2 โดยใช้เทคนิคเคมีน (K-mean) ผลลัพธ์ที่ได้จากการประยุกต์ใช้วิธีการวิเคราะห์หลายตัวแปรที่เรียกว่า HJ-Biplot แสดงให้เห็นถึงความสัมพันธ์ของตัวแปรทางเคมี องค์ประกอบทางกายภาพและชีวภาพ นอกจากนั้นแล้วผลลัพธ์ยังสอดคล้องกับกลุ่มตัวอย่างสองกลุ่มที่ตรงกับฤดูกาลของภูมิภาค โดยกลุ่มที่ 1 ประกอบด้วยตัวแปร ความเป็นกรดเป็นเบส (PH) ความใสของน้ำ (Transparency) คลอโรฟิลล์ (Chlorophyll) ออกซิเจนละลายน้ำ (Dissolved Oxygen) และอุณหภูมิ (Temperature) ในกลุ่มที่ 2 ประกอบด้วยตัวแปร ไนเตรด (Nitrates) ออโธฟอสเฟต (Orthophosphates) ความขุ่นของน้ำ (Turbidity) และที่เหลือทั้งหมด ค่าพารามิเตอร์ทั้งหมดจะมีความแปรปรวนในฤดูฝน และจะมีผลทำให้เกิด ไซยาโนแบคทีเรีย (Ciabobacteria) ที่สามารถก่อให้เกิดสารพิษ (Toxigenic) ขึ้นได้ นอกจากนี้ยังแสดงให้เห็นว่าวิธีการทางสถิติหลายตัวแปรมีผลสำหรับการตีความชุดข้อมูลที่ซับซ้อนโดยเฉพาะสำหรับการตรวจสอบคุณภาพน้ำ

Cloete และคณะ[53] ทำการออกแบบและพัฒนาระบบตรวจสอบคุณภาพน้ำโดยมีวัตถุประสงค์เพื่อแจ้งผู้ใช้ให้ทราบถึงค่าพารามิเตอร์คุณภาพน้ำแบบเรียลไทม์ เช่นการไหลของน้ำ อุณภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) การนำไฟฟ้า (Conductivity) และการลดออกซิเดชั่น (Oxidation Reduction Potential) โดยค่าพารามิเตอร์ทางกายภาพและเคมีเหล่านี้ใช้สำหรับตรวจจับสิ่งปนเปื้อนในน้ำ เซ็นเซอร์ได้รับการออกแบบนำไปใช้กับวงจรปรับสภาพสัญญาณเชื่อมต่อกับไมโครคอนโทรลเลอร์ สำหรับการประมวลผลและวิเคราะห์ข้อมูล ในการออกแบบใช้ Zigbee Protocal ในการรับละส่งข้อมูลระหว่างโมดูล ระหว่างโหนดการวัดและโหนดการแจ้งเตือน โหนดการแจ้งเตือนทำหน้าที่ในการอ่านค่าจากเซ็นเซอร์และส่งสัญญาณเตือนเมื่อพารามิเตอร์คุณภาพน้ำถึงระดับที่ไม่ปลอดภัย โดยโหนดที่ทำการวัดค่าส่งข้อมูลโดย ZigBee Protocal ไปยังโหนดสำหรับ

การแจ้งเตือนเพื่อแสดงภาพและเสียง ผลการทดลองแสดงให้เห็นว่าระบบที่พัฒนาสามารถอ่านพารามิเตอร์ทางกายภาพ ทางเคมี สามารถประมวลผลและแสดงผลได้

Di และคณะ[54] กล่าวว่า น้ำจากโรงงานอุตสาหกรรมส่วนใหญ่ถูกปล่อยออกสู่แม่น้ำโดยไม่มีการบำบัด เนื่องจากขาดข้อมูลและกระบวนการบางอย่าง โดยความสัมพันธ์ระหว่างสารมลพิษที่ปล่อยออกมากับน้ำเสียและน้ำผิวดินยังไม่ได้รับการตีแผ่อย่างกว้างขวางและรวมถึงเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน เช่น อัลกอลิทึมของการจัดกลุ่ม (Clustering Algorithm) ที่ถูกละเลยในสาขาการวิจัยที่เกี่ยวข้อง หลังจากนั้นได้ทำการตรวจสอบข้อมูลแบบเรียลไทม์เพื่อต้องการทราบค่าออกซิเจนทางเคมี (Chemical Oxygen Demand) แอมโมเนียไนโตรเจน (Ammonia Nitrogen) ความเป็นกรดเป็นเบส (PH) และออกซิเจนละลายในน้ำ (Dissolved Oxygen) ในน้ำเสียที่ถูกปล่อยออกมาจาก 2,213 โรงงานและลงสู่น้ำผิวดิน จำนวน 18 พื้นที่ในการตรวจสอบ ใน 7 เขตการปกครองของลุ่มแม่น้ำแยงซี โดยทำการรวบรวมข้อมูลจากปี 2016-2017 และทำการวิเคราะห์ข้อมูลด้วยอัลกอลิทึมการจัดกลุ่มแบบ Partitioning Around Medoids (PAM) Expectation–Maximization (EM) Wilcoxon Test และ Spearman Correlation ผลการวิจับพบว่า เมื่อทำการเปรียบเทียบกับพื้นที่ปลอดมลภาวะสารปนเปื้อน กลุ่มพื้นที่ที่มีสารปนเปื้อนมีน้ำเน่าเสียมากขึ้น ค่าออกซิเจนทางเคมี (Chemical Oxygen Demand) ถูกปล่อยออกมาค่อนข้างสูง มากกว่า 100 มิลลิกรัมต่อลิตร และแอมโมเนียไนโตรเจน (Ammonia nitrogen) มากกว่า 6 มิลลิกรัมต่อลิตร ค่าความเป็นกรดเป็นเบส (PH) ค่อนข้างต่ำโดยมีค่าที่วัดได้น้อยกว่า 6 จากโรงงานอุตสาหกรรม 15 แห่ง ความสัมพันธ์ระหว่างคุณภาพของน้ำเสียและน้ำผิวดินลดลงทำให้ที่มีมลพิษเพิ่มขึ้น โดยใช้การจัดกลุ่มด้วย อัลกอลิทึม (Expectation–Maximization) EM ในการประเมินคุณภาพน้ำ ผลการวิจัยพบว่าอัลกอลิทึมการเรียนรู้ของเครื่องแบบไม่มีผู้สอน มีบทบาทและมีประสิทธิภาพในการตรวจสอบข้อมูลแบบเรียลไทม์โดยเน้นความสัมพันธ์เชิงพื้นที่ระหว่างมลพิษที่มาจากการระบายน้ำเสียกับน้ำผิวดินเพื่อเป็นระบบสนับสนุนการจัดการทรัพยากรน้ำทางวิทยาศาสตร์

Gao และคณะ[16] ได้กล่าวไว้ว่า คุณภาพน้ำและความปลอดภัยของผลผลิตจากสัตว์น้ำที่เลี้ยงมีความสำคัญอย่างมาก และได้ทำการพัฒนาระบบเพาะเลี้ยงปลาอัจฉริยะ ด้วยเทคโนโลยี Internet of Thing (IoT) โดยมีการติดตั้งโหนดเซนเซอร์ (Sensor Node) สองจุด แต่ละโหนดทำการติดตั้งเซนเซอร์อุณหภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) ออกซิเจนละลายน้ำ (Dissolved Oxygen) การนำไฟฟ้าของน้ำ (Water Electrical Conductivity) และความขุ่นของน้ำ (Water Electrical Conductivity) เพื่อทำการรวมรวมพารามิเตอร์ของคุณภาพน้ำสำหรับวิเคราะห์ข้อมูลเพื่อการควบคุมอุปกรณ์ที่เชื่อมต่อกับระบบ เช่น ระบบปั๊มน้ำ เครื่องเติมอากาศ เครื่องให้อาหาร เครื่องกรองน้ำ เป็นต้น โดยระบบควบคุมติดตามและรวมถึงการพยากรณ์ที่ช่วยในการจัดการคุณภาพน้ำอัตโนมัติ และกระบวนการขายปลา โดยระบบช่วยให้เกษตรกรผู้เลี้ยงปลาสามารถควบคุมและ

จัดการอุปกรณ์การบำบัดคุณภาพน้ำในบ่อปลาได้เป็นอย่างดี อีกทั้งยังช่วยให้ผู้บริโภคสามารถติดตามและดูข้อมูลกระบวนการเลี้ยงในที่ผ่านมาโดยใช้รหัสคิวอาร์โค้ด (QR Code) ซึ่งสามารถเพิ่มรายได้ของเกษตรกรผู้เลี้ยงปลาและความปลอดภัยของอาหารสำหรับผู้บริโภค นอกจากนี้ได้นำเสนอวิธีการพยากรณ์คุณภาพน้ำสำหรับการจัดการแบบอัจฉริยะสำหรับตรวจจับและลบข้อมูลที่ผิดปกติโดยใช้อัลกอลิทึม LOF จากนั้นข้อมูลจะถูกวิเคราะห์ และทำนายผลโดยใช้อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) ช่วยให้สามารถทำนายคุณภาพน้ำล่วงหน้าและควบคุมให้อยู่ในช่วงที่ปลอดภัยและเป็นไปตามมาตรฐาน จากผลการทดลองการทำนายคุณภาพน้ำโดยใช้อัลกอลิทึมต้นไม้ตัดสินใจแบบ (M5) ได้ผลการทดลองและความสัมพันธ์ดีกว่าเมื่อเปรียบเทียบกับ อัลกอลิทึม Cubist RF และGBM อีกทั้งค่าเฉลี่ยความคลาดเคลื่อนของอัลกอลิทึมต้นไม้ตัดสินใจแบบ (M5) มีเพียงเล็กน้อย สามารถสรุปได้ว่าวิธีการที่นำเสนอสามารถทำนายคุณภาพน้ำได้อย่างแม่นยำและมีประสิทธิภาพ

Geetha และ Gouthami [55] ได้นำเสนอระบบตรวจสอบคุณภาพน้ำอัจฉริยะ สำหรับการตรวจสอบคุณภาพน้ำด้วยเทคโนโลยี Internet of Thing IoT โดยใช้แบบจำลองที่พัฒนาขึ้นสำหรับการทดสอบตัวอย่างน้ำและการวิเคราะห์ข้อมูลผ่านระบบอินเทอร์เน็ต โดยระบบมีการแจ้งเตือนไปยังผู้ใช้ เมื่อมีการเปลี่ยนแปลงของค่าพารามิเตอร์คุณภาพน้ำจากค่ามาตรฐานที่ทำการกำหนดเอาไว้ล่วงหน้า โดยใช้ค่าพารามิเตอร์ในการตรวจสอบคุณภาพน้ำ 5 ตัว ประกอบด้วย ค่าการนำไฟฟ้า (Electrical Conductivity) ค่าความเป็นกรดเป็นด่าง (PH) ค่าความขุ่น (Turbidity) ค่าอุณหภูมิของน้ำ (Temperature) และระดับน้ำ (Water Level) โดยใช้เทคโนโลยีไร้สาย Zigbee Protocal ซึ่งเป็นเทคโนโลยีที่ใช้พลังงานต่ำตามมาตรฐาน IEEE 802.15.4 สำหรับการสื่อสารระหว่างเซ็นเซอร์กับตัวควบคุม และจากตัวควบคุมไปยังคลาวด์ที่จัดเก็บข้อมูล ใช้การสื่อสารแบบ 3G หรือระบบเครือข่ายอินเทอร์เน็ต ค่าพารามิเตอร์ทั้ง 5 ตัวที่ถูกตั้งค่าให้เชื่อมต่อกับแพลตฟอร์ม Ubidots ที่ใช้ในการวิเคราะห์ข้อมูล โดยที่ผลลัพธ์จากค่าพารามิเตอร์ที่ได้จะถูกนำไปเปรียบเทียบกับมาตรฐานคุณภาพน้ำที่กำหนดโดย WHO ที่ตั้งค่าไว้ใน Ubidots และทำการแจ้งเตือนไปยังผู้ใช้งานในรูปแบบของ SMS หรืออีเมลลแบบอัตโนมัติในกรณีที่ค่าพารามิเตอร์ที่วัดได้มีค่าเกินกว่ากำหนดที่ตั้งไว้

Chowdury และคณะ[11] ได้นำเสนอระบบตรวจสอบคุณภาพน้ำโดยใช้เซนเซอร์ในการวัดค่าพารามิเตอร์ทางกายภาพและทางเคมีของน้ำซึ่งมีส่วนประกอบหลักที่ประกอบด้วยระบบเครือข่ายเซนเซอร์ไร้สาย Wireless Sensor Network (WSN) ไมโครคอนโทรลเลอร์สำหรับประมวลผลระบบระบบการสื่อสารสำหรับการสื่อสารภายในโหนดและเซ็นเซอร์ การเข้าถึงข้อมูลแบบเรียลไทม์โดยใช้การตรวจสอบระยะไกลและเทคโนโลยี Internet of Thing (IoT) โดยใช้เซนเซอร์ในการวัดค่าความเป็นกรดเป็นเบสของน้ำ (PH) ความขุ่นของน้ำ (Turbidity) และอุณภูมิของน้ำ (Temperature) จัดเก็บข้อมูลบนเครื่องคอมพิวเตอร์แม่ข่ายและทำการวิเคราะห์ข้อมูลแบบ Spark Streaming Analysis ด้วย Spark MLlib แบบจำลองโครงข่ายประสาทการเรียนรู้เชิงลึก (Deep Learning

Neural Network Models) Belief Rule Based (BRB) ของเทคโนโลยีฮาดูป (Hadoop) เพื่อเปรียบเทียบกับค่ามาตรฐาน หากค่าที่วัดได้สูงกว่าเกณฑ์มาตรฐานระบบจะทำการแจ้งเตือนผ่านระบบ SMS แบบอัตโนมัติ โดยการนำเสนอระบบตรวจสอบน้ำที่มีความถี่สูงนี้ มีความคล่องตัวและใช้พลังงานต่ำ โดยมีความมุ่งหวังเพื่อให้ประชากรตระหนักถึงสารปนเปื้อนในน้ำและการรณรงค์การหยุดมลพิษทางน้ำ ผลการวิจัยพบว่าระบบสามารถแสดงผลการตรวจวัดค่าและทำการแสดงผลผ่านจอแอลซีดี พีซี หรือมือถือได้แบบเรียลไทม์ หากค่าที่ได้จากเซนเซอร์ไม่อยู่ในเกณฑ์ที่กำหนดไว้ระบบจะแสดงค่าเป็น BAD หากค่าที่ได้จากเซนเซอร์อยู่ในเกณฑ์ที่กำหนดไว้ระบบจะแสดงค่าเป็น GOOD และทำการแสดงผลในรูปแบบของกราฟ

Muangthong [56] ได้นำเสนอการวิเคราะห์มลพิษทางน้ำของแม่น้ำลำตะคอง โดยมีวัตถุประสงค์เพื่อลดสถานีและพารามิเตอร์คุณภาพน้ำให้สามารถจัดการได้ โดยการวิเคราะห์กลุ่มซึ่งเป็นวิธีการทางสถิติสำหรับเรียงลำดับข้อมูลที่คล้ายกันลงในกลุ่มและปรับนำไปใช้ในการจัดกลุ่มของสถานีที่มีลักษณะเดียวกันของคุณภาพน้ำ จากทั้งหมด 15 สถานี โดยมี 7 สถานีถูกเลือกใช้ในการวิเคราะห์กลุ่ม การวิเคราะห์ใช้การเลือกตัวแปรเพื่อเป็นตัวอย่างในการลดพารามิเตอร์ของคุณภาพน้ำ โดยใช้พารามิเตอร์ในการวิเคราะห์ทั้งหมด 17 พารามิเตอร์ประกอบด้วย อุณหภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) ความขุ่น (Turbidity) การนำไฟฟ้า (Conductivity) ความเค็ม (Salinity) ออกซิเจนละลายน้ำ (Dissolved Oxygen) ความต้องการออกซิเจนทางชีวเคมี (Biochemical Oxygen Deman) แบคทีเรียโคลิฟอร์มทั้งหมด (Total Coliform Bacteria) แบคทีเรียโคลิฟอร์มจากอุจจาระ (Fecal Coliform Bacteria) ฟอสฟอรัส (Total Phosphorus) ไนเตรตไนโตรเจน (Nitrate Nitrogen NO3-N) ไนเตรตไนโตรเจน (Nitrate Nitrogen NO2-N) แอมโมเนียไนโตรเจน (Ammonia Nitrogen) ของแข็ง (Total Solid) ของแข็งที่ละลายทั้งหมด (Total Dissolved Solids) ของแข็งที่ถูกระงับ (Suspended Solids) และแคลเซียมคาร์บอเนต (Calcium carbonate) จากกรมควบคุมมลพิษที่ทำการเก็บข้อมูลเป็นรายเดือนทั้ง 7 สถานีเป็นเวลาทั้งสิ้น 17 ปี ในช่วงปี 1996-2012 พารามิเตอร์ส่วนใหญ่มีหน่วยวัดที่แตกต่างกัน จึงต้องทำการแปลงค่าให้เป็นมาตรฐานก่อนทำการวิเคราะห์ โดยค่าเมทริกซ์มาตรฐานได้มาจาก เมทริกซ์คุณภาพน้ำโดยการเปลี่ยนค่าพารามิเตอร์ไปเป็นค่า Z Score และทำการแปลงเมทริกซ์คุณภาพน้ำที่ได้มาตรฐานเป็นเมทริกซ์แบบยุคลิด (Euclidean distant matrix) ซึ่งทำการตรวจสอบระยะห่างของสถานีคุณภาพน้ำในการตรวจสอบแต่ละคู่สถานี เพื่อทำการจัดกลุ่มให้แต่ละสถานี และทำการลดค่าพารามิเตอร์คุณภาพน้ำเพื่อให้ได้ค่าพารามิเตอร์ที่เป็นตัวแทนน้อยที่สุดที่มีผลต่อมลพิษในแม่น้ำโดยใช้วิธีการทางสถิติคือการวิเคราะห์ปัจจัย (Factor Analysis) โดยใช้การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) และการวิเคราะห์ปัจจัยร่วม (Common Factor Analysis) ผลการวิจัยพบว่า แม่น้ำลำตะคองสามารถจำแนกได้ตามวัตถุประสงค์ออกเป็น 2 ส่วนคือ กลุ่มต้นน้ำและกลุ่มปลายน้ำ โดยใช้เทคนิค

การวิเคราะห์กลุ่ม คือส่วนต้นน้ำประกอบด้วยสถานีตรวจวัดน้ำที่ LT03 04 05 06 และ07 ปลายน้ำประกอบด้วย สถานีตรวจวัดที่ LT01 และ 02

Pathak และคณะ[57] กล่าวไว้ว่า ปัจจุบันระบบชลประทานตามธรรมชาติอยู่ในสถานการณ์ที่น่าเป็นห่วงเนื่องจากการขาดแคลนน้ำเพิ่มขึ้น โดยส่วนใหญ่เกิดจากการเพิ่มขึ้นของประชากรและการเปลี่ยนแปลงของสภาพภูมิอากาศ ดังนั้นการควบคุมทรัพยากรน้ำ เพื่อเพิ่มการจัดสรรน้ำจึงมีความสำคัญเป็นอย่างมาก การเปลี่ยนแปลงภูมิอากาศส่งผลกระทบต่อการผลิตและพืชผลทางการเกษตรอย่างมีนัยสำคัญ และได้นำเสนอขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบนกดุเหว่า (Cuckoo Search Algorithm) ในการจัดสรรน้ำเพื่อการเกษตรภายใต้เงื่อนไขต่าง ๆ โดยมีการรวบรวมพารามิเตอร์ต่าง ๆ เช่นอุณหภูมิ (Temperature) ความขุ่น (Turbidity) ค่าความเป็นกรดเป็นเบส (PH) โดยใช้แพลตฟอร์ม Internet of Thing (IoT) เข้ามาช่วยพร้อมกับเซ็นเซอร์และระบบการสื่อสารแบบไร้สาย ข้อมูลที่รับจากเซ็นเซอร์ถูกแสดงบนระบบคลาวด์ผ่าน ThingSpeak ผลการทดลองพบว่า ระบบสามารถรักษาสภาพพื้นที่เพาะปลูกและช่วยในการบริหารจัดการน้ำได้เป็นอย่างดี

Myint และคณะ[58] กล่าวว่าการติดตามตรวจสอบคุณภาพน้ำที่มีประสิทธิภาพและประสิทธิผล นั้นมีความสำคัญอย่างยิ่งต่อการดำเนินการแก้ไขปัญหาน้ำเสียทั่วโลก ด้วยการพัฒนาเพิ่มขึ้นอย่างรวดเร็วของเทคโนโลยี Wireless Sensor Network (WSN) ภายใต้สภาพแวดล้อมของ Internet of Thing (IoT) การตรวจสอบคุณภาพน้ำแบบเรียลไทม์ทำการตรวจสอบจากระยะไกลโดยการเก็บข้อมูล การส่งข้อมูลและการประมวลผลตามเวลาจริง จากนั้นได้นำเสนอระบบตรวจสอบคุณภาพน้ำอัจฉริยะที่สามารถกำหนดค่าได้ตามสภาพแวดล้อมของ Internet of Thing (IoT) โดยนำเสนอระบบ WQM (Water Quality Monitoring) ซึ่งประกอบประกอบด้วยการออกแบบบอร์ดเซนเซอร์ โมดูลการสื่อสารไร้สายที่ใช้ Zigbee โปรโตคอล และคอมพิวเตอร์ส่วนบุคคล โดยบอร์ดที่ออกแบบเป็นบอร์ดแบบชิปเดียวใช้เชื่อมต่อเครื่องแปลงสัญญาณไปยังเครือข่ายเซ็นเซอร์โดยใช้ FPGA (Field Programmable Gate Array) เชื่อมต่อกับโมดูล XBee แบบไร้สาย เพื่อทำการรวบรวมพารามิเตอร์ 5 ตัวของข้อมูลน้ำประกอบด้วย ความเป็นกรดเป็นเบสของน้ำ (PH) ระดับน้ำ (Water Level) ความขุ่นของน้ำ (Turbidity) คาร์บอนไดออกไซด์ (Carbon Dioxide) และอุณภูมิของน้ำ (Temperature) ผลการทดลองปรากฏว่าการออกแบบระบบ WQM โดยค่าพารามิเตอร์น้ำที่เก็บรวบรวมจะถูกส่งแบบไร้สายไปยังอุปกรณ์ตรวจสอบโดยใช้พีซีที่ติดตั้ง Nios II ของซอฟต์แวร์ Quartus II โดยใช้ภาษา VHDL (Very High Speed Integrated Circuit Hardware Description Language) ในการพัฒนา จากนั้นทำการแสดงผลข้อมูลที่รวบรวมทั้ง 5 ค่าจากเซนเซอร์บนแผงควบคุม Grafana ที่ติดตั้งบนเครื่องคอมพิวเตอร์PC และทำการแสดงผลของกราฟที่พัฒนาด้วยภาษาไพธอน (Python) ในงานวิจัยนี้ข้อมูลถูกตรวจสอบอย่างต่อเนื่องและแสดงผลแบบเรียลไทม์ทุก ๆ 5

วินาที โดยระบบ WQM ที่นำเสนอสามารถลดพลังงานและมีประสิทธิภาพสูงกว่าระบบ WSN (Wireless Sensor Networks) ที่ใช้ไมโครคอนโทรลเลอร์แบบเดิม

Pranata และคณะ[59] นำเสนอระบบการตรวจสอบคุณภาพน้ำแบบเรียลไทม์โดยใช้เฟรมเวิร์คซึ่งเป็นสถาปัตยกรรมของ MQTT broker ในระบบเซ็นเซอร์จะตรวจวัด อุณหภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) และออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen) ข้อมูลที่รวบรวมทั้งหมดจะถูกเก็บไว้ในฐานข้อมูลและประมวลผลเพื่อวิเคราะห์คุณภาพน้ำ การทดลองนำเสนอสถาปัตยกรรม Pub/Sub และ MQTT เป็นโปรโตคอลที่เหมาะกับการใช้งานกับระบบ Internet of Thing (IoT) เพื่อแสดงประสิทธิภาพของสถาปัตยกรรมบนระบบเครือข่ายและการทำงานสำหรับขนาดของข้อมูลที่หลากหลาย จากนั้นทำการวิเคราะห์ความสัมพันธ์ระหว่างอุณหภูมิ (Temperature) ความเป็นกรดเป็นเบส (PH) และออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen) ผลการการทดลองสรุปว่าอุณหภูมิของน้ำ (Temperature) แปรผกผันกับค่าความเป็นกรดเป็นเบส (PH) และค่าออกซิเจนละลายน้ำ (Dissolved Oxygen)

ในขณะที่ Ding และคณะ [60] กล่าวไว้ว่า การตรวจจับความผิดปกติเป็นปัญหาที่เกิดขึ้นมานานของระบบ การออกแบบการตรวจจับความผิดปกติที่มีคุณภาพสูงนั้นจะมีประโยชน์ต่อแอปพลิเคชันเป็นอย่างมาก เช่นการตรวจสอบระบบการป้องกันภัยพิบัติและการตรวจจับการบุกรุกต่าง ๆ เป็นต้น อัลกอริทึมการตรวจจับความผิดปกติส่วนใหญ่นั้น มีความสามารถน้อยกว่าทั้งในด้านประสิทธิภาพและความสามารถแบบเรียลไทม์พร้อมกัน จากนั้นได้นำเสนอ LGMAD อัลกอริทึมที่เป็นการตรวจจับความผิดปกติแบบเรียลไทม์โดยใช้ Long-Short Term Memory (LSTM) และ Gaussian Mixture Model (GMM) โดยทำการตรวจสอบประเมินความผิดปกติแบบเรียลไทม์ ของแต่ละช่วงเวลาโดยใช้ Long-Short Term Memory (LSTM) จากนั้น Gaussian Mixture Model (GMM) ถูกนำมาใช้เพื่อให้การตรวจสอบแบบหลายมิติของความผิดปกติที่เป็นไปได้ โดยใช้ชุดข้อมูลของ NAB และชุดข้อมูลที่สร้างขึ้นเองเพื่อใช้ในการตรวจสอบ โดยการทดลองแสดงให้เห็นถึงความสามารถของ LGMAD อัลกอลิทึม ที่ดีกว่าอัลกอริทึมการตรวจจับความผิดปกติที่มีอยู่

Zhang และคณะ[61] ได้กล่าวถึงการปรับปรุงความแม่นยำในการพยากรณ์พลังงานกังหันลมให้เป็นไปมาตรฐานและมีประสิทธิภาพ โดยใช้เครือข่ายการเรียนรู้เชิงลึกเพื่อคาดการณ์พลังงานกังหันลมโดยใช้อัลกอริทึม Long Short-Term Memory Network (LSTM) และ Gaussian Mixture Model (GMM) เพื่อทำการวิเคราะห์ลักษณะการกระจายข้อผิดพลาดของการพยากรณ์พลังงานกังหันลมในระยะสั้น อัลกอริทึม Long Short-Term Memory Network (LSTM) ใช้ในการคาดการณ์พลังงานและความไม่แน่นอนสำหรับกังหันลม 3 จุดภายในฟาร์มกังหันลม ตามข้อมูลพยากรณ์อากาศเชิงตัวเลข (Numerical Weather Prediction) และข้อมูลพลังงานย้อนหลังของกังหันทั้งสามจุด ความแม่นยำในการพยากรณ์ของกังหันที่มีจำนวนมาก ทั้งนี้จึงนำเอากังหันสามจุดซึ่งเป็นตัวอย่างที่ดีที่สุด

ของการฝึก โดยใช้อัลกอลิทึม LSTM RBF Wavelet DBN BPNN และ ELMAN นำมาการพยากรณ์พลังงานกังหันลม ผลปรากฏว่า LSTM สามารถปรับปรุงความแม่นยำในการพยากรณ์ได้ดีที่สุด นอกจากนี้ ยังพบความแตกต่างของความเชื่อมั่นทั้ง 3 จุดซึ่งเป็นผลการวิเคราะห์ของแบบจำลอง GMM MDN และ RVM สรุปผลการทดลองปรากฏว่าวิธีการ LSTM แสดงให้เห็นถึงความแม่นยำที่สูงกว่าและมีการลู่เข้าที่รวดเร็วกว่าวิธีอื่น ๆ และอัลกอลิทึม GMM มีประสิทธิภาพและการประเมินผลที่ดีกว่าวิธีอื่น

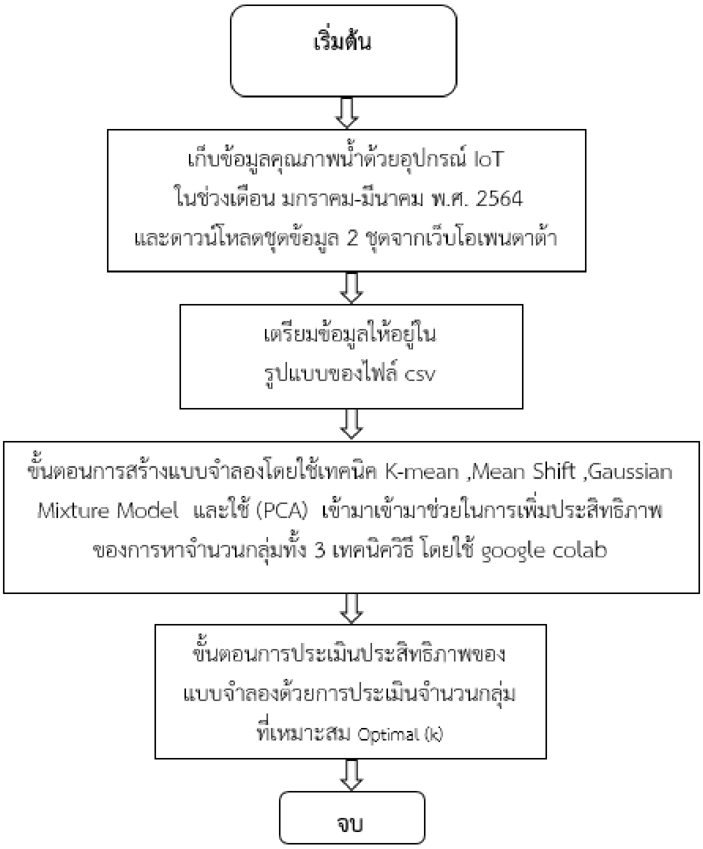
Jia และคณะ[36] ได้นำเสนอรูปแบบการทำนายตามลำดับขั้นเพื่อใช้ในการทำนายค่าเช่าใช้จักรยาน และผลตอบแทนให้กับแต่ละกลุ่ม ประการแรกใช้อัลกอริทึมการจัดกลุ่มแบบจำลองการผสมแบบเกาส์สองระดับ (Two-Level Gaussian Mixture Model Clustering Algorithm) เพื่อแบ่งสถานีเช่าออกเป็นกลุ่ม ตามการพิจารณาแนวโน้มของจักรยานระหว่างสถานีรวมถึงตำแหน่งทางภูมิศาสตร์ ประการที่สองใช้แผนผังการถดถอยแบบต้นไม้ (Regression Tree) เพื่อทำนายค่าเช่าทั้งหมด ประการที่สามใช้แบบจำลองการอนุมานที่มีความคล้ายคลึงกันหลายรูปแบบ (Multi-Similarity-Based Inference Model) เพื่อคาดการณ์สัดส่วนการคืนและการเปลี่ยนแปลงระหว่างกลุ่ม จากข้างต้นจะมีการสรุปค่าเช่า และผลตอบแทนของจักรยานไปยังแต่ละกลุ่ม เพื่อตรวจสอบประสิทธิภาพของรูปแบบการทำนาย โดยใช้ระบบเช่าจักรยานจากเมืองนิวยอร์กและเมืองวอชิงตัน และทำการเปรียบเทียบผลลัพธ์กับวิธีการอื่น ๆ จากผลการทดลองสามารถสรุปได้ว่าวิธีการที่นำเสนอให้ผลดีกว่าวิธีการอื่น ๆ โดยสามารถลดอัตราข้อผิดพลาด 8% และ 22% ตามลำดับของการตรวจสอบและการทำนายของเมืองนิวยอร์ก 3% และ 2% ตามลำดับของการคืนและการเช่าของเมืองวอร์ชิงตัน เมื่อทำการเปรียบเทียบกับวิธีการอื่น โมเดลที่นำเสนอสามารถลดอัตราข้อผิดพลาดได้ประมาณ 20%และ 30% ตามลำดับในการทำนายการคืนและการเช่าสำหรับสองเมือง

จากทฤษฎีและงานวิจัยที่เกี่ยวข้อง การวิเคราะห์ข้อมูลคุณภาพน้ำ เป็นหนึ่งในประเด็นที่กำลังได้รับความสนใจในการวิจัย เนื่องจากการประสบปัญหาด้านมลพิษและการขาดแคลนทรัพยากรน้ำทั่วโลก จึงจำเป็นที่จะต้องให้ความสนใจอย่างจริงจังกับการปรับปรุงและบำรุงรักษาเพื่อให้เกิดการพัฒนาที่ยั่งยืน รวมทั้งการทำการเกษตรด้านการเพาะเลี้ยงสัตว์น้ำที่ต้องอาศัยคุณภาพน้ำเป็นหลักในการดำเนินการ เพื่อระบุหาสาเหตุ ปัญหาและหาแนวทางแก้ไขปัญหาของคุณภาพน้ำในอนาคต ในงานนี้ผู้วิจัยใช้เทคนิคการวิเคราะข้อมูลคือการจัดกลุ่ม (Clustering) โดยใช้ เทคนิคในการจัดกลุ่ม 3 เทคนิคคือ เทคนิคการจัดกลุ่มแบบเคมีน (K-mean) โดยที่ Hamed [62] ได้ใช้ K-means สำหรับการวิเคราะห์การจัดกลุ่ม เพื่อหาปัจจัยที่สำคัญในการเปลี่ยนแปลงของคุณภาพน้ำ ผลการวิเคราะห์กลุ่มจากสถานีกลุ่มตัวอย่าง 21 แห่งของแม่น้ำไนล์ สามารถแบ่งออกเป็น 3 กลุ่ม ตามความคล้ายคลึงกันของคุณภาพน้ำ เทคนิคการจัดกลุ่มแบบมีนชิป (Mean Shift) Xiao-Tong และคณะ [63] กล่าวถึง

เทคนิคการจัดกลุ่มแบบมีนชิป ว่าเป็นวิธีการจัดกลุ่มแบบไม่อิงพารามิเตอร์ที่มีประสิทธิภาพ และมีการใช้อย่างแพร่หลาย หลักสำคัญคือการใช้ Density Function เป็นตัวสำคัญในการทำ Feature Space Analysis ซึ่งอัลกอลิทึมนี้จะทำการกำหนดจำนวนกลุ่มให้เองโดยอัตโนมัติ เทคนิคการจัดกลุ่มแบบจำลองผสมเกาส์เชียน (Gaussian Mixture Model) Chengchang และคณะ[64] ได้ทำการหาจำนวนกลุ่มที่เหมาะสมของชั้นแนวดิ่งของน้ำทะเลในทะเลจีนใต้ตอนใต้ โดยใช้จำนวนพารามิเตอร์ 4 พารามิเตอร์ และใช้เทคนิควิธี Gaussian Mixture Model และ K-means พบว่าการแบ่งกลุ่มได้จำนวนกลุ่มเป็น 4 กลุ่มตามค่าเกณฑ์จากทั้งสองวิธีการ อีกทั้งงานวิจัยนี้ยังใช้การวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) เข้ามาเข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธีโดยที่ Celestino และคณะ[65] กล่าวว่าการจัดกลุ่มแบบ K-means และการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) สามารถนำไปใช้ในการวิเคราะห์และจัดการคุณภาพน้ำ และได้ทำการทดลองกับพารามิเตอร์จำนวน 28 พารามิเตอร์จากข้อมูลของอุทกธรณีเคมี จากพื้นที่ทั้งหมด 582 แหล่ง ของน้ำบริเวณชายฝั่งทะเลซานโตโดมิงโก ประเทศเม็กซิโก ผลการจัดกลุ่มแบบ K-means+PCA โดยการลดมิติของข้อมูลด้วย PCA สามารถลดตัวแปรอุทกธรณีเคมี เหลือเพียง 16 ตัวแปร และผลของการจัดกลุ่มข้อมูลได้จำนวน 3 กลุ่ม ซึ่งจากที่กล่าวมานี้เป็นที่มาของการเลือกวิธีการจัดกลุ่มด้วยคือ เทคนิคการจัดกลุ่มแบบเคมีน (K-mean) เทคนิคการจัดกลุ่มแบบมีนชิป (Mean Shift) เทคนิควิธี Gaussian Mixture Model และใช้วิธีการวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) เข้ามาเข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธี

บทที่ 3
วิธีการดำเนินการวิจัย

ในการวิจัยเรื่อง การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล ผู้วิจัยได้ดำเนินการศึกษารายละเอียดเกี่ยวกับวิธีดำเนินการวิจัยและได้นำเสนอตามตามผังการดำเนินงาน ดังนี้



ภาพประกอบที่ 3.1 ผังการดำเนินงาน

3.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองของการวิจัยในครั้งนี้ มีจำนวน 3 ชุดข้อมูลประกอบด้วยชุดข้อมูลจากเขื่อนลำปะทาว ชุดข้อมูลจาก Charles River Buoy Data และชุดข้อมูลจาก Fitzroy River Data ซึ่งในการทดลองวิจัยนี้ผู้วิจัยได้นำเอาชุดข้อมูลจำนวน 2 ชุดข้อมูลที่เป็น Open data คือชุดข้อมูลจาก Charles River Buoy Data และชุดข้อมูลจาก Fitzroy River Data มาใช้ร่วมในการทดลองกับชุดข้อมูลของเขื่อนลำปะทาวที่เก็บข้อมูลจริง เพื่อเปรียบเทียบผลของการให้จำนวนกลุ่มระหว่างชุด

ข้อมูลของเขื่อนลำปะทาวกับอีก 2 ชุดข้อมูลเพื่ออธิบาย ว่าชุดข้อมูลที่นำมาไม่ได้ให้ผลการหาจำนวนกลุ่มที่แตกต่างกันมาก ซึ่งรายละเอียดของชุดข้อมูลแสดงรายละเอียดได้ต่อไปนี้

1. ชุดข้อมูลจากเขื่อนลำปะทาว

ชุดข้อมูลจากเขื่อนลำปะทาวเป็นชุดข้อมูลของคุณภาพน้ำได้จากการเก็บข้อมูลด้วย อุปกรณ์ IoT จำนวน 5 พารามิเตอร์ประกอบด้วยค่าปริมาณออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen concentration) ค่าอุณภูมิของน้ำ(Temperature) ค่าความเป็นกรดเป็นเบสของน้ำ(PH) ค่าความขุ่นของน้ำ (Turbidity) และค่าการนำไฟฟ้าของน้ำ(Electric Conductivity) ซึ่งมีหน่วยวัดคุณภาพน้ำดังตารางที่ 3.1 ที่ได้จากการเก็บข้อมูลโดยใช้ทุ่นที่ติดตั้งอุปกรณ์เซนเซอร์ บริเวณเขื่อนลำปะทาว จังหวัดชัยภูมิ ประเทศไทย จำนวน 5 จุด บริเวณที่เกษตรกรเลี้ยงปลานิลในกระชัง และบริเวณริมฝั่งของเขื่อนลำปะทาว ซึ่งได้ทำการเก็บข้อมูลจำนวนทั้งสิ้น 13,608 ชุดข้อมูล ที่จัดเก็บในช่วงเดือนมกราคม-มีนาคม พ.ศ. 2564

ตารางที่ 3.1 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว

พารามิเตอร์	หน่วยวัด
Dissolved Oxygen concentration	milligrams per liter (mg/L)
Temperature	Celsius (°C)
pH	Standard Units
Total Dissolved Solids	Part Per Million (ppm)
Electric Conductivity	percent saturation (%)

2. ชุดข้อมูลจาก Charles River Buoy Data

ชุดข้อมูล Charles River Buoy Data ชุดนี้ได้มาจากเว็บไซต์ของสำนักงานคุ้มครองสิ่งแวดล้อมแห่งสหรัฐอเมริกา (United States Environmental Protection Agency) ที่ได้ติดตั้งทุ่น (Buoy) ไว้ใน The Lower Charles River เพื่อตรวจวัดคุณภาพน้ำ ข้อมูลคุณภาพน้ำได้เผยแพร่ในเว็บไซต์ https://www.epa.gov/charlesriver/live-water-quality-data-lower-charles-river โดย Water Monitoring Buoy ได้ถูกปล่อยให้ลอยอยู่ใน Charles River ใกล้กับ Museum of Science ซึ่งเซนเซอร์ที่ใช้วัดค่าจะอยู่ใต้ผิวน้ำลึกลงไป 1 เมตร เพื่อทำหน้าที่เก็บข้อมูลในทุก ๆ 15 นาที ข้อมูลที่จัดเก็บประกอบไปด้วย 8 พารามิเตอร์ ดังนี้ Water Temperature, Specific Conductance, PH, Dissolved Oxygen Concentration, Dissolved Oxygen % Saturation, Turbidity, Chlorophyll และ Phycoerythrin [66] งานวิจัยนี้ได้เลือกใช้ Charles River Buoy

Data จำนวนทั้งสิ้น 28,116 ชุดซึ่งจัดเก็บในช่วงปี ค.ศ. 2018-2019 รายละเอียดของพารามิเตอร์แสดงดังตารางที่ 3.2

ตารางที่ 3.2 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดของ Charles River Buoy Data

พารามิเตอร์	**หน่วยวัด**
Temperature	Celsius (°C)
Specific conductance	millisiemens per centimeter (mS/cm)
pH	Standard Units
Dissolved Oxygen concentration	milligrams per liter (mg/L)
Dissolved Oxygen % saturation	percent saturation (%)
Turbidity	Formazin Nephelometric Units (FNUs)
Chlorophyll	Raw Fluorescent Units (RFUs)
Phycoerythrin	Raw Fluorescent Units (RFUs)

3. ชุดข้อมูลจาก Fitzroy River Data

ข้อมูล Fitzroy River Data ชุดนี้เป็น Open data ที่จัดเก็บและเผยแพร่โดย Queensland Government ประเทศออสเตรเลีย โดยเป็นการวัดคุณภาพน้ำจาก Fitzroy river, Central coast Queensland เป็นข้อมูลที่จัดเก็บในช่วงปี ค.ศ. 1993-2003 ประกอบด้วย 11,014 ชุดข้อมูล โดยได้เผยแพร่ผ่านเว็บไซต์ https://www.data.qld.gov.au ซึ่งประกอบไปด้วย 7 พารามิเตอร์ ประกอบด้วย Dissolved Oxygen concentration, Dissolved Oxygen % saturation, pH, Salinity, Specific Conductance และ Temperature[67] รายละเอียดของพารามิเตอร์แสดงดังตารางที่ 3.3

ตารางที่ 3.3 พารามิเตอร์ของคุณภาพน้ำและหน่วยวัดของ Fitzron river data

พารามิเตอร์	**หน่วยวัด**
Dissolved Oxygen (DO)	milligrams per liter (mg/L)
Dissolved Oxygen (DO)	percent saturation (%)
pH	Standard Units
Salinity	practical salinity Units (PSU)
Specific conductance	millisiemens per centimeter (mS/cm)
Temperature	Celsius (°C)
Turbidity	Formazin Nephelometric Units (FNUs)

3.2 เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

ในการศึกษาวิจัยครั้งนี้ทำการเก็บรวบรวมข้อมูลคุณภาพน้ำบริเวณเขื่อนลำปะทาว จังหวัดชัยภูมิและเพื่อหาโมเดลทางการเรียนรู้ของเครื่องและนำมาใช้วิเคราะห์กลุ่มข้อมูลคุณภาพน้ำ โดยการออกแบบและพัฒนาอุปกรณ์ระบบ IoT ในการวัดข้อมูลคุณภาพน้ำ ประกอบด้วยอุปกรณ์และเซ็นเซอร์ในการตรวจวัดข้อมูลดังต่อไปนี้

1) NodeMcu ESP8266 เป็นแพลตฟอร์มแบบซอร์ฟแวร์รหัสเปิดซึ่งเป็นชื่อเรียกของชิฟโมดูล ESP8266 สำหรับติดต่อสื่อสารบนมาตรฐานการทำงานแบบไร้สาย ทำงานที่แรงดันไฟฟ้า 3.0-3.6V ใช้กระแสโดยเฉลี่ย 80 mA สามารถรองรับคำสั่ง Deep Sleep ในการประหยัดพลังงาน ใช้กระแสน้อยกว่า 10 ไมโครแอมป์ สามารถ Wake Up กลับมาส่งข้อมูลใช้เวลาน้อยกกว่า 2 มิลลิวินาที ภายในมี Low Power MCU 32bit ทำให้สามารถเขียนโปรแกรมสั่งงานได้ มีวงจร Analog Digital Converter ทำให้สามารถอ่านค่าแบบอนาล็อคได้ ทำงานได้ที่อุณหภูมิ -40 ถึง 125 องศาเซลเซียส

2) อุปกรณ์เซนเซอร์วัดปริมาณออกซิเจนที่ละลายในน้ำ (DO Analog Dissolved Oxygen Sensor Meter) เป็นเซนเซอร์ที่ใช้ในการวัดค่าปริมาณออกซิเจนที่ละลายในน้ำ สำหรับตรวจวัดคุณภาพน้ำ โดยการทำงานเป็นการส่งค่าแบบอนาล็อค ใช้หัววัดค่าประเภท Galvanic Probe แบบ BNC ช่วงการตรวจจับอยู่ที่ 0-20 มิลลิกรัมต่อลิตร ใช้แรงดันไฟฟ้า 3.3-5.5 โวลต์ สัญญาณเอาต์พุต 0-3.0 โวลต์

ภาพประกอบที่ 3.2 เซนเซอร์วัดปริมาณออกซิเจนที่ละลายในน้ำ [68]

3) อุปกรณ์เซนเซอร์วัดค่าอุณภูมิของน้ำ (DS18B20 Waterproof Temperature Sensor) เป็นเซนเซอร์ที่ใช้ในการวัดค่าอุณภูมิของน้ำ ส่งค่าเอาท์พุตแบบดิจิตอลเซนเซอร์ ใช้ชิปโพรบเซ็นเซอร์ในการวัดอุณหภูมิ ช่วงอุณหภูมิในการทำงานอยู่ที่ -55 องศาเซลเซียส ถึง + 125

องศาเซลเซียส ใช้แรงดันไฟฟ้า 3.0- 5.5 โวลต์ ขนาดท่อสแตนเลส 6 × 45 มิลลิเมตร ความละเอียดตั้งโปรแกรมได้ตั้งแต่ 9 ถึง 12 บิต ค่าความถูกต้องอยู่ในช่วง -10 ถึง 85 องศาเซลเซียส: ± 0.5 องศาเซลเซียส ช่วงอุณหภูมิการจัดเก็บอยู่ที่ -55 องศาเซลเซียส ถึง + 125 องศาเซลเซียส

ภาพประกอบที่ 3.3 เซนเซอร์วัดค่าอุณภูมิของน้ำ [69]

4) อุปกรณ์เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ (PH Meter E-201-C PH Module) เป็นเซนเซอร์ที่ใช้ในการวัดค่าความเป็นกรดเป็นเบสของน้ำ ใช้แรงดันไฟฟ้า 5 โวลต์ ช่วงการวัดค่าความเป็นกรดเป็นเบสอยู่ที่ 0-14 ช่วงอุณภูมิในการวัดค่าอยู่ที่ 0-60 องศาเซลเซียส ค่าความแม่นยำ: ± 0.1pH (25 องศาเซลเซียส)

ภาพประกอบที่ 3.4 เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ[70]

5) อุปกรณ์เซนเซอร์วัดค่าความขุ่นของน้ำ (EC Sensor Turbidity) เป็นเซนเซอร์สำหรับใช้วัดระดับความขุ่นของน้ำโดยใช้หลักการตรวจสอบด้วยแสง ใช้แรงดันไฟฟ้า 5 โวลต์ อุณหภูมิในการทำงานอยู่ที่ระหว่าง 5-90 องศาเซลเซียส ส่งค่าแบบอนาล็อค

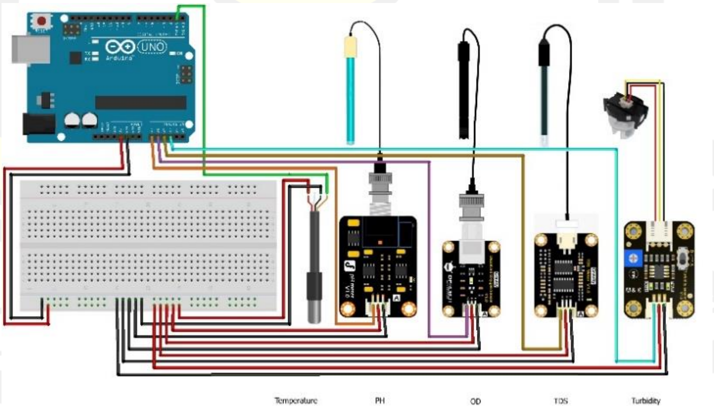
ภาพประกอบที่ 3.5 เซนเซอร์วัดค่าความขุ่นของน้ำ[71]

6) อุปกรณ์เซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ (Analog EC Sensor TDS) เป็นเซ็นเซอร์วัดค่าการนำไฟฟ้าของน้ำ เพื่อตรวจสอบคุณภาพน้ำ ใช้แรงดันไฟฟ้าระหว่าง 3.3-5.5 โวลต์ ความแม่นยำในการวัด ± 10% FS (25 องศาเซลเซียส)

ภาพประกอบที่ 3.6 เซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ[72]

โครงสร้างการทำงานของระบบจัดเก็บข้อมูลคุณภาพน้ำ

การพัฒนาระบบจัดเก็บข้อมูลคุณภาพน้ำ โดยใช้เซนเซอร์ในการวัดคุณภาพน้ำ เชื่อมต่อกับบอร์ดไมโครคอนโทรลเลอร์ โดยใช้เซนเซอร์จำนวน 5 ตัวคือ เซนเซอร์วัดค่าปริมาณออกซิเจนที่ละลายน้ำ (Dissolved Oxygen) เซนเซอร์วัดค่าความเป็นกรดเป็นเบสของน้ำ (PH) เซนเซอร์วัดค่าอุณหภูมิของน้ำ(Temperature) เซนเซอร์วัดค่าความขุ่น (Turbidity) และเซนเซอร์วัดค่าการนำไฟฟ้าของน้ำ (Total Dissolved Solids) ส่งข้อมูลไปจัดเก็บยังอุปกรณ์บันทึกข้อมูล (Memory Card) ดังภาพประกอบที่ 3.7



ภาพประกอบที่ 3.7 ไดอะแกรมของอุปกรณ์ที่ใช้ในการจัดเก็บข้อมูล

3.3 การเก็บรวบรวมข้อมูล

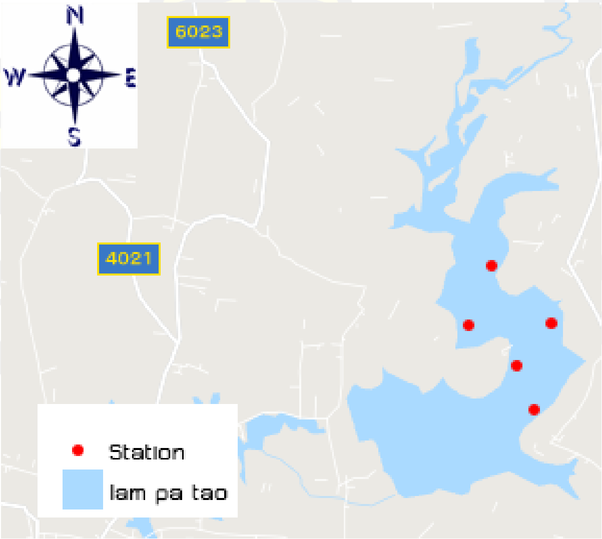
3.3.1 ระยะเวลาในการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการศึกษาวิจัยครั้งนี้เก็บข้อมูลโดยใช้ทุ่นอุปกรณ์ติดตั้งเซนเซอร์ที่ใช้ในการตรวจวัดค่า จำนวน 5 ประเภทที่ใช้วัดค่าดังต่อไปนี้นี้ ค่าปริมาณออกซิเจนที่ละลายในน้ำ ค่าอุณภูมิของน้ำ ค่าความเป็นกรดเป็นเบสของน้ำ ค่าความขุ่นของน้ำ ค่าการนำไฟฟ้าของน้ำ จำนวนทั้งสิ้น 13,608 ชุดข้อมูล ดังภาพประกอบที่ 3.8 ที่จัดเก็บในช่วงเดือนมกราคม-มีนาคม พ.ศ. 2564

ภาพประกอบที่ 3.8 ทุ่นอุปกรณ์ที่ใช้สำหรับเก็บข้อมูลจากเขื่อนลำปะทาวและส่วนประกอบอุปกรณ์เซนเซอร์สำหรับจัดเก็บข้อมูลคุณภาพน้ำ

3.3.2 พื้นที่ในการเก็บข้อมูลคุณภาพน้ำ

บริเวณพื้นที่ในการจัดเก็บคุณภาพน้ำ บริเวณเขื่อนลำปะทาว อำเภอแก้งคร้อ จังหวัดชัยภูมิ จำนวน 5 จุด ดังภาพประกอบที่ 3.9



ภาพประกอบที่ 3.9 จุดในการจัดเก็บคุณภาพน้ำบริเวณพื้นที่เขื่อนลำปะทาว

3.4 การสร้างแบบจำลอง

นำชุดข้อมูลที่ได้จากการเก็บข้อมูลคุณภาพน้ำและข้อมูลจากแหล่งเปิดเผยข้อมูลคุณภาพน้ำ มาสร้างแบบจำลอง โดยนำชุดข้อมูลทั้ง 3 ชุดข้อมูลมาสร้างแบบจำลองเพื่อหาแบบจำลองที่เหมาะสมที่สุดในการจัดกลุ่มซึ่งผู้วิจัยได้เลือกเทคนิคในการจัดกลุ่ม 3 เทคนิคคือ เทคนิคการจัดกลุ่มแบบเคมีน (K-mean) เทคนิคการจัดกลุ่มแบบมีนชิป (Mean Shift) เทคนิคการจัดกลุ่มแบบจำลองผสมเกาส์เชียน (Gaussian Mixture Model) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธี

3.5 การวิเคราะห์ข้อมูลและการประเมินประสิทธิภาพของแบบจำลอง

เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและการประเมินประสิทธิภาพของแบบจำลอง มีดังต่อไปนี้

1) ในการวิเคราะห์ข้อมูล เป็นการวิเคราะห์ชุดข้อมูลคุณภาพน้ำจาก 3 แหล่งข้อมูล คือ ชุดข้อมูลแรกจากเขื่อนลำปะทาว จังหวัดชัยภูมิ โดยข้อมูลได้จากการเก็บข้อมูลจากทุ่นอุปกรณ์ จำนวน 5 พารามิเตอร์ คือ ค่าปริมาณออกซิเจนที่ละลายในน้ำ(Dissolved Oxygen Concentration) ค่าอุณภูมิของน้ำ(Temperature) ค่าความเป็นกรดเป็นเบสของน้ำ(PH) ค่าความขุ่นของน้ำ(Turbidity) และค่าการนำไฟฟ้าของน้ำ(Electric Conductivity) จำนวน 13,608 ชุดข้อมูล ชุดข้อมูลที่ 2 จากชุดข้อมูลของ Charles River Buoy Data ที่เป็นชุดข้อมูล open data ประกอบไปด้วย 8 พารามิเตอร์ ดังนี้ Water Temperature, Specific Conductance, PH, Dissolved Oxygen Concentration, Dissolved Oxygen % Saturation, Turbidity, Chlorophyll จำนวนทั้งสิ้น 28,116 ชุด และชุดข้อมูลสุดท้าย จากFitzroy River Dataที่เป็น Open data ที่ จำนวน 11,014 ชุดข้อมูล ซึ่งประกอบไปด้วย 7 พารามิเตอร์ ประกอบด้วย Dissolved Oxygen Concentration, Dissolved Oxygen % Saturation, PH, Salinity, Specific Conductance และ Temperature

2) เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ใช้ Google Colab

3) ในการวิเคราะห์ข้อมูลของการวิจัยครั้งนี้ เป็นการแบ่งกลุ่มคุณภาพน้ำว่าควรแบ่งออกเป็นกี่กลุ่ม โดยใช้ชุดข้อมูลจาก 3 แหล่งข้อมูลตามที่ได้กล่าวมาในข้างต้น แล้วนำเอาข้อมูลทั้งหมดมาวิเคราะห์ เพื่อให้ทราบถึงจำนวนกลุ่มของคุณภาพน้ำ โดยใช้เทคนิคการจัดกลุ่มแบบเคมีน (k-mean) เทคนิคการจัดกลุ่มแบบมีนชิป (Mean Shift) เทคนิคการจัดกลุ่มแบบจำลองผสมเกาส์เชียน (Gaussian Mixture Model) และนำวิธีการการวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 เทคนิควิธี

4) ในการประเมินประสิทธิภาพของแบบจำลองเพื่อประเมินผลการจัดกลุ่มที่ไม่ทราบกลุ่มที่แท้จริงมาก่อนโดยใช้การประเมินประสิทธิภาพ [73] ของการประเมินจำนวนกลุ่มที่เหมาะสม Optimal (k)

บทที่ 4
ผลการดำเนินการวิจัย

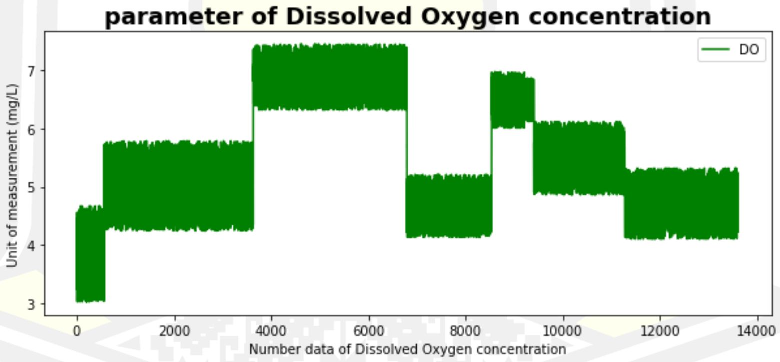
ผลการดำเนินการวิจัยประกอบด้วย การหาจำนวนกลุ่มของคุณภาพน้ำ ผลของการจัดกลุ่มคุณภาพน้ำ จากชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว ชุดข้อมูล Charles River Buoy Data และชุดข้อมูล Fitzroy River Data ตามลำดับต่อไปนี้

4.1 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว

การทดลองเพื่อหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มคุณภาพน้ำจากชุดข้อมูลที่ได้จากการเก็บข้อมูลบริเวณเขื่อนลำปะทาว จังหวัดชัยภูมิ จำนวน 13,608 ชุดข้อมูลสามารถแสดงค่าพารามิเตอร์ของคุณภาพน้ำในรูปแบบของกราฟ ดังต่อไปนี้

4.1.1 พารามิเตอร์ค่าออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen concentration)

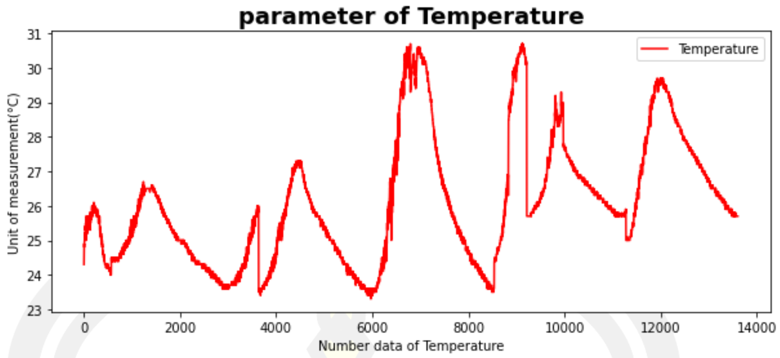
พารามิเตอร์ค่าออกซิเจนที่ละลายในน้ำ (Dissolved Oxygen concentration) หรือค่า DO ที่มีหน่วยวัดพารามิเตอร์เป็น milligrams per liter (mg/L) จากค่าของพารามิเตอร์นี้สามารถแสดงค่าของข้อมูล ดังภาพที่ 4.1



ภาพประกอบที่ 4.1 กราฟแสดงข้อมูลค่าออกซิเจนที่ละลายในน้ำ
จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว

4.1.2 พารามิเตอร์ค่าอุณหภูมิ (Temperature)

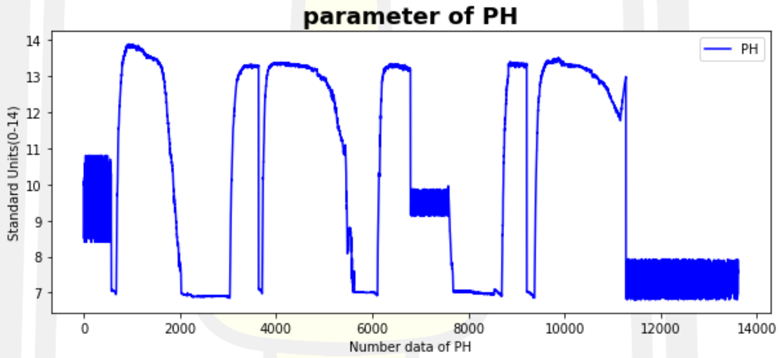
พารามิเตอร์ค่าอุณหภูมิ (Temperature) ที่มีหน่วยวัดพารามิเตอร์เป็นองศาเซลเซียส Celsius (°C) จากค่าของพารามิเตอร์นี้สามารถแสดงค่าของข้อมูล ดังภาพที่ 4.2



ภาพประกอบที่ 4.2 กราฟแสดงข้อมูลค่าอุณหภูมิ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว

4.1.3 พารามิเตอร์ค่า PH

พารามิเตอร์ค่า PH ที่มีหน่วยวัดพารามิเตอร์ของความเป็นกรดเป็นด่างอยู่ระหว่างช่วง 1-14 จากค่าของพารามิเตอร์นี้สามารถแสดงค่าของข้อมูล ดังภาพที่ 4.3

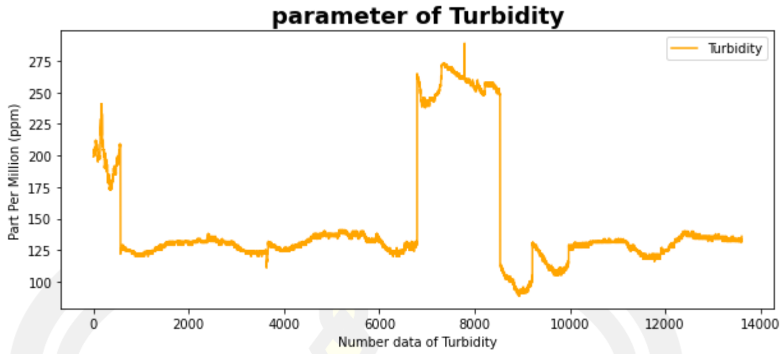


ภาพประกอบที่ 4.3 กราฟแสดงข้อมูลค่า PH จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว

4.1.4 พารามิเตอร์ค่าความขุ่นของน้ำ (Turbidity)

พารามิเตอร์ค่าความขุ่นของน้ำ (Turbidity) ที่มีหน่วยวัดพารามิเตอร์ของความขุ่นของน้ำ เป็น Part Per Million (ppm) จากค่าของพารามิเตอร์นี้สามารถแสดงค่าของข้อมูล ดังภาพที่ 4.4

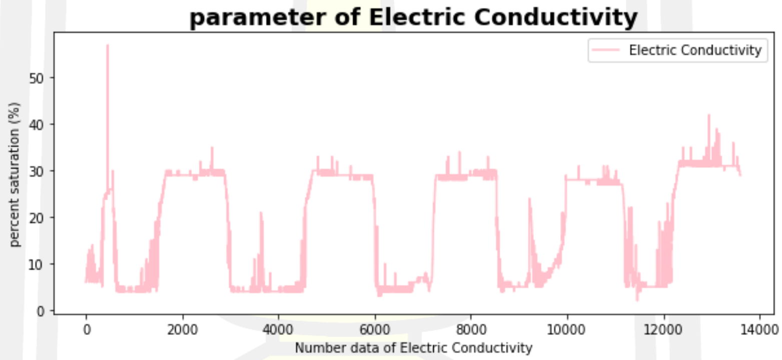
พหุน ปณ ฑิโต ชีเว



ภาพประกอบที่ 4.4 กราฟแสดงข้อมูลค่าความขุ่นของน้ำ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว

4.1.5 พารามิเตอร์ค่าการนำไฟฟ้าของน้ำ(Electric Conductivity)

พารามิเตอร์ค่าการนำไฟฟ้าของน้ำ(Electric Conductivity) มีหน่วยวัดพารามิเตอร์เป็นเปอร์เซ็น (percent saturation (%)) จากค่าของพารามิเตอร์นี้สามารถแสดงค่าของข้อมูล ดังภาพที่ 4.5



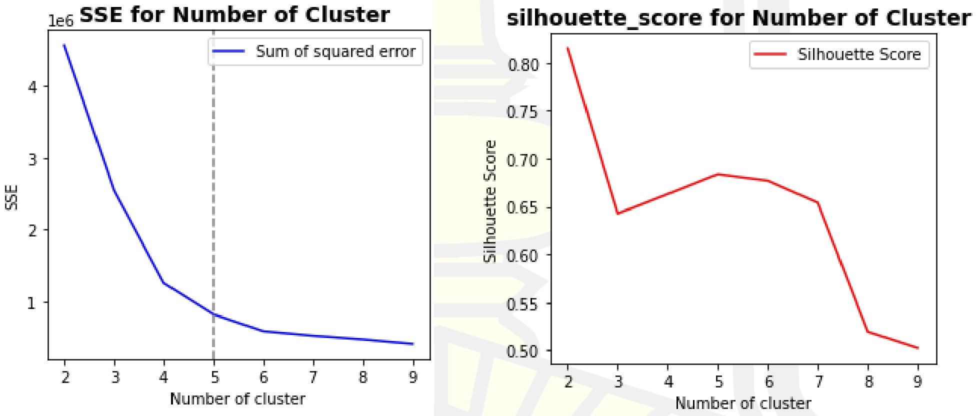
ภาพประกอบที่ 4.5 กราฟแสดงข้อมูลค่าการนำไฟฟ้าของน้ำ จากชุดข้อมูลคุณภาพน้ำเขื่อนลำปะทาว

จากการแสดงค่าของพารามิเตอร์ทั้ง 5 พารามิเตอร์ของชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาวขั้นตอนต่อไปจะเป็นการพิจารณาการหาจำนวนกลุ่มจาก 3 วิธีการ คือ K-Means , MeanShift และ Gaussian Mixture Model (GMM) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 วิธีการ ดังต่อไปนี้

ในการวิจัยนี้การหาจำนวนกลุ่มของวิธีการ K-Means จะใช้การพิจารณาจากการหักศอก ซึ่งได้จากค่า Sum of Squared Error และการพิจารณาค่าของ Silhouette Score มาใช้ในการพิจารณาหาจำนวนกลุ่มร่วมกัน[74] การทดลองด้วยวิธีการ K-means จากการพิจารณาจากกราฟ ที่ได้จากค่า Sum of Squared Error โดยเมื่อค่าของความผิดพลาดลดน้อยลง เส้นที่มีความชันจะเริ่มโค้งและราบเรียบ จนเกิดเป็นมุมลักษณะเหมือนการหักศอก ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่ม ที่ดีที่สุด[75] และในกรณีเดียวกัน การพิจารณาค่า Silhouette Score เป็นการหาจำนวนกลุ่มที่

เหมาะสมเช่นกัน โดยค่าที่ได้จะอยู่ในช่วง -1 ถึง 1 ถ้ามีค่าเข้าใกล้ 1 แสดงว่าจำนวนกลุ่มมีความเหมาะสม[76] ซึ่งในการทดลองนี้จะหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากเขื่อนลำปะทาว โดยพิจารณาจากกราฟ ของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.6a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.6b ในการพิจารณากลุ่มร่วมกัน จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.6a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 4 กลุ่ม 5 กลุ่ม และ 6 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.6b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 2กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.8151 5กลุ่ม ค่า Silhouette Score อยู่ที่ 0.6839 และ 6กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.6771 ตามลำดับ

ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากเขื่อนลำปะทาวเมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.6a ได้จำนวนกลุ่มเท่ากับ 5 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6839 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.6b

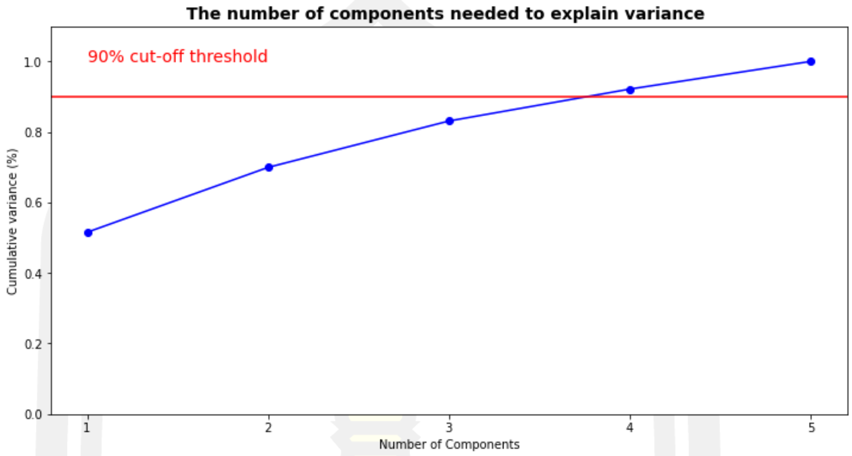


a) b)

ภาพประกอบที่ 4.6 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-Means กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 80% ตามที่ Li และคณะได้กล่าวเอาไว้[77] พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสม

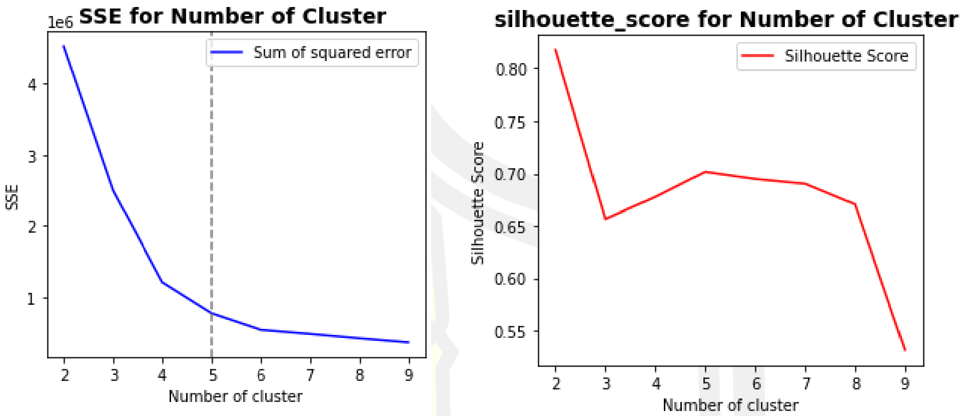
อยู่เท่ากับ 3 Component โดยค่าความแปรปรวน (Explained Variance) อยู่ที่ 90% ดังภาพประกอบที่ 4.7



ภาพประกอบที่ 4.7 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูลเขื่อนลำปะทาว จังหวัดชัยภูมิ

หลังจากการหาจำนวน Component ได้จำนวน 3 Component แล้วทำการพิจารณาหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากเขื่อนลำปะทาวโดยพิจารณาจากกราฟ ของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.8a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.8b จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.8a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 4 กลุ่ม 5 กลุ่ม และ 6 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.8b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่จำนวน 2กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.8176 5กลุ่ม ค่า Silhouette Score อยู่ที่ 0.7015 และ 6กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.6947 ตามลำดับ

ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากเขื่อนลำปะทาวเมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.8a ได้จำนวนกลุ่มเท่ากับ 5 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.7015 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.8b



a) b)

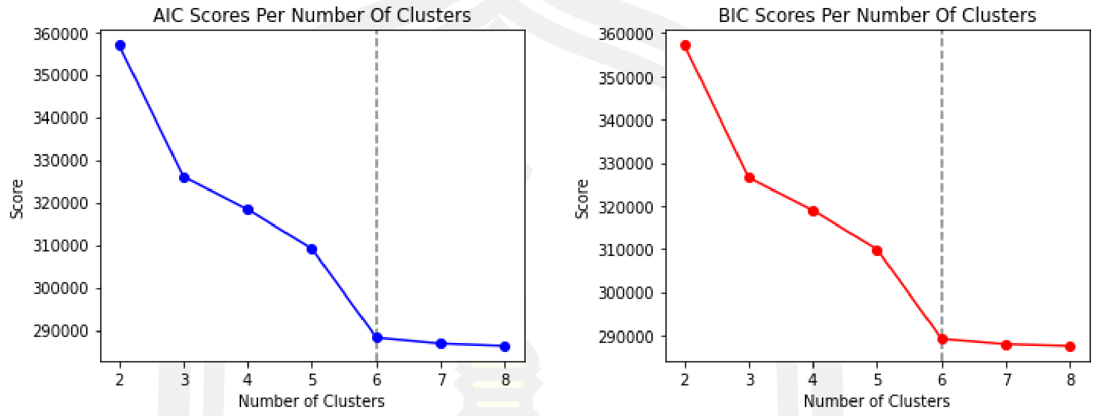
ภาพประกอบที่ 4.8 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-Means กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

การหาจำนวนกลุ่ม ด้วย Mean Shift เป็นวิธีการหาจำนวนกลุ่มแบบอัตโนมัติ โดยอาศัยการหาจำนวนกลุ่มแบบวนซ้ำ ซึ่งไม่อ้างอิงค่าพารามิเตอร์ เหมาะสำหรับการหาจำนวนกลุ่มที่มีรูปลักษณ์แบบไม่ตายตัว ซึ่งจำนวนกลุ่ม (Cluster) ของวิธีนี้จะให้ผลลัพธ์ออกมาแบบอัตโนมัติ โดยการค้นหาจุดศูนย์กลางของบริเวณที่หนาแน่นที่สุด ซึ่งเป็นวิธีที่ไม่ใช้พารามิเตอร์ในการประมาณการฟังก์ชันในการสุ่มตัวแปร [78]

ในการทดลองนี้ผลการทดลองการหาจำนวนกลุ่มด้วยวิธีการ Mean Shift กับชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว จะให้คำตอบออกมาเป็นจำนวนค่าตัวเลข ซึ่งสามารถบ่งบอกถึงจำนวนกลุ่มที่หาออกมาได้โดยอัตโนมัติ พบว่าได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 80% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 3 Component แล้วหาจำนวนกลุ่มด้วย Mean shift ซึ่งได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม

การหาจำนวนกลุ่มด้วย การจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian Mixture Model) หรือ GMM กับชุดข้อมูลคุณภาพน้ำจากเขื่อนลำปะทาว จังหวัดชัยภูมิ โดยการพิจารณาจากค่าของ AIC (Akaike Information Criterion) และ BIC (Bayesian Information Criterion) โดยการวัดหาจำนวนกลุ่มทั้งสองวิธีการนี้มักจะเลือกจำนวนกลุ่มเดียวกันสำหรับการหาจำนวนกลุ่ม แต่ถ้าในกรณีที่ค่าต่างกันให้พิจารณาจากค่า BIC ก่อน โดยพิจารณาจากกราฟที่มีค่าต่ำสุดหรือมีการลดระดับลงและมีลักษณะเส้นกราฟคงที่[79] จากการทอลองพบว่าการหาจำนวนกลุ่มด้วยวิธี GMM ได้

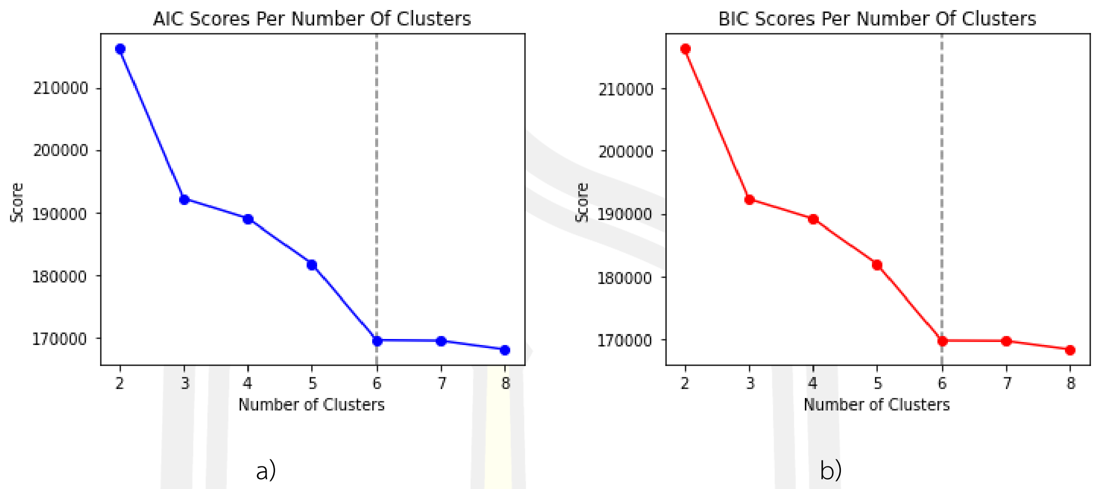
จำนวนกลุ่ม เท่ากับ 6 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.9a ตามเส้นประและการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.9b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC



a) b)

ภาพประกอบที่ 4.9 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า AIC b) ค่า BIC

จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 80% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 3 Component ซึ่งได้ทำการหาจำนวน Component ไปแล้วนั้น แล้วจึงทำการหาจำนวนกลุ่มด้วยวิธีการ GMM พบว่า ได้จำนวนกลุ่มเท่ากับ 6 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.10a ตามเส้นประและการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.10b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC



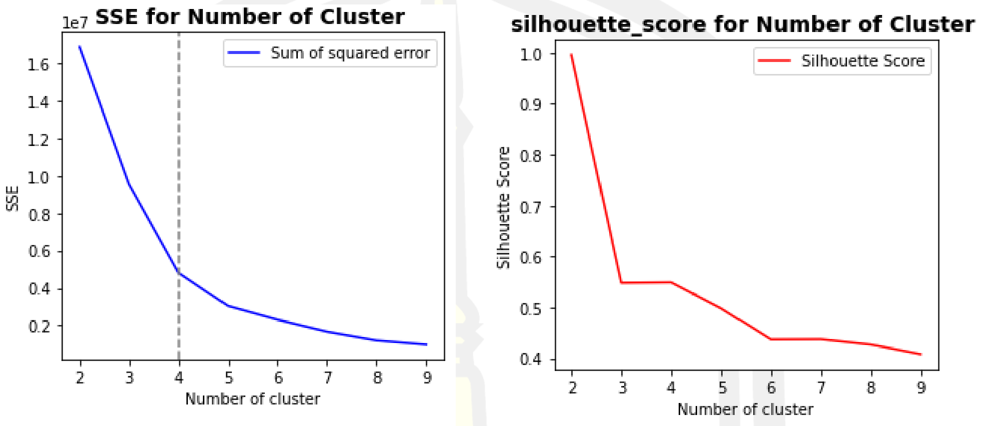
ภาพประกอบที่ 4.10 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูลของเขื่อนลำปะทาว a) ค่า AIC b) ค่า BIC

4.2 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำ Charles River Buoy Data

การทดลองหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data โดยการพิจารณาจาก 3 วิธีการ คือ K-means , MeanShift และ Gaussian Mixture Model (GMM) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 วิธีการ ผลการทดลองดังต่อไปนี้

ในการวิจัยนี้การหาจำนวนกลุ่มของวิธีการ K-means จะใช้การพิจารณาจากการหักศอก ซึ่งได้จากค่า Sum of Squared Error และการพิจารณาค่าของ Silhouette Score มาใช้ในการพิจารณาหาจำนวนกลุ่มร่วมกัน[70] การทดลองด้วยวิธีการ K-Means จากการพิจารณาจากกราฟ ที่ได้จากค่า Sum of Squared Error โดยเมื่อค่าของความผิดพลาดลดน้อยลง เส้นที่มีความชันจะเริ่มโค้งและราบเรียบ จนเกิดเป็นมุมลักษณะเหมือนการหักศอก ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่ม ที่ดีที่สุด[71] และในกรณีเดียวกัน การพิจารณาค่า Silhouette Score เป็นการหาจำนวนกลุ่มที่เหมาะสมเช่นกัน โดยค่าที่ได้จะอยู่ในช่วง -1 ถึง 1 ถ้ามีค่าเข้าใกล้ 1 แสดงว่าจำนวนกลุ่มมีความเหมาะสม[72] ซึ่งในการทดลองนี้จะหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data โดยพิจารณาจากกราฟ ของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.11a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.11b ในการพิจารณากลุ่มร่วมกัน จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.11a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 3 กลุ่ม 4 กลุ่ม และ 5 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.6b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 2กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.9956 3กลุ่ม ค่า Silhouette Score อยู่ที่ 0.5480 และ 4กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.5489 ตามลำดับ

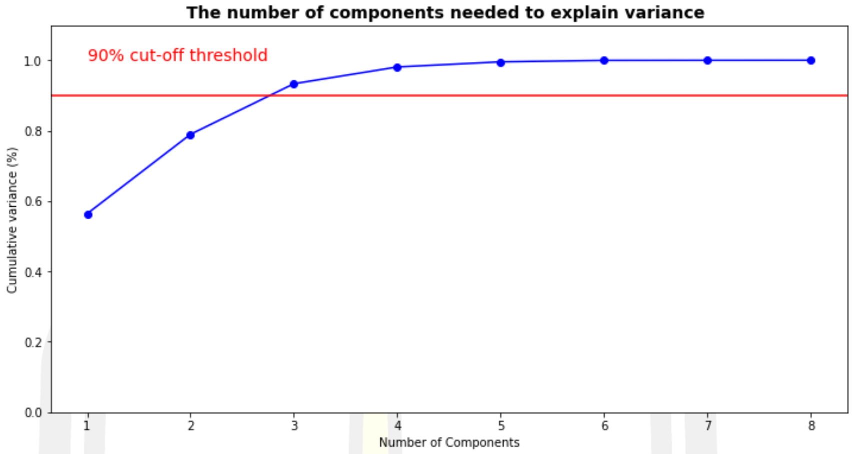
ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data เมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.11a ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.5489 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.11b



a) b)

ภาพประกอบที่ 4.11 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-means กับชุดข้อมูล Charles River Buoy Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

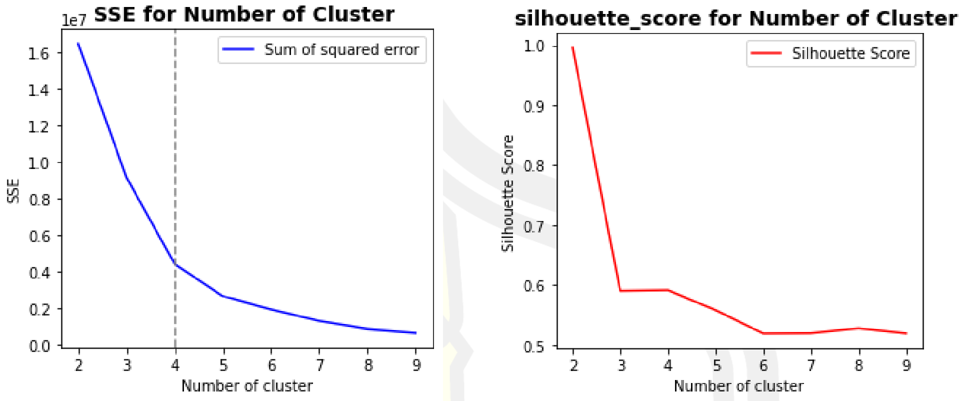
จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% ตามที่ Li และคณะได้กล่าวเอาไว้[73] พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component โดยค่าความแปรปรวน (explained variance) อยู่ที่ 90% ดังภาพประกอบที่ 4.12



ภาพประกอบที่ 4.12 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูล Charles River Buoy Data

หลังจากการหาจำนวน Component ได้จำนวน 2 Component แล้วทำการพิจารณาหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำชุดข้อมูล Charles River Buoy Data โดยพิจารณาจากกราฟของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.13a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.13b จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.13a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 3 กลุ่ม 4 กลุ่ม และ 5 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.13b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่จำนวน 2กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.9959 3กลุ่ม ค่า Silhouette Score อยู่ที่ 0.5897 และ 4กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.5909 ตามลำดับ

ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data เมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.13a ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.5909 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.13b

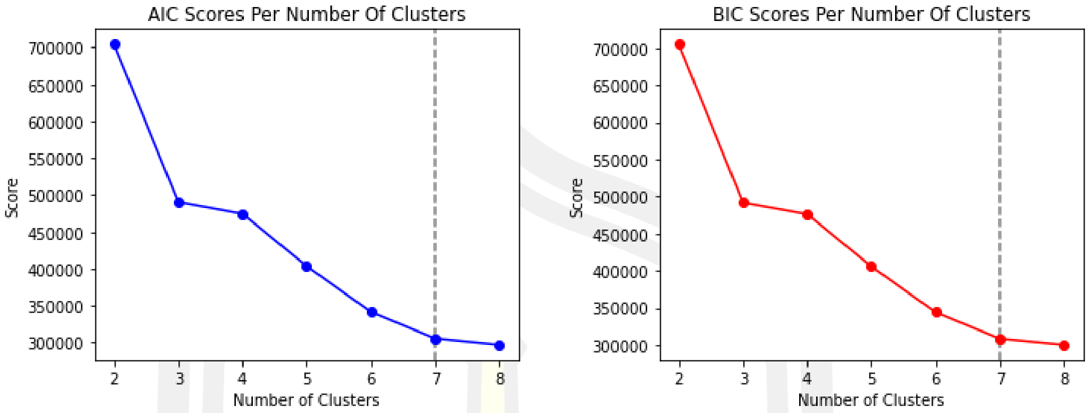


a) b)

ภาพประกอบที่ 4.13 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-means กับชุดข้อมูล Charles River Buoy Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

การหาจำนวนกลุ่ม ด้วย Mean shift เป็นวิธีการหาจำนวนกลุ่มแบบอัตโนมัติ โดยอาศัยการหาจำนวนกลุ่มแบบวนซ้ำ ซึ่งไม่อ้างอิงค่าพารามิเตอร์ เหมาะสำหรับการหาจำนวนกลุ่มที่มีรูปลักษณ์แบบไม่ตายตัว ซึ่งจำนวนกลุ่ม (Cluster) ของวิธีนี้จะให้ผลลัพธ์ออกมาแบบอัตโนมัติ โดยการค้นหาจุดศูนย์กลางของบริเวณที่หนาแน่นที่สุด ในการทดลองนี้ผลการทดลองการหาจำนวนกลุ่มด้วยวิธีการ Mean shift กับชุดข้อมูลคุณภาพน้ำจาก Charles River Buoy Data จะให้คำตอบออกมาเป็นจำนวนค่าตัวเลข ซึ่งสามารถบ่งบอกถึงจำนวนกลุ่มที่หาออกมาได้โดยอัตโนมัติ พบว่าได้จำนวนกลุ่มเท่ากับ 7 กลุ่ม จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component แล้วหาจำนวนกลุ่มด้วย Mean shift ซึ่งได้จำนวนกลุ่มเท่ากับ 9 กลุ่ม

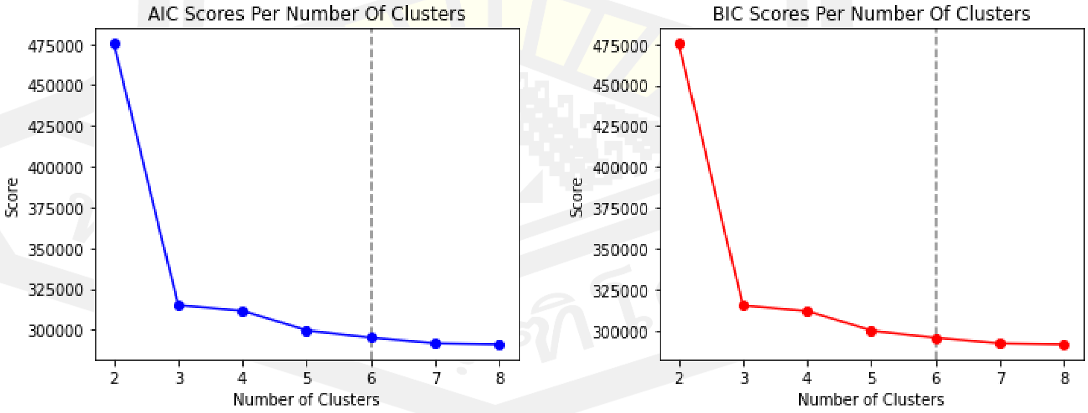
การหาจำนวนกลุ่มด้วย การจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian mixture model) หรือ GMM กับชุดข้อมูลคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data โดยการพิจารณาจากค่าของ AIC และ BIC โดยการวัดหาจำนวนกลุ่มทั้งสองวิธีการนี้มักจะเลือกจำนวนกลุ่มเดียวกันสำหรับการหาจำนวนกลุ่ม แต่ถ้าในกรณีที่ค่าต่างกันให้พิจารณาจากค่า BIC ก่อน โดยพิจารณาจากกราฟที่มีค่าต่ำสุดหรือมีการลดระดับลงและมีลักษณะเส้นกราฟคงที่[75] จากการทดลองพบว่าการหาจำนวนกลุ่มด้วยวิธี GMM ได้จำนวนกลุ่ม เท่ากับ 7 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.14a ตามเส้นประและการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.14b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC



a) b)

ภาพประกอบที่ 4.14 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูล Charles River Buoy Data a) ค่า AIC b) ค่า BIC

จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component ซึ่งได้ทำการหาจำนวน Component ไปแล้วนั้น แล้วจึงทำการหาจำนวนกลุ่มด้วยวิธีการ GMM พบว่า ได้จำนวนกลุ่มเท่ากับ 6 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.10a ตามเส้นประและการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.10b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC



a) b)

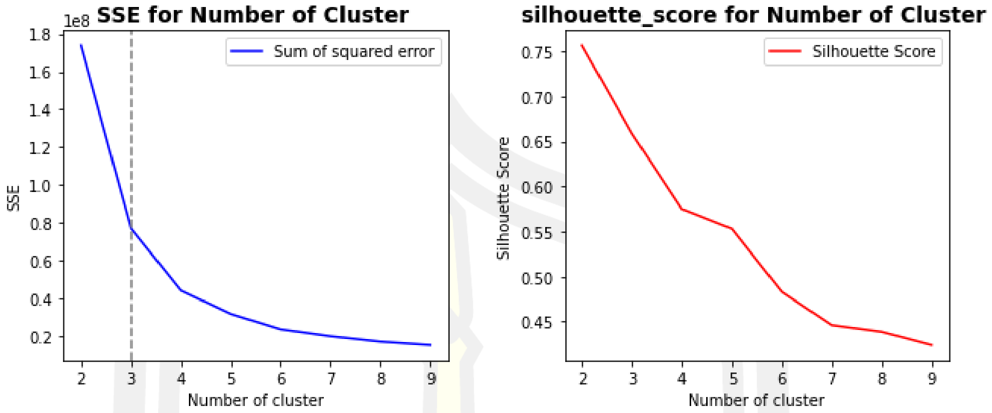
ภาพประกอบที่ 4.15 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูล Charles River Buoy Data a) ค่า AIC b) ค่า BIC

4.3 ผลของการจัดกลุ่มคุณภาพน้ำ ชุดข้อมูลคุณภาพน้ำ Fitzroy River Data

การทดลองหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data โดยการพิจารณาจาก 3 วิธีการ คือ K-means , MeanShift และ Gaussian Mixture Model (GMM) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 วิธีการ ผลการทดลองดังต่อไปนี้

ในการวิจัยนี้การหาจำนวนกลุ่มของวิธีการ K-means จะใช้การพิจารณาจากการหักศอก ซึ่งได้จากค่า Sum of squared error และการพิจารณาค่าของ Silhouette Score มาใช้ในการพิจารณาหาจำนวนกลุ่มร่วมกัน[70] การทดลองด้วยวิธีการ K-Means จากการพิจารณาจากกราฟ ที่ได้จากค่า Sum of Squared Error โดยเมื่อค่าของความผิดพลาดลดน้อยลง เส้นที่มีความชันจะเริ่มโค้งและราบเรียบ จนเกิดเป็นมุมลักษณะเหมือนการหักศอก ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่ม ที่ดีที่สุด[71] และในกรณีเดี่ยวกัน การพิจารณาค่า Silhouette Score เป็นการหาจำนวนกลุ่มที่เหมาะสมเช่นกัน โดยค่าที่ได้จะอยู่ในช่วง -1 ถึง 1 ถ้ามีค่าเข้าใกล้ 1 แสดงว่าจำนวนกลุ่มมีความเหมาะสม[72] ซึ่งในการทดลองนี้จะหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data โดยพิจารณาจากกราฟ ของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.16a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.16b ในการพิจารณากลุ่มร่วมกัน จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.16a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 3 กลุ่ม 4 กลุ่ม และ 6 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.16b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 2กลุ่ม โดยมีค่า Silhouette Score อยุ่ที่ 0.7567 3กลุ่ม ค่า Silhouette Score อยุ่ที่ 0.6589 และ 4กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.5749 ตามลำดับ

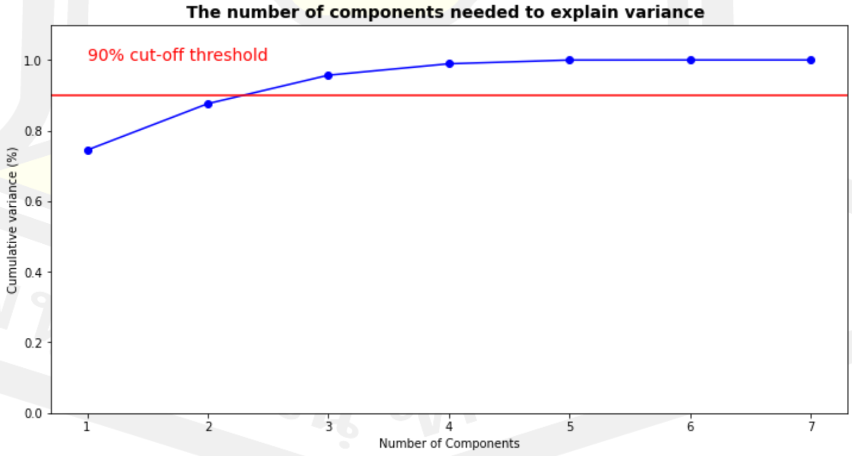
ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data เมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.16a ได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6589 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.16b



a) b)

ภาพประกอบที่ 4.16 ผลการหาจำนวนกลุ่มด้วยวิธีการ K-Means กับชุดข้อมูล Fitzroy River Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

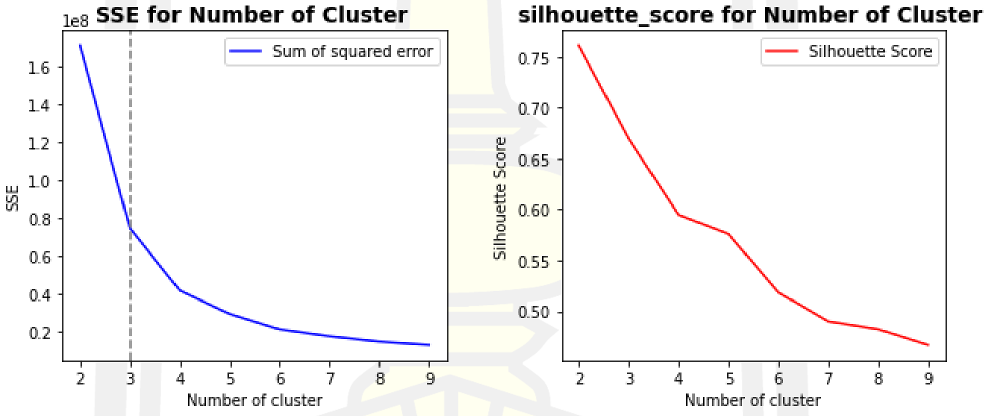
จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% ตามที่ Li และคณะได้กล่าวเอาไว้[73] พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component โดยค่าความแปรปรวน (Explained Variance) อยู่ที่ 90% ดังภาพประกอบที่ 4.17



ภาพประกอบที่ 4.17 การหาจำนวน Component ที่เหมาะสมของชุดข้อมูล Fitzroy River Data

หลังจากการหาจำนวน Component ได้จำนวน 2 Component แล้วทำการพิจารณาหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data โดยพิจารณาจากกราฟของการหักศอกซึ่งได้มาจากการคำนวณของค่า Sum of Squared Error ดังแสดงในภาพประกอบที่ 4.18a และกราฟที่ได้จากการหาค่า Silhouette Score ดังแสดงในภาพประกอบที่ 4.18b จากการวิเคราะห์กราฟจากภาพประกอบที่ 4.18a ซึ่งพิจารณาจากการหักศอก จำนวนกลุ่มที่เหมาะสมจะอยู่ที่ 3 กลุ่ม 4 กลุ่ม และ 6 กลุ่ม และเมื่อพิจารณากราฟของค่า Silhouette Score จากภาพประกอบที่ 4.13b จำนวนกลุ่มที่เหมาะสมเมื่อพิจารณาจากค่าที่เข้าใกล้ 1 จำนวนกลุ่มที่เหมาะสมจะอยู่ที่จำนวน 2กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.7612 3กลุ่ม ค่า Silhouette Score อยู่ที่ 0.6698 และ 4กลุ่ม โดยมีค่า Silhouette Score อยู่ที่ 0.5947 ตามลำดับ

ดังนั้นการหาจำนวนกลุ่มที่เหมาะสมของคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data เมื่อพิจารณาจากการหักศอกที่ได้จากค่า Sum of Squared Error ตามภาพประกอบที่ 4.18a ได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม ตามรอยเส้นประ และสอดคล้องกับ ค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6698 ซึ่งมีค่าเข้าใกล้ 1 เป็นลำดับที่ 2 ตามรอยเส้นประที่แสดงในภาพประกอบที่ 4.18b



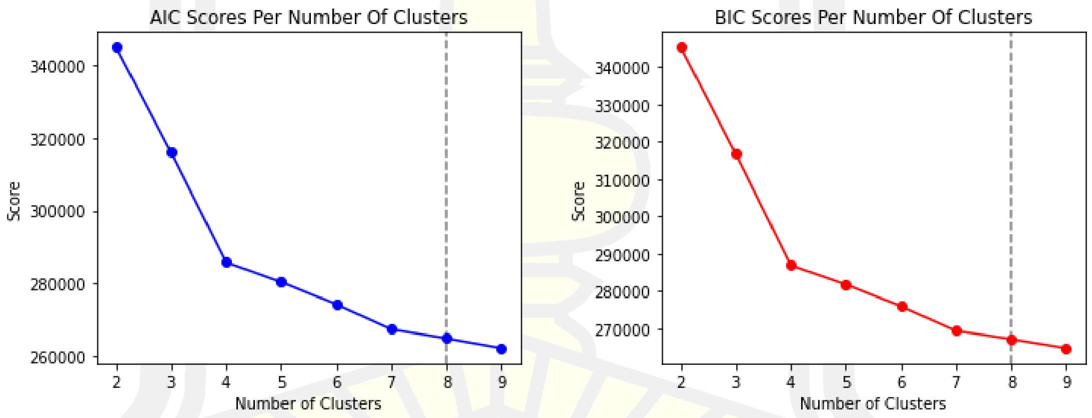
a) b)

ภาพประกอบที่ 4.18 ผลการหาจำนวนกลุ่มด้วยวิธีการ PCA+K-means กับชุดข้อมูล Fitzroy River Data a) ค่า Sum of Squared Error และ b) ค่า Silhouette Score

การหาจำนวนกลุ่ม ด้วย Mean shift เป็นวิธีการหาจำนวนกลุ่มแบบอัตโนมัติ โดยอาศัยการหาจำนวนกลุ่มแบบวนซ้ำ ซึ่งไม่อ้างอิงค่าพารามิเตอร์ เหมาะสำหรับการหาจำนวนกลุ่มที่มีรูปลักษณ์แบบไม่ตายตัว ซึ่งจำนวนกลุ่ม ของวิธีนี้จะให้ผลลัพธ์ออกมาแบบอัตโนมัติ โดยการค้นหาจุดศูนย์กลางของบริเวณที่หนาแน่นที่สุด ซึ่งเป็นวิธีที่ไม่ใช้พารามิเตอร์ในการประมาณการฟังก์ชันในการสุ่มตัวแปร [74] ในการทดลองนี้ผลการทดลองการหาจำนวนกลุ่มด้วยวิธีการ Mean shift กับชุดข้อมูลคุณภาพน้ำจาก Fitzroy River Data จะให้คำตอบออกมาเป็นจำนวนค่าตัวเลข ซึ่งสามารถบ่งบอกถึงจำนวนกลุ่มที่หาออกมาได้โดยอัตโนมัติ พบว่าได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม จากนั้นได้ทำการเพิ่ม

ประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component แล้วหาจำนวนกลุ่มด้วย Mean shift ซึ่งได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม

การหาจำนวนกลุ่มด้วย การจัดกลุ่มแบบจำลองการผสมแบบเกาส์ (Gaussian mixture model) หรือ GMM กับชุดข้อมูลคุณภาพน้ำจากชุดข้อมูล Fitzroy River Data โดยการพิจารณาจากค่าของ AIC และ BIC โดยการวัดหาจำนวนกลุ่มทั้งสองวิธีการนี้มักจะเลือกจำนวนกลุ่มเดียวกันสำหรับการหาจำนวนกลุ่ม แต่ถ้าในกรณีที่ค่าต่างกันให้พิจารณาจากค่า BIC ก่อน โดยพิจารณาจากกราฟที่มีค่าต่ำสุดหรือมีการลดระดับลงและมีลักษณะเส้นกราฟคงที่[75] จากการทดลองพบว่าการหาจำนวนกลุ่มด้วยวิธี GMM ได้จำนวนกลุ่ม เท่ากับ 8 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.19a ตามเส้นประและการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.19b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC

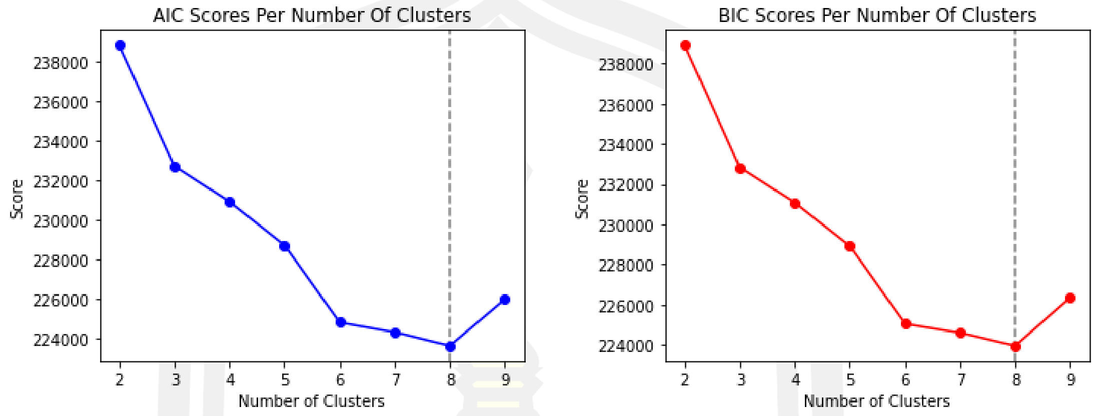


a) b)

ภาพประกอบที่ 4.19 ผลการหาจำนวนกลุ่มวิธีการ GMM กับชุดข้อมูล Fitzroy River Data a) ค่า AIC b) ค่า BIC

จากนั้นได้ทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยวิธีการทำ กระบวนการปรับปรุงโครงสร้างข้อมูล (Normalization) และการวิเคราะห์องค์ประกอบหลัก (PCA) ซึ่งเป็นการวิเคราะห์องค์ประกอบหลักที่บ่งบอกถึงคุณลักษณะที่สำคัญของตัวแปรที่มีความสัมพันธ์กัน เพื่อทำการลดจำนวนตัวแปรลง โดยในการทดลองนี้จะใช้ค่าความแปรปรวนในการเลือกจำนวน Component อยู่ที่ 90% พบว่าการทำ PCA ได้จำนวน Component ที่เหมาะสมอยู่เท่ากับ 2 Component ซึ่งได้ทำการหาจำนวน Component ไปแล้วนั้น แล้วจึงทำการหาจำนวนกลุ่มด้วยวิธีการ GMM พบว่า ได้

จำนวนกลุ่มเท่ากับ 8 กลุ่ม โดยใช้การพิจารณาจากค่า AIC ดังภาพประกอบที่ 4.20a ตามเส้นประ และการพิจารณาจากค่าของ BIC ดังภาพประกอบที่ 4.20b ตามรอยเส้นประที่ผู้วิจัยพล็อตแสดงเอาไว้เช่นกันกับค่า AIC



a) b)

ภาพประกอบที่ 4.20 ผลการหาจำนวนกลุ่มวิธีการ PCA+GMM กับชุดข้อมูล Fitzroy River Data Data a) ค่า AIC b) ค่า BIC

จากผลการทดลองการหาจำนวนกลุ่มคุณภาพน้ำจากชุดข้อมูลทั้ง 3 ชุดข้อมูลคือจากชุดข้อมูลของ เขื่อนลำปะทาว ชุดข้อมูลของCharles River Buoy Data และ ชุดข้อมูลของFitzroy River Data โดยใช้วิธีการ K-means MeanShift และ Gaussian Mixture Model (GMM) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 วิธีการ โดยผลของการหาจำนวนกลุ่มแต่ละวิธีการแสดงดังในตารางที่ 4.1

ตารางที่ 4.1 ผลการหาจำนวนกลุ่มจากชุดข้อมูลทั้ง 3 ชุดข้อมูล

ชุดข้อมูล	วิธีการ	จำนวนกลุ่ม
เขื่อนลำปะทาว	K-Means	5
	PCA+K-Means	5
	MeanShift	4
	PCA+MeanShift	3
	GMM	6
	PCA+GMM	6
Charles River Buoy Data	K-Means	4
	PCA+K-Means	4
	MeanShift	7
	PCA+MeanShift	8
	GMM	7
	PCA+GMM	7
Fitzroy River Data	K-Means	3
	PCA+K-Means	3
	MeanShift	3
	PCA+MeanShift	3
	GMM	8
	PCA+GMM	8

จากตารางแสดงผลการจัดกลุ่มคุณภาพน้ำของเขื่อนลำปะทาวจาก วิธีการ คือ K-means , MeanShift และ Gaussian Mixture Model (GMM) และนำวิธีการลดมิติของข้อมูล Principal Component Analysis (PCA) เข้ามาช่วยในการเพิ่มประสิทธิภาพของการหาจำนวนกลุ่มทั้ง 3 วิธีการ ในการทดลองนี้ จะมุ่งเน้นการหาจำนวนกลุ่มคุณภาพน้ำไปที่เขื่อนลำปะทาว จังหวัดชัยภูมิ ในการทดลองนี้จำนวนกลุ่มของคุณภาพน้ำในการจัดกลุ่มข้อมูลที่เหมาะสมอยู่ที่จำนวน 5 กลุ่ม หรือ 6 กลุ่ม ที่มีระยะเวลาในการเก็บข้อมูลอยู่ในช่วงเข้าสู่ฤดูร้อนระหว่างเดือน มกราคม-มีนาคม พ.ศ. 2564 ซึ่งในการทดลองนี้จะทำการเลือกจำนวนกลุ่มของเขื่อนลำปะทาวออกเป็น 5 กลุ่มจากวิธีการ PCA+K-Means ดังแสดงในภาพประกอบที่ 4.21 โดยแต่ละกลุ่มมีค่ามีช่วงของค่าแต่ละพารามิเตอร์ ดังนี้

กลุ่มที่ 1 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 4.5 – 5 mg/L ค่า TDS อยู่ในช่วง 122-124 ppm ค่า EC อยู่ในช่วง 21-22 % และค่า Temp อยู่ในช่วง 24 °C

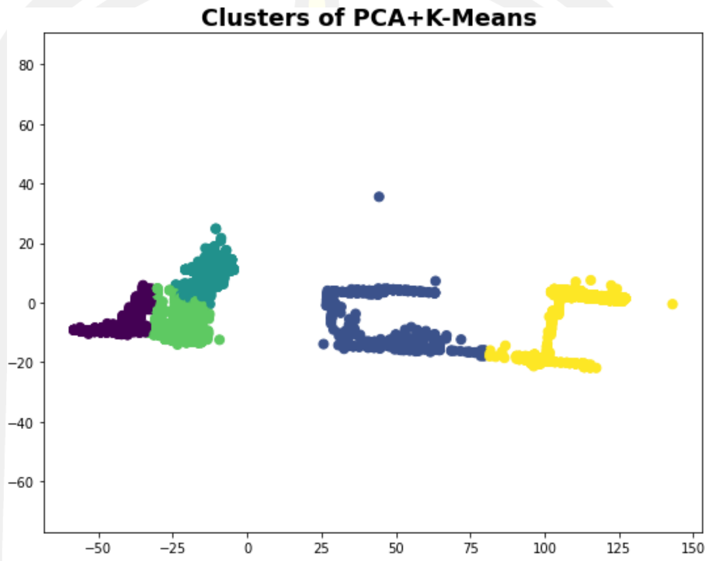
กลุ่มที่ 2 ค่า PH มีค่าอยู่ในช่วง 8-10 ค่า DO อยู่ในช่วง 3 – 4 mg/L ค่า TDS อยู่ในช่วง 229-234 ppm ค่า EC อยู่ในช่วง 6-8 % และค่า Temp อยู่ในช่วง 25-26 °C

กลุ่มที่ 3 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 4.5 – 5.6 mg/L ค่า TDS อยู่ในช่วง

125-127 ppm ค่า EC อยู่ในช่วง 13-18 % และค่า Temp อยู่ในช่วง 24 ℃

กลุ่มที่ 4 ค่า PH มีค่าอยู่ในช่วง 8-10 ค่า DO อยู่ในช่วง 3.6 – 4.4 mg/L ค่า TDS อยู่ในช่วง 199 ppm ค่า EC อยู่ในช่วง 6 % และค่า Temp อยู่ในช่วง 24 ℃

กลุ่มที่ 5 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 6 – 7 mg/L ค่า TDS อยู่ในช่วง 111-114 ppm ค่า EC อยู่ในช่วง 16-21 % และค่า Temp อยู่ในช่วง 23-24 ℃



ภาพประกอบที่ 4.21 ผลการจัดกลุ่มคุณภาพน้ำของเขื่อนลำปะทาว 5 กลุ่ม

ตารางที่ 4.2 ตารางเกณฑ์คุณภาพของน้ำ เพื่อวิเคราะห์กลุ่มคุณภาพน้ำ

พารามิเตอร์	**หน่วยวัด**	**ค่าคุณภาพน้ำที่เหมาะสม**	**ค่าคุณภาพน้ำที่มีความเสี่ยง**
Dissolved Oxygen concentration	milligrams per liter (mg/L)	5-8	<5 , >8
Temperature	Celsius (℃)	20-30	<20 , >30
pH	Standard Units	6.5-9	<6.5 , >9
Total Dissolved Solids	Part Per Million (ppm)	50-250	<50 , >250
Electric Conductivity	percent saturation (%)	30-80	<30 , >80

หมายเหตุ อ้างอิงเกณฑ์คุณภาพน้ำจาก [22]

จากผลการจัดกลุ่มคุณภาพน้ำของเขื่อนลำปะทาว ซึ่งแบ่งออกเป็น 5 กลุ่ม ผู้วิจัยได้ทำการเปรียบเทียบกับเกณฑ์คุณภาพน้ำที่แสดงในตารางที่ 4.2 พบว่ากลุ่มที่ 5 เป็นกลุ่มข้อมูลคุณภาพน้ำที่ดี

และเหมาะสมกว่าทุกกลุ่มเมื่อเทียบค่ากับตารางเกณฑ์คุณภาพของน้ำ ซึ่งมีค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 6 – 7 mg/L ค่า TDS อยู่ในช่วง 111-114 ppm ค่า EC อยู่ในช่วง 16-21 % และค่า Temp อยู่ในช่วง 23-24 °C

และในขณะเดียวกันการหาจำนวนกลุ่มคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data ที่ได้มาจากเว็บไซต์ของสำนักงานคุ้มครองสิ่งแวดล้อมแห่งสหรัฐอเมริกา (United States Environmental Protection Agency) โดยมีเซนเซอร์ที่ใช้วัดค่าจะอยู่ใต้ผิวน้ำลึกลงไป 1 เมตร เพื่อทำหน้าที่เก็บข้อมูลในทุก ๆ 15 นาที ข้อมูลที่จัดเก็บประกอบไปด้วย 8 พารามิเตอร์ ดังนี้ Water temperature, Specific conductance, pH, Dissolved Oxygen concentration, Dissolved Oxygen % saturation, Turbidity, Chlorophyll และ Phycoerythrin งานวิจัยนี้ได้เลือกใช้ Charles river buoy data จำนวนทั้งสิ้น 28,116 ชุดซึ่งจัดเก็บในช่วงปี ค.ศ. 2018-2019 มีจำนวนกลุ่มที่เหมาะสมอยู่ที่ 7 กลุ่ม จากการวิเคราะห์จากผลการหาจำนวนกลุ่มของแต่ละวิธีดังแสดงในตารางที่ 4.1

และชุดข้อมูล Fitzroy River Data ที่เป็นชุดนี้เป็น Open data ที่จัดเก็บและเผยแพร่โดย Queensland Government ประเทศออสเตรเลีย โดยเป็นการวัดคุณภาพน้ำจาก Fitzroy river, Central coast Queensland เป็นข้อมูลที่จัดเก็บในช่วงปี ค.ศ. 1993-2003 ประกอบด้วย 11,014 ชุดข้อมูล โดยได้เผยแพร่ผ่านเว็บไซต์ https://www.data.qld.gov.au ซึ่งประกอบไปด้วย 7 พารามิเตอร์ คือ Dissolved Oxygen concentration, Dissolved Oxygen % saturation, pH, Salinity, Specific Conductance และ Temperature มีจำนวนกลุ่มที่เหมาะสมอยู่ที่ 3 กลุ่ม จากการวิเคราะห์ผลการหาจำนวนกลุ่มของแต่ละวิธีดังแสดงในตารางที่ 4.1 เช่นกัน

ตารางที่ 4.3 ผลการเปรียบเทียบค่า Sum of Squared Error และ Silhouette Score สำหรับวิธีการ K-Mean และ PCA+ K-Mean

ชุดข้อมูล	จำนวนกลุ่ม	ค่า SSE K-Mean	ค่า SSE PCA+ K-Mean	ค่า SS K-Mean	ค่า SS PCA+ K-Mean
เขื่อนลำปะทาว	2	4573247.00006	**4522435.67257**	0.81519	**0.81761**
	3	2552826.33739	**2502111.69503**	0.64286	**0.65642**
	4	1260133.48652	**1211106.15656**	0.66321	**0.67752**
	5	823283.83657	**777183.34746**	0.68392	**0.70152**
	6	587970.97446	**546111.68159**	0.67717	**0.69479**
	7	525132.34224	**489091.04860**	0.65420	**0.69015**
	8	473825.04540	**427182.80670**	0.51937	**0.67050**
	9	411747.10624	**371179.00585**	0.50244	**0.53153**
Charles River Buoy Data	2	16859120.46322	**16441739.04066**	0.99569	**0.99590**
	3	9541669.35172	**9128579.47573**	0.54807	**0.58976**
	4	4807409.23826	**4394350.78679**	0.54898	**0.59092**
	5	3029735.59017	**2631431.48617**	0.49825	**0.55799**
	6	2305477.66406	**1911521.38961**	0.43696	**0.51899**
	7	1674662.00487	**1280712.81557**	0.43738	**0.51946**
	8	1215132.43166	**844638.32428**	0.42696	**0.52753**
	9	1000204.98615	**638124.08215**	0.40807	**0.51917**
Fitzroy River Data	2	173854317.39092	**171379336.85742**	0.75674	**0.76123**
	3	77251272.59971	**74784865.41915**	0.65893	**0.66984**
	4	44318367.29160	**41852825.87244**	0.57496	**0.59471**
	5	31853489.88920	**29385808.32834**	0.55316	**0.57594**
	6	23658293.14260	**21198825.24216**	0.48308	**0.51845**
	7	20010729.97071	**17553386.44326**	0.44589	**0.48987**
	8	17159679.23788	**14707452.79192**	0.43843	**0.48206**
	9	15390935.04830	**12938654.53202**	0.42384	**0.46663**

จากผลการทดลองตารางที่ 4.3 แสดงผลการเปรียบเทียบค่า Sum of Squared Error และ Silhouette Score สำหรับวิธีการ K-Mean และ PCA+ K-Mean เพื่อลดมิติของข้อมูล กับชุดข้อมูลทั้ง 3 ชุด โดยชุดข้อมูลเขื่อนลำปะทาวมีจำนวน 5 Component เมื่อทำ PCA แล้วเหลือจำนวน 3 Component ชุดข้อมูล Charles River Buoy Data มีจำนวน 8 Component เมื่อทำ PCA แล้วเหลือจำนวน 2 Component ชุดข้อมูล Fitzroy River Data มีจำนวน 7 Component เมื่อทำ PCA แล้วเหลือจำนวน 2 Component ผลการทดลองแสดงให้เห็นว่า วิธีการ Principal

Component Analysis (PCA) เพื่อลดมิติของข้อมูล แล้วหาจำนวนกลุ่มด้วย K-Mean ให้ผลของค่า Sum of Squared Error และ Silhouette Score ดีกว่า วิธีการ K-Mean กับชุดข้อมูลทั้ง 3 ชุดข้อมูล โดยค่า Sum of Squared Error ลดลง และค่า Silhouette Score มีค่าเข้าใกล้ 1 ดีกว่าวิธีการแบบ K-Mean

บทที่ 5
สรุปผลและอภิปรายผล

5.1 สรุปผล

งานวิจัยนี้ มุ่งเน้นในศึกษาการวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล ของเขื่อนลำปะทาว จังหวัดชัยภูมิ ในการทดลองนี้ผู้วิจัยได้ออกแบบทุ่นลอยน้ำที่ติดตั้งเซนเซอร์ตรวจวัดคุณภาพน้ำ จำนวน 5 พารามิเตอร์ประกอบด้วยเซนเซอร์วัดปริมาณออกซิเจนที่ละลายในน้ำ ค่าอุณภูมิของน้ำ ค่าความเป็นกรดเป็นเบสของน้ำ ค่าความขุ่นของน้ำ และค่าการนำไฟฟ้าของน้ำ เก็บรวบรวมข้อมูลคุณภาพน้ำในช่วงเดือน มกราคม-มีนาคม พ.ศ. 2564 จำนวนทั้งสิ้น 13,608 ชุดข้อมูล โดยกระจายรอบบริเวณเขื่อนจำนวน 5 จุด และได้ทดสอบกับชุดข้อมูลเพิ่มเติมอีก 2 ชุดข้อมูล คือ Charles River Buoy Data และ Fitzroy River Data โดยใช้ 3 วิธีการคือ K-Means Algorithm วิธีการ Mean Shift Algorithm และ Gaussian Mixture Model algorithm และทำการเพิ่มประสิทธิภาพการหาจำนวนกลุ่มด้วยการนำเอาวิธีการลดมิติของข้อมูลด้วย Principal Component Analysis (PCA) มาใช้กับวิธีการหาจำนวนกลุ่มทั้ง 3 วิธีการ

จากการทดลองพบว่าการหาจำนวนกลุ่มจากข้อมูลเขื่อนลำปะทาว จังหวัดชัยภูมิ ด้วย K-Means ได้ 5 กลุ่ม PCA+ K-Means ได้ 5 กลุ่ม วิธีการ Mean shift ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม PCA+ Mean shift ได้ 3 กลุ่ม วิธีการ GMM ได้จำนวนกลุ่ม 6 กลุ่ม PCA+GMM ได้ 6 กลุ่ม ซึ่งจากการวิเคราะห์การหาจำนวนกลุ่มจากวิธีที่กล่าวมานั้น ในงานนี้ได้ทำการเลือกจำนวนกลุ่มในการแบ่งกลุ่มคุณภาพน้ำออกเป็น 5 กลุ่ม โดยวิธีการ PCA+K-Means เป็นวิธีที่ให้จำนวนกลุ่มที่ดีที่สุดเมื่อทำการพิจารณาจาก Elbow Method และ Silhouette Score

ทั้งนี้ ยังได้ทดสอบกับชุดข้อมูลอีก 2 ชุดข้อมูลซึ่งเป็นชุดข้อมูล Open Data คือชุดข้อมูล Charles River Buoy Data และ Fitzroy River Data พบว่าการหาจำนวนกลุ่มคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data ด้วยวิธีการ K-Means ได้จำนวนกลุ่ม 4 กลุ่ม วิธีการ PCA+ K-Means ได้จำนวนกลุ่ม 4 กลุ่ม วิธีการ Mean shift ได้จำนวนกลุ่มเท่ากับ 7 กลุ่ม วิธีการ PCA+ Mean shift ได้จำนวนกลุ่ม 8 กลุ่ม วิธีการ GMM ได้จำนวนกลุ่ม 7 กลุ่มและวิธีการ PCA+GMM ได้จำนวนกลุ่ม 7 กลุ่ม จากผลการพิจารณาในการหาจำนวนกลุ่มคุณภาพน้ำจากชุดข้อมูล Charles River Buoy Data ได้จำนวนกลุ่มที่เหมาะสมอยู่ที่ 7 กลุ่มเมื่อพิจารณาจากผลการให้จำนวนกลุ่มที่ดีที่สุดเมื่อเปรียบเทียบกับวิธีการทั้งหมด และสำหรับการหาจำนวนกลุ่มของชุดข้อมูล Fitzroy River Data ด้วยวิธีการ K-Means ได้จำนวนกลุ่ม 3 กลุ่ม วิธีการ PCA+K-Means ได้ 3 กลุ่ม วิธีการ Mean shift ได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม วิธีการ PCA+ Mean shift ได้ 3 กลุ่ม วิธีการ GMM ได้จำนวนกลุ่ม 4 กลุ่ม และวิธีการ PCA+GMM ได้ 4 กลุ่มจากผลการพิจารณาในการหาจำนวนกลุ่มคุณภาพน้ำ

จากชุดข้อมูล Fitzroy River Data ได้จำนวนกลุ่มที่เหมาะสมอยู่ที่ 3 กลุ่มเมื่อพิจารณาจากผลการให้จำนวนกลุ่มที่ดีที่สุดเมื่อเปรียบเทียบกับวิธีการทั้งหมด

5.2 อภิปรายผล

จากผลการทดลองการวิเคราะห์กลุ่มคุณภาพน้ำของเขื่อนลำปะทาว จังหวัดชัยภูมิ จากผลการทดลองได้จำนวนกลุ่มเท่ากับ 5 กลุ่ม ซึ่งวิธีการ K-Means และวิธีการ PCA+ K-Means ให้จำนวนกลุ่มเท่ากับ 5 กลุ่มเท่ากัน แต่เมื่อพิจารณาค่า Sum of Squared Error ที่ให้ผลสำหรับวิธีการ Elbow Method จากค่าของความผิดพลาดลดลง จะทำให้เส้นที่มีความชันจะเริ่มโค้งและราบเรียบจนเกิดเป็นมุมลักษณะเหมือนการหักศอก ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่มที่ดีที่สุด และค่า Silhouette Score เป็นวิธีการหาจำนวนกลุ่มที่เหมาะสมเช่นกัน โดยค่าที่ได้จะอยู่ในช่วง -1 ถึง 1 ถ้ามีค่าเข้าใกล้ 1 แสดงว่าจำนวนกลุ่มมีความเหมาะสม ซึ่งจากวิธีการ K-Means จะให้ค่า Sum of Squared Error เท่ากับ 823283.83657 ค่า Silhouette Score เท่ากับ 0.68392 วิธีการ PCA+K-Mean โดย ได้ค่า Sum of Squared Error เท่ากับ 777183.34746 และค่า Silhouette Score เท่ากับ 0.70152 ดังนั้นเมื่อพิจารณาแล้ว วิธีการ PCA+K-Mean จะให้ค่า Sum of Squared Error และ Silhouette Score ดีกว่าวิธีการ K-Mean โดยที่ PCA จะลดมิติของข้อมูล โดยที่ยังคงไม่ทำให้สูญเสีย information ในการทดลองกับชุดข้อมูลเขื่อนลำปะทาวได้กำหนดค่าความแปรปรวน (Explained Variance) ที่ 90 % ดังนั้นในการทดลองการหาจำนวนกลุ่มของคุณภาพน้ำจากเขื่อนลำปะทาว ควรเลือกใช้วิธีการ PCA+K-Mean เพราะจากการลดมิติของข้อมูลโดยใช้ PCA ให้ผลของค่าค่า Sum of Squared Error และ Silhouette Score ดีกว่าวิธีการ K-Mean

และเมื่อพิจารณาถึงจำนวนกลุ่ม 5 กลุ่มแล้วสามารถแยกค่าพารามิเตอร์ของแต่ละกลุ่มโดยมีช่วงของค่าแต่ละพารามิเตอร์ ดังนี้

กลุ่มที่ 1 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 4.5 – 5 mg/L ค่า TDS อยู่ในช่วง 122-124 ppm ค่า EC อยู่ในช่วง 21-22 % และค่า Temp อยู่ในช่วง 24 °C

กลุ่มที่ 2 ค่า PH มีค่าอยู่ในช่วง 8-10 ค่า DO อยู่ในช่วง 3 – 4 mg/L ค่า TDS อยู่ในช่วง 229-234 ppm ค่า EC อยู่ในช่วง 6-8 % และค่า Temp อยู่ในช่วง 25-26 °C

กลุ่มที่ 3 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 4.5 – 5.6 mg/L ค่า TDS อยู่ในช่วง 125-127 ppm ค่า EC อยู่ในช่วง 13-18 % และค่า Temp อยู่ในช่วง 24 °C

กลุ่มที่ 4 ค่า PH มีค่าอยู่ในช่วง 8-10 ค่า DO อยู่ในช่วง 3.6 – 4.4 mg/L ค่า TDS อยู่ในช่วง 199 ppm ค่า EC อยู่ในช่วง 6 % และค่า Temp อยู่ในช่วง 24 °C

กลุ่มที่ 5 ค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 6 – 7 mg/L ค่า TDS อยู่ในช่วง 111-114 ppm ค่า EC อยู่ในช่วง 16-21 % และค่า Temp อยู่ในช่วง 23-24 °C

จากการหาจำนวนกลุ่มได้ 5 กลุ่มของชุดข้อมูลเขื่อนลำปะทาว ผู้วิจัยได้ทำการวิเคราะห์หาว่ากลุ่มไหนที่มีคุณภาพที่เหมาะสมกับการเลี้ยงปลา โดยกลุ่มที่ 5 เป็นกลุ่มที่มีความเหมาะสมในการเลี้ยงปลาในบริเวณเขื่อนลำปะทาว ซึ่งมีค่า PH มีค่าอยู่ในช่วง 7 ค่า DO อยู่ในช่วง 6 – 7 mg/L ค่า TDS อยู่ในช่วง 111-114 ppm ค่า EC อยู่ในช่วง 16-21 % และค่า Temp อยู่ในช่วง 23-24 °C

5.3 ข้อเสนอแนะ

1 ในการเก็บข้อมูลควรมีการระบุจุด วันที่ และเวลาของการเก็บข้อมูลลงในชุดข้อมูลที่เก็บ

2 หลังจากการหาจำนวนกลุ่มจากจุดเก็บข้อมูลจำนวน 5 จุดและทำการวิเคราะห์ออกมาว่ากลุ่มไหนเหมาะสำหรับการเพาะเลี้ยงปลา ควรวิเคราะห์ด้วยว่าจุดที่เก็บข้อมูลจุดไหนคุณภาพน้ำเหมาะสมหรือไม่เหมาะสมสำหรับการเลี้ยงปลา

บรรณานุกรม

บรรณานุกรม

[1] สำนักนายกรัฐมนตรี, "แผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 12," 2559.

[2] กรมการประมง กระทรวงเกษตรและสหกรณ์, "สถิติประมงแห่งประเทศไทย พ.ศ.2560," 2562.

[3] N. W. Patcharawalai Sryasak , Chanagan Chitmanat , Jongkon Promya , Louis Lebe, "Impacts of Climate and Season on Water Quality in Aquaculture Ponds," *KKU Res.j,* vol. 19(5), pp. 743-751, 2014.

[4] P. Gokhale, O. Bhat, and S. Bhat, "Introduction to IOT," vol. 5, pp. 41-44, 01/01 2018.

[5] P. P. Jayaraman, A. Yavari, D. Georgakopoulos, A. Morshed, and A. Zaslavsky, "Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt," *Sensors,* vol. 16, no. 11, 2016.

[6] J. Nord, A. Koohang, and J. Paliszkiewicz, "The Internet of Things: Review and Theoretical Framework," *Expert Systems with Applications,* vol. 133, 05/01 2019.

[7] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, and P. Nillaor, "IoT and agriculture data analysis for smart farm," *Computers and Electronics in Agriculture,* vol. 156, pp. 467-474, 2019/01/01/ 2019.

[8] M. Mehra, S. Saxena, S. Sankaranarayanan, R. J. Tom, and M. Veeramanikandan, "IoT based hydroponics system using Deep Neural Networks," *Computers and Electronics in Agriculture,* vol. 155, pp. 473-486, 2018/12/01/ 2018.

[9] X. Yao, J. Wang, M. Shen, H. Kong, and H. Ning, "An improved clustering algorithm and its application in IoT data analysis," *Computer Networks,* vol. 159, pp. 63-72, 2019/08/04/ 2019.

[10] J. Chen and A. Yang, "Intelligent Agriculture and Its Key Technologies Based on Internet of Things Architecture," *IEEE Access,* vol. 7, pp. 77134-77141, 2019.

[11] M. S. U. Chowdury *et al.*, "IoT Based Real-time River Water Quality Monitoring System," *Procedia Computer Science,* vol. 155, pp. 161-168, 2019/01/01/ 2019.

[12] A. Najah Ahmed *et al.*, "Machine learning methods for better water quality prediction," *Journal of Hydrology,* vol. 578, p. 124084, 2019/11/01/ 2019.

[13] M. Hajigholizadeh and A. M. Melesse, "Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses," *CATENA,* vol. 151, pp. 247-258, 2017/04/01/ 2017.

[14] L. Parra, J. Rocher, J. Escrivá, and J. Lloret, "Design and development of low cost smart turbidity sensor for water quality monitoring in fish farms," *Aquacultural Engineering,* vol. 81, pp. 10-18, 2018/05/01/ 2018.

[15] G. Carrasco, J.-L. Molina, C. Patino-Alonso, M. Castillo, M.-P. VicenteGalindo, and M.-P. Galindo-Villardón, "Water quality evaluation through a multivariate statistical HJ-Biplot approach," *Journal of Hydrology,* 07/26 2019.

[16] G. Gao, K. Xiao, and M. Chen, "An intelligent IoT-based control and traceability system to forecast and maintain water quality in freshwater fish farms," *Computers and Electronics in Agriculture,* vol. 166, p. 105013, 2019/11/01/ 2019.

[17] ฉัตรชัย โรจน์เรืองไร , กีฬา หนูยศ , สุรัช สะราคำ และ วันชัย ฉอ้อนศรี, "ศึกษาศักยภาพชุมชนลุ่มน้ำลำปะทาวตอนบน," สำนักงานคณะกรรมการส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม2553.

[18] แสงอรุณ เนื่องสิทธิ์ "ความต้องการการส่งเสริมการเลี้ยงปลานิลในกระชังของเกษตรกรในบริเวณอ่างเก็บน้ำลำปะทาว จังหวัดชัยภูมิ," มหาวิทยาลัยขอนแก่น, 2551.

[19] สุนทร ปัญญะพงษ์ "การคลี่คลายความขัดแย้งในการจัดการทรัพยากรน้ำในบริบทสังคมไทย: กรณีศึกษาเขื่อนลำปะทาว จังหวัดชัยภูมิ," มหาวิทยาลัยขอนแก่น, 2554.

[20] อมรรัตน์ นาคเสพ "นิเวศสังคมของชุมชนลำปะทาวภายหลังการก่อสร้างเขื่อนไฟฟ้าพลังน้ำลำปะทาว จังหวัดชัยภูมิ," มหาวิทยาลัยเชียงใหม่, 2554.

[21] งานประมง ศูนย์ศึกษาการพัฒนาเขาหินซ้อนอันเนื่องมาจากพระราชดำริ. (2002). คุณภาพน้ำเพื่อการเพาะเลี้ยงสัตว์น้ำ. Available: https://www.fisheries.go.th/shrimp/km/QualityWaterCoastal.pdf

[22] P. D. Anita Bhatnagar, "Water quality guidelines for the management of pond fish culture " *INTERNATIONAL JOURNAL OF ENVIRONMENTAL SCIENCES* vol. 3, no. 6, pp. 1980-2009, 2013.

[23] G. O. C. N.C. Ezeanya, K.N. Nwaigwe,C.C. Egwuonwu, "STANDARD WATER QUALITY REQUIREMENTS AND MANAGEMENT STRATEGIES FOR FISH FARMING," *International Journal of Research in Engineering and Technology,* vol. 04, no. 03, Mar-2015 2015.

[24] I. Hussain, M. Shakeel, M. Faisal, Z. A. Soomro, M. Hussain, and T. Hussain, "Distribution of Total Dissolved Solids in Drinking Water by Means of Bayesian Kriging and Gaussian Spatial Predictive Process," *Water Quality, Exposure and Health,* vol. 6, no. 4, pp. 177-185, 2014/12/01 2014.

[25] M. Tavana, V. Hajipour, and S. Oveisi, "IoT-based enterprise resource planning: Challenges, open issues, applications, architecture, and future research directions," *Internet of Things,* vol. 11, p. 100262, 2020/09/01/ 2020.

[26] Sathish and S. Smys, "A Survey on Internet of Things (IoT) based Smart Systems," 2020.

[27] นงเยาว์ สอนจะโปะ "รูปแบบการจำแนกกลุ่มข้อความภาษาไทยแบบอัตโนมัติโดยใช้การเรียนรู้ของเครื่อง (Machine Learning) ด้วยเทคนิค Unsupervised Learning ร่วมกับการประมวลผลภาษาธรรมชาติ (Natural Language Processing)," วารสารวิชาการ ศรีปทุม ชลบุรี*,* vol. 4, no. 14, pp. 95-106, 2561.

[28] T. Cheng-Fa, W. Han-Chang, and T. Chun-Wei, "A new data clustering approach for data mining in large databases," in *Proceedings International Symposium on Parallel Architectures, Algorithms and Networks. I-SPAN'*02, 2002, pp. 315-320.

[29] ปริญญา สงวนสัตว์ *Artificail Intelligence with Machine Learning*. 2562, p. 356.

[30] javatpoint. (2021, 24/7/2021). *Clustering in Machine Learning*. Available: https://www.javatpoint.com/clustering-in-machine-learning

[31] J. Petrus, Ermatita, and Sukemi, "Soft and Hard Clustering for Abstract Scientific Paper in Indonesian," in 2019 *International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2019, pp. 131-136.

[32] K. Charansiriphaisan, S. Chiewchanwattana, and K. Sunut, "K - Inverse Harmonic Means Clustering Algorithm," *Khon Kaen University Journal (Graduate Studies),* vol. 11, pp. 21-30, 04/01 2011.

[33] Y. A. Ghassabeh and F. Rudzicz, "Modified Subspace Constrained Mean Shift Algorithm," *Journal of Classification,* vol. 38, no. 1, pp. 27-43, 2021/04/01 2021.

[34] D. M. de Brito, V. Maracaja-Coutinho, S. T. de Farias, L. V. Batista, and T. G. do Rêgo, "A Novel Method to Predict Genomic Islands Based on Mean Shift Clustering Algorithm," *PLOS ONE,* vol. 11, no. 1, p. e0146352, 2016.

[35] calveeen. (2020). *Mean shift algorithm formula*. Available: https://dsp.stackexchange.com/questions/70743/mean-shift-algorithm-formula

[36] W. Jia, Y. Tan, L. Liu, J. Li, H. Zhang, and K. Zhao, "Hierarchical prediction based on two-level Gaussian mixture model clustering for bike-sharing system," *Knowledge-Based Systems,* vol. 178, pp. 84-97, 2019/08/15/ 2019.

[37] X. He, D. Cai, Y. Shao, H. Bao, and J. Han, "Laplacian Regularized Gaussian Mixture Model for Data Clustering," *IEEE Transactions on Knowledge and Data Engineering,* vol. 23, no. 9, pp. 1406-1418, 2011.

[38] วรางคณา เหนือคูเมือง และ เกศศักดิ์ดา ศีโคตร, "ระบบเตือนภัยดินถล่มแบบใช้เวลาจริงโดยใช้แบบจำลองการผสมเกาส์เซียนที่เรียนรู้รูปแบบจากวิธีค่าสูงสุดที่คาดหวัง," *การประชุมวิชาการระดับชาติมหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 7,* vol. 1, pp. 145-153, 2558.

[39] R. Nainggolan, R. Perangin-angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *Journal of Physics: Conference Series,* vol. 1361, p. 012015, 2019/11 2019.

[40] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conference Series: Materials Science and Engineering,* vol. 336, p. 012017, 2018/04 2018.

[41] X. Du, F. Shao, S. Wu, H. Zhang, and S. Xu, "Water quality assessment with hierarchical cluster analysis based on Mahalanobis distance," *Environmental Monitoring and Assessment,* vol. 189, 06/13 2017.

[42] N. U. Godwin Ogbuabor "Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value," *International Journal of Computer Science and Information Technology,* vol. 10, pp. 27-37, 2018.

[43] ทิพยา ถินสูงเนิน, มาโนช ถินสูงเนิน, กิตติศักดิ์ เกิดประสพ, และนิตยา เกิดประสพ, "การตรวจสอบความเหมาะสมของการจัดกลุ่มข้อมูลอนุกรมเวลา," *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี,* vol. 2, 2561.

[44] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Comparative Analysis of Electricity Consumption at Home through a Silhouette-score prospective," in 2019 *21st International Conference on Advanced Communication Technology (ICACT),* 2019, pp. 589-591.

[45] N. António, P. Rita, and P. Saraiva, "COVID-19: Worldwide Profiles during the First 250 Days," *Applied Sciences,* vol. 11, no. 8, 2021.

[46] J. Ponciano and M. Taper, "Model Projections in Model Space: A Geometric Interpretation of the AIC Allows Estimating the Distance Between Truth and Approximating Models," *Frontiers in Ecology and Evolution,* vol. 7, p. 413, 11/01 2019.

[47] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and

refinements," *WIREs Computational Statistics,* vol. 11, no. 3, p. e1460, 2019/05/01 2019.

[48] K.-H. Pho, S. Ly, S. Ly, and T. M. Lukusa, "Comparison among Akaike Information Criterion, Bayesian Information Criterion and Vuong's test in Model Selection: A Case Study of Violated Speed Regulation in Taiwan," *Journal of Advanced Engineering and Computation; Vol 3, No* 1 (2019)*DO* - 10.25073/*jaec.*201931.220*,* 03/31/ 2019.

[49] N. M. Tran, P. Burdejová, M. Ospienko, and W. K. Härdle, "Principal component analysis in an asymmetric norm," *Journal of Multivariate Analysis,* vol. 171, pp. 1-21, 2019/05/01/ 2019.

[50] Y. He, Y. Pang, Q. Zhang, Z. Jiao, and Q. Chen, "Comprehensive evaluation of regional clean energy development levels based on principal component analysis and rough set theory," *Renewable Energy,* vol. 122, pp. 643-653, 2018/07/01/ 2018.

[51] D. Shen, H. Wu, B. Xia, and D. Gan, "A Principal Component Analysis-Based Dimension Reduction Method for Parametric Power Flow," in 2020 12*th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2020, pp. 1-5.

[52] G. Carrasco, J.-L. Molina, C. Patino-Alonso, M. Castillo, M.-P. VicenteGalindo, and M.-P. Galindo-Villardón, "Water quality evaluation through a multivariate statistical HJ-Biplot approach.," *Journal of Hydrology,* 07/26 2019.

[53] N. A. Cloete, R. Malekian, and L. Nair, "Design of Smart Sensors for Real-Time Water Quality Monitoring," *IEEE Access,* vol. 4, pp. 3975-3990, 2016.

[54] Z. Di, M. Chang, P. Guo, Y. Li, and Y. Chang, "Using Real-Time Data and Unsupervised Machine Learning Techniques to Study Large-Scale Spatio–Temporal Characteristics of Wastewater Discharges and their Influence on Surface Water Quality in the Yangtze River Basin," *Water,* vol. 11, no. 6, 2019.

[55] S. Geetha and S. Gouthami, "Internet of things enabled real time water quality monitoring system," *Smart Water,* vol. 2, p. 1, 07/01 2017.

[56] S. Muangthong, "Water quality parameters reduction of the Lamtakong river basin (Thailand) using statistical analyses," *Engineering and Applied Science Research,* pp. 235-242, 2015.

[57] A. Pathak, M. Uddin, M. J. Abedin, K. Andersson, R. Mustafa, and M. Hossain, "IoT based Smart System to Support Agricultural Parameters: A Case Study," *Procedia Computer Science,* vol. 155, pp. 648-653, 09/13 2019.

[58] C. Z. Myint, L. Gopal, and Y. L. Aung, "Reconfigurable smart water quality monitoring system in IoT environment," in 2017 *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, 2017, pp. 435-440.

[59] A. A. Pranata, L. Jae Min, and K. Dong Seong, "Towards an IoT-based water quality monitoring system with brokerless pub/sub architecture," in 2017 *IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, 2017, pp. 1-6.

[60] N. Ding, H. Ma, H. Gao, Y. Ma, and G. Tan, "Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model," *Computers & Electrical Engineering,* vol. 79, p. 106458, 2019/10/01/ 2019.

[61] J. Zhang, J. Yan, D. Infield, Y. Liu, and F.-s. Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Applied Energy,* vol. 241, pp. 229-244, 2019/05/01/ 2019.

[62] M. Reda, "Application of Surface Water Quality Classification Models Using Principal Components Analysis and Cluster Analysis," *Journal of Geoscience and Environment Protection,* vol. 07, pp. 26-41, 01/01 2019.

[63] X. Yuan, B. Hu, and R. He, "Agglomerative Mean-Shift Clustering," *IEEE Transactions on Knowledge and Data Engineering,* vol. 24, no. 2, pp. 209-219, 2012.

[64] C. Wan, X. Cao, and W. Wu, "Vertical Stratification of Seawater in the Southern South China Sea in Winter Based on Cluster Analysis," in 2019 *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2019, pp. 452-455.

[65] A. E. Marín Celestino, D. A. Martínez Cruz, E. M. Otazo Sánchez, F. Gavi Reyes, and D. Vásquez Soto, "Groundwater Quality Assessment: An Improved Approach to K-Means Clustering, Principal Component Analysis and Spatial Analysis: A Case Study," *Water,* vol. 10, no. 4, 2018.

[66] U. S. E. P. Agency. (2020, 11/7/2020). *Live Water Quality Data for the Lower Charles River*.

[67] Q. G. O. D. Portal. (2014, 12/7/2020). *Fitzroy River (Drainage basin* 130). Available: https://www.data.qld.gov.au/dataset/ambient-estuarine-water-quality-monitoring-data-1993-to-2013/resource/831dc300-2e2c-4a90-a87e-3dcd74947025

[68] botland. (2020, 30/6/2021). *DFRobot Gravity analog X dissolved oxygen sensor* Available: https://botland.store/gravity-gas-dust-sensors/9968-dfrobot-gravity-analog-x-dissolved-oxygen-sensor-sen0237-a.html

[69] elecsensor. (2021, 30/6/2021). อุปกรณ์เซ็นเซอร์. Available: http://www.elecsensor.com/product/1108/ds18b20

[70] myarduino. (2021, 30/6/2021). *PH Sensor arduino Analog pH Meter* Available: https://www.myarduino.net/product/854/ph-sensor-arduino-analog-ph-meter

[71] myarduino. (2021, 30/6/2021). เซ็นเซอร์วัดความขุ่นของน้ำ *EC Sensor Turbidity sensor*. Available: https://www.myarduino.net/product/1162/

[72] arduinostep. (2021, 30/6/2021). *TDS Meter* เซนเซอร์วัดคุณภาพน้ำ. Available: http://www.arduinostep.com/product/446/tds-meter

[73] ท. ถินสูงเนิน, "การแทนอนุกรมเวลาเพื่อการจดักลุ่มที่มีประสิทธิภาพ " Phd, มหาวิทยาลัยเทคโนโลยีสุรนารี, 2560.

[74] H. H. R. Rasyidah, "Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm," *European Union Digital Library,* 2018.

[75] M. Q. D. L. H. L. B. T. M. T. G. Q. H. N. H. Hieu, "A Combination of K-Mean Clustering and Elbow Technique in Mitigating Losses of Distribution Network," *GMSARN International Journal,* vol. 13, pp. 153-185, 2019.

[76] T. Gupta and S. P. Panda, "Clustering Validation of CLARA and K-Means Using Silhouette & DUNN Measures on Iris Dataset," in 2019 *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 10-13.

[77] Z. Li, W. Zhang, D. Pan, and Q. Li, "Power Calculation of Multi-step Combined Principal Components with Applications to Genetic Association Studies," *Scientific Reports,* vol. 6, no. 1, p. 26243, 2016/05/18 2016.

[78] L. Abdallah and I. Shimshoni, *Mean Shift Clustering Algorithm for Data with Missing Values*. 2014.

[79] V. DiFrancesco. (2021, 30/6/2021). *Gaussian Mixture Models for Clustering*. Available: https://towardsdatascience.com/gaussian-mixture-models-for-clustering-3f62d0da675

ภาคผนวก

ภาคผนวก ก

ได้รับการตอบรับการนำเสนอผลงานวิชาการระดับนานาชาติ
จากงาน Research, Invention and Innovation Congress (RI²C 2021)
และได้ตีพิมพ์ใน IEEE Xplore

พหุ่ ปณ ฑิโต ชีเว

Water Quality Assessment in the Lam Pa Thao Dam, Chaiyaphum, Thailand with K-Means Clustering Algorithm

Phukkaraphon Ardarsa
Department of Information Technology
Faculty of Informatics, Mahasarakham University
Mahasarakham, Thailand
62011284504@msu.ac.th

Olarik Surinta
Multi-agent Intelligent Simulation Laboratory (MISL)
Department of Information Technology, Faculty of Informatics
Mahasarakham University, Mahasarakham, Thailand
olarik.s@msu.ac.th

***Abstract*— Water resource management is one of the biggest challenges that are being faced, such as a warming climate, arid land, and toxic chemicals in the water. It is essential to deal with water resource management urgently. In this article, researchers mainly focus on monitoring the water quality in the Lam Pa Thao dam, Chaiyaphum, Thailand. The farmer in that area directly affected by the water quality in the dam because they raise fish in floating fish cages. To prevent losses from fish farming, they should have the ability to monitor and control the factors that affect the water quality. As a result, the farmer can monitor the water quality and the monitor system can report to the farmer in time. In this case, to monitor the water quality, researchers designed the buoys, which is the internet of things device, to collect data from the Lam Pa Thao dam. researchers collected the water quality data from January - March 2021, including 13,608 instances. The five important parameters were obtained, including dissolved oxygen, temperature, pH, total dissolved solids, and electric conductivity. Due to the number of parameters, researchers decided not to apply dimension reduction. In these experiments, researchers proposed using K-means clustering algorithms to group the water data into appropriate clusters. For the K-Means algorithm, we calculated the silhouette coefficient to analyze the effectiveness of cluster separation. The best cluster that was grouped using the K-means algorithm achieved the silhouette score of 0.6839. Furthermore, researchers evaluated the K-means algorithm on Charles river and Fitzroy river datasets. It obtained the silhouette score of 0.5489 and 0.6589, respectively.**

Keywords*—*water quality segmentation, K-means clustering, silhouette coefficient, cluster analysis, internet of things

I. INTRODUCTION

Nowadays, the water supply problem is one of the public issues which affects human life. Water is used in agriculture, consumption, fishery, and public health. Sometimes, water from public health is the source of pathogens and becomes spoilage water.[1] Water is the fundamental factor for human life and the ecosystem, so water management is an urgent issue. [2] It is also the preparation for water problem in the future, For example, solving the water problem in time by using the present technology. [3] Data analysis is one of the technologies helping people to find the hidden data in a specific case and also help us to prove the hypothesis. Water quality analysis is one of the issues which gain attention from several researchers because water still has pollution problems and water supply problems globally. Maintaining water supply to ensure safety for consumption is essential. The fishery and agriculture are also needed high-quality water in their activities. For those reasons, this is an important issue for researchers. [4] To specify the problem and solutions to problems of water quality, water quality data is essential for improving water quality in the long term. [5]

For water data analysis, researchers need to know the water properties in changing environmental conditions and different conditions. Clustering is one of the data analysis techniques. Clustering divides the parameters into groups and parameters are grouped by basic characteristics defined by similarity. [6] The partition of the data is divided in order to define the same elements in the same group for finding the structure of the dataset.[7] This technique is Unsupervised Learning which process objects or values to the appropriate group. Distance of between objects is used as an indicator to identify similarities and differences. Water quality analysis using clustering techniques to find patterns of data distribution is the learning process from samples that is no target value and label value. The datasets are divided into the cluster. The data which contains the same characteristics will be set in the same group.

Water quality segmentation is an important process to make us know the quality of the water source. The quality data also used as a tool to improve water quality, solving the water problem, predicting water quality. This research uses three water quality datasets including, Charles River Buoy Data which was retrieved from the United States Environmental Protection Agency, Fitzroy River Data which is open data recorded by the Queensland government, Australia. The Fitzroy river data, Central coast Queensland was retrieved from the website. The third dataset is Lam Pa Thao dam, Chaiyaphum, Thailand which the researchers have created. The third dataset was created for water quality analysis.

Outline of the paper: This paper is organized in the following way. Section II presents a review of related work. Section III describes the water quality datasets, clustering algorithm, and cluster analysis. The experimental results and conclusion are presented in Section IV and V.

II. RELATED WORK

Carrasco et al. [8] claim that water quality is a sensitive issue. Water quality relates to the characteristics of physical, chemical, and biological elements. Complexities of the area are one thing researchers must concern about. Research in water quality tries to find a model to determine the variables

978-1-6654-0300-9/21/$31.00 ©2021 IEEE

ภาพประกอบที่ ก1 บทความที่ได้รับการตีพิมพ์

that most affect water quality. The multiple variable analysis help researcher to find relationship and solution for determining the biological, physical and chemical property of water. Water quality evaluation includes several factors such as temperature, pH, transparency, turbidity, nitrates, orthophosphates, phosphorus, total nitrogen, chlorophyll, solar radiation, dissolved oxygen, and microcystins. All of those factors are used as the water quality parameters in Gamboa and Paraiso river in Panama. The research includes three-step. the first step is basic statistical analysis including mean and standard deviation. The second step is HJ-Biplot analysis. The HJ-Biplot technique is the statistical analysis for detecting the correlation of chemical, physical and biological variables. The third step is group analysis. This step identifies the data from the random segmentation of the sampling and identifies affected variables. The segmentation process is calculated by the K-mean method using Biplot data from step two. The result from multiple variable analysis called HJ-Biplot which shows the relationship of chemical, physical, and biological elements. Moreover, the analysis shows two sample data relate to the seasons of the regions. The first group includes variables as follows, pH, transparency, chlorophyll, dissolved oxygen, and temperature. The second group includes variables as follows, nitrates, orthophosphates, turbidity, and others. All parameters normally vary in rainy seasons. The variances cause cyanobacteria and toxigenic in the water.

Hamed [9] claims that water quality monitoring is the priority of surface water protection. There are several methodologies for variable analysis which determine the variability of water quality from different sources. The important part of the variable analysis is the statistical technique and multiple variable statistics. The research decreases the number of variables when testing water in Nile river which are used for Cairo drinking water company. The relationship of each variable is analyzed using Fuzzy C-Means. K-means is used for Segmentation in order to find important factors which change water quality. The result shows that twenty-one water stations can be divided into three segments.

Celestino et al. [10] claims that segmentation by K-means and principal component analysis can be used for water quality analysis and water quality management. However, the previous study claims that K-means which use for finding Euclidean distance is not suitable for high dimensional data. The research uses the algorithm K-means+PCA for the water quality test. The PCA has been used to reduce the dimensions of the data for improving the grouping efficiency of K-means. The huge twenty-eight parameter datasets from hydrogeo chemistry are used. The data are retrieved from 582 water sources of the coastal area of Santo Domingo, Mexico. Segmentation by K-means+PCA using PCA for reducing the dimensions of the data can reduce the attributes of the hydrogeochemical variables from twenty-eight attributes to sixteen attributes. The result shows data are divided into three segments as follows. Segment one includes 160 water sources which are shallow and deep water sources near the coast and near the city. The second segment includes 166 water sources which are mid-level deep and deep water sources located near an industrial area and urban area. The third segment includes 256 water sources which are deep-level water sources near agricultural and urban areas. The result shows the methodology is effective in water source analysis.

Hajigholizadeha and Melesse [11] uses Cluster Analysis (CA) and Discriminant Analysis (DA) in water quality assessment and spatial change assessment in south Florida. The fifteen years dataset from 2000-2014 includes twelve water quality variables from sixteen water stations. The result from sixteen stations shows 35,000 datasets. Water quality from the data can be divided into three segments including low pollution, moderate pollution, and high pollution.

III. Materials and Methods

A. Water Quality Datasets

- *Lam Pa Thao Dam Water Quality Dataset*

The data in this research are collected by buoys equipped with sensors located at Lam Pa Thao Dam, Chaiyaphum, Thailand. Buoys are located at different five points near tilapia cages, the banks of the Lam Pa Thao dam. Total of 13,608 datasets were collected. The data were collected from January – March 2021. The buoy location is shown in Fig. 1.

The water quality of Lam Pa Thao dam, Chaiyaphum, Thailand collected from five floating buoys during January – March 2021 are shown in Fig.2. The data includes five parameters including, dissolved oxygen concentration (DO), temperature, pH, total dissolved solids (TDS), and electric conductivity (EC). The details of the parameters have shown in Table 1.

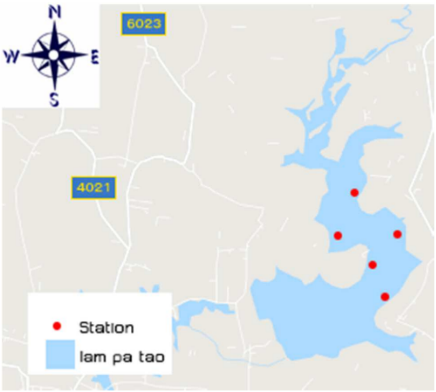


Fig. 1. The position where researchers collect data at Lam Pa Thao dam, Chaiyaphum, Thailand

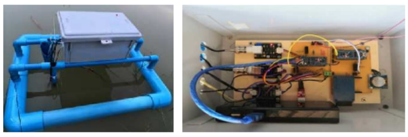


Fig. 2. The picture of buoys is used as data collector instruments. The buoys are used as data collectors near tilapia fish cages in the dam. The picture also identifies the IoT instruments in each buoy

ภาพประกอบที่ ก2 บทความที่ได้รับการตีพิมพ์

TABLE I. PARAMETERS WHICH USED TO MEASURE THE WATER QUALITY AND DATA MEASUREMENT UNIT

Parameters	Unit of Measurement
Dissolved Oxygen concentration	milligrams per liter (mg/L)
Temperature	Celsius (°C)
pH	Standard Units
Total Dissolved Solids	Part Per Million (ppm)
Electric Conductivity	percent saturation (%)

- *Charles River Buoy Data*

These data are retrieved from the United States Environmental Protection Agency. They set the buoys at the Lower Charles River for water quality measurement. The water quality data are distributed via the website "https://www.epa.gov/charlesriver/live-water-quality-data-lower-charles-river." Water monitoring buoys are floated at Charles river near the Museum of Science. The sensors are placed one meter below the water surface for collecting data every fifteen minutes. (Table II) shows the 8 collected parameter data including, water temperature, specific conductance, pH, dissolved oxygen concentration, dissolved Oxygen % saturation, turbidity, chlorophyll, and phycoerythrin [12]. This research uses 28,116 Charles river buoy data which are collected during 2018-2019.

TABLE II. PARAMETER FOR WATER MEASUREMENT FROM CHARLES RIVER BUOY

Parameters	Unit of Measurement
Temperature	Celsius (°C)
Specific conductance	millisiemens per centimeter (mS/cm)
pH	Standard Units
Dissolved Oxygen concentration	milligrams per liter (mg/L)
Dissolved Oxygen % saturation	percent saturation (%)
Turbidity	Formazin Nephelometric Units (FNUs)
Chlorophyll	Raw Fluorescent Units (RFUs)
Phycoerythrin	Raw Fluorescent Units (RFUs)

- *Fitzroy River Data*

This data is open data which are collected and distributed by Queensland Government, Australia. The data is water assessment of Fitzroy River, central coast Queensland which collected during 1993-2003. The data include 11,014 datasets. The data are retrieved from https://www.data.qld.gov.au. The data consist of seven parameters including, dissolved oxygen concentration, dissolved oxygen % saturation, pH, salinity, specific Conductance at 25 degrees Celsius, and temperature [13] (Table III)

TABLE III. PARAMETER FOR WATER MEASUREMENT FROM FITZRON RIVER DATA

Parameters	Unit of Measurement
Dissolved Oxygen concentration	milligrams per liter (mg/L)
Dissolved Oxygen (DO)	Percent saturation (%)
pH	Standard Units
Salinity	Practical Salinity Units (PSU)
Specific conductance	millisiemens per centimeter (mS/cm)
Temperature	Celsius (°C)

B. K-Means Clustering Algorithm

K-Means clustering algorithm is an unsupervised learning algorithm developed by MacQueen in 1967. The algorithm uses non-hierarchical cluster analysis which K-Means algorithm automatically divides data into a segment (k) by calculating the distance between Centroid with the target data [14]. It is shown in the following equation.

$$J = \sum_{j=i}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \quad (1)$$

where k is number of segment (class), n is the number of instances, and c is the (centroid) of the group j.

C. Cluster Analysis

- *The Elbow Method*

The elbow method is the error assessment of the sum of the distance between the centroid and instance for choosing the proper k value. It can be called the optimal cluster number. The appropriate point of the elbow method is where the elbow bend in the graph, as shown in Fig. 3. This point gives the best cluster numbers. [23] [24] Sum of Square Error (SSE) can be calculated as follows.

$$SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (2)$$

where n is the number of instances, y_i is the value of the i^{th} instance, and $\bar{y}$ is the mean of all instances.

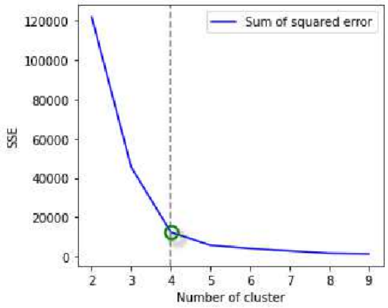


Fig. 3. Illustration of the elbow point

ภาพประกอบที่ ก3 บทความที่ได้รับการตีพิมพ์

- *Silhouette Coefficient*

Silhouette Coefficient is used for calculating the appropriate number of clusters. Silhouette value range between -1 to +1. The high value indicates the cluster is appropriately divided from the other cluster while a low value means the cluster is divided incorrectly. Silhouette coefficient equations are as follows.

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}},\ \text{if } |C_i| > 1 \qquad (3)$$

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j)$$

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} d(i,j)$$

where $d(i,j)$ is the distance between instances i and j in cluster (C), k is number of clusters, $a(i)$ is average intra-cluster distance, and $b(i)$ is average inter-cluster distance.

IV. Experimental Results

A. Experimental results on Lum Pa Thao Dam Water Quality Dataset

K-Means algorithm is used in this experiment for dividing the water quality of Lam Pa Thao dam. After that, water data in each cluster is analyzed by the elbow method and Silhouette Coefficient. The result from elbow method shows that the water quality is divided into four clusters. The elbow point is shown in Fig. 4. The calculation result from the Silhouette score shows the appropriate value is 0.6839 which divides the data into 4 clusters as well.

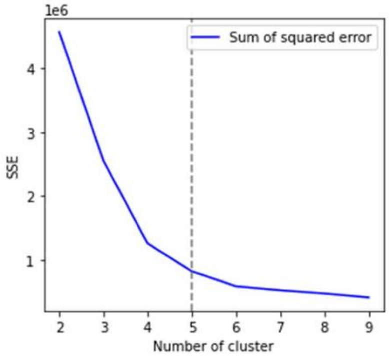


Fig. 4. analysis of suitable segmentation by elbow method

B. Experimental results on Charles River Buoy and Fitzroy River Datasets

The calculation for SSE value with Charles river buoy data and segmentation method by elbow method shows that four clusters are the best point for dividing water quality (See Fig. 5) The result corresponds to the optimal Silhouette value at four clusters as well. The Silhouette value is 0.5489.

Finally, the Fitzroy River Data is segmented by the K-Means algorithm method. The elbow method shows that the optimal number of groups for this data is three clusters, as shown in Fig. 6. The result shows the optimal Silhouette value is 0.6589 which divides the data into three clusters as well.

V. Conclusion

This research aims to study the water quality of Lam Pa Thao dam, Chaiyaphum Province, Thailand because there are a group of tilapia fish farmers around the dam area.

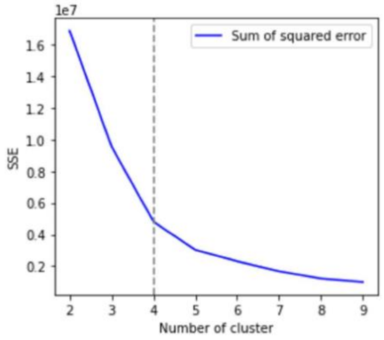


Fig. 5. Optimal number of segmentation analysis by Charles River Buoy Data using elbow method

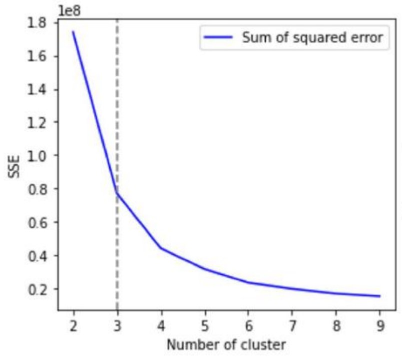


Fig. 6. Analysis of the optimal number of segmentation of Fitzroy River Data using elbow method

In the experiment, the researchers have designed a buoy that can measure the water quality in five parameters including, dissolved oxygen concentration (DO), temperature, pH, total dissolved solids (TDS), and electric conductivity (EC). The buoys are used for collecting data from January to March 2021. 13,608 datasets (instance) are collected from five different locations around the dam area. K-Means algorithm method is used for clustering water quality data. Elbow method and silhouette coefficient are used to analyze the effectiveness of water quality clustering. From the experiment, water quality in Lam Pa Thao dam, Chaiyaphum

ภาพประกอบที่ ก4 บทความที่ได้รับการตีพิมพ์

is divided into four groups. The optimal point of Silhouette value is 0.6839. The research method is also used with other two datasets including Charles river buoy data and Fitzroy river data. The result shows that Charles river buoy data are divided into four groups, the Silhouette value is 0.5489. the water quality of Fitzroy river is divided into three groups. The Silhouette value is 0.6589.

In Future work, a researcher can test water quality in each cluster to identify the appropriate cluster for tilapia farming.

REFERENCES

[1] K. Rishitha and S. Ullas, "IoT based Automation in Domestic Sewage Treatment Plant to Optimize Water Quality and Power Consumption," in *3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 306-310.

[2] L. N. Nthunya, N. P. Khumalo, A. R. Verliefde, B. B. Mamba, and S. D. Mhlanga, "Quantitative analysis of phenols and PAHs in the Nandoni Dam in Limpopo Province, South Africa: A preliminary study for dam water quality management," *Physics and Chemistry of the Earth, Parts A/B/C,* vol. 112, pp. 228-236, 2019/08/01/ 2019.

[3] W.-J. Syu, T.-K. Chang, and S.-Y. Pan, "Establishment of an Automatic Real-Time Monitoring System for Irrigation Water Quality Management," *International Journal of Environmental Research and Public Health,* vol. 17, no. 3, 2020.

[4] M. Tripathi and S. K. Singal, "Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India," *Ecological Indicators,* vol. 96, pp. 430-436, 2019/01/01/ 2019.

[5] P. Luo *et al.*, "Water quality trend assessment in Jakarta: A rapidly growing Asian megacity," *PLOS ONE,* vol. 14, no. 7, p. e0219009, 2019.

[6] F. Ustaoğlu and Y. Tepe, "Water quality and sediment contamination assessment of Pazarsuyu Stream, Turkey using multivariate statistical methods and pollution indicators," *International Soil and Water Conservation Research,* vol. 7, no. 1, pp. 47-56, 2019/03/01/ 2019.

[7] T. Vo-Van, A. Nguyen-Hai, M. V. Tat-Hong, and T. Nguyen-Trang, "A New Clustering Algorithm and Its Application in Assessing the Quality of Underground Water," *Scientific Programming,* vol. 2020, p. 6458576, 2020/03/07 2020.

[8] N. M. Tran, P. Burdejová, M. Ospienko, and W. K. Härdle, "Principal component analysis in an asymmetric norm," *Journal of Multivariate Analysis,* vol. 171, pp. 1-21, 2019/05/01/ 2019.

[9] D. Shen, H. Wu, B. Xia, and D. Gan, "A Principal Component Analysis-Based Dimension Reduction Method for Parametric Power Flow," in *2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2020, pp. 1-5.

[10] J. Gan, A. Li, Q. Lei, H. Ren, and Y. Yang, "K-means based on active learning for support vector machine," in *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, 2017, pp. 727-731.

[11] R. Nainggolan, R. Perangin-angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *Journal of Physics: Conference Series,* vol. 1361, p. 012015, 2019/11 2019.

[12] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conference Series: Materials Science and Engineering,* vol. 336, p. 012017, 2018/04 2018.

[13] N. U. Godwin Ogbuabor "Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value," *International Journal of Computer Science and Information Technology,* vol. 10, pp. 27-37, 2018.

[14] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Comparative Analysis of Electricity Consumption at Home through a Silhouette-score prospective," in *21st International Conference on Advanced Communication Technology (ICACT)*, 2019, pp. 589-591.

[15] U. S. E. P. Agency. (2020, 11/7/2020). Live Water Quality Data for the Lower Charles River. Available: https://www.epa.gov/charlesriver/live-water-quality-data-lower-charles-river

[16] Q. G. O. D. Portal. (2014, 12/7/2020). *Fitzroy River (Drainage basin 130)*. Available: https://www.data.qld.gov.au/dataset/ambient-estuarine-water-quality-monitoring-data-1993-to-2013/resource/831dc300-2e2c-4a90-a87e-3ded74947025

ภาพประกอบที่ ก5 บทความที่ได้รับการตีพิมพ์

ภาคผนวก ข

ตัวอย่างข้อมูลที่ใช้ในการทดลอง

ตารางที่ ข1 ชุดข้อมูลที่ใช้ในการทดลอง

PH	DO	TDS	EC	Temp
10	4.44	199	6	24.3
10.1	4.39	199	6	24.5
10.1	3.66	199	6	24.6
8.5	4.14	199	6	24.7
8.9	4.15	199	6	24.8
8.8	4.04	200	7	24.9
9.9	3.24	200	7	24.8
10.2	4.45	202	7	24.9
8.4	4.13	204	7	24.9
9.7	3.56	202	7	25
8.4	4.56	205	7	25
10.2	4.15	204	7	25
9.6	4.25	204	7	25
10.2	3.37	205	7	25
10.1	3.02	205	7	25
8.7	3.4	205	7	25.1
9.7	4.27	204	7	25.2
10.4	3.71	202	7	25.1
10.1	4.44	205	7	25.1
8.7	3.05	205	8	25.2
8.8	3.38	205	8	25.2
9.5	3.1	202	9	25.2
8.6	4.03	202	9	25.2
9.1	3.92	202	9	25.2
9.5	4.48	204	8	25.2
10	4.07	205	8	25.3
9.6	4.37	204	8	25.4
9.3	3.28	204	8	25.3
10.6	4.16	204	8	25.3

ประวัติผู้เขียน

ชื่อ	นายภัครพล อาจอาษา
วันเกิด	19 ธันวาคม 2526
สถานที่เกิด	จังหวัดชัยภูมิ
สถานที่อยู่ปัจจุบัน	19/1 หมู่ 5 ตำบลนาฝาย อำเภอเมือง จังหวัดชัยภูมิ
ตำแหน่งหน้าที่การงาน	นักวิชาการคอมพิวเตอร์
สถานที่ทำงานปัจจุบัน	สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏชัยภูมิ
ประวัติการศึกษา	พ.ศ.2548 วิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิทยาการคอมพิวเตอร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยราชภัฏชัยภูมิ พ.ศ.2564 วิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

พหูนํ ปณฺฑิโต ชีเว