



การศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง สำหรับการวิเคราะห์คุณภาพน้ำ กรณีศึกษา
การเลี้ยงปลา

นางสาวศมลรดา รัตนภักดี

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต

สาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีการศึกษา 2568

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

การศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง สำหรับการวิเคราะห์คุณภาพน้ำ กรณีศึกษา
การเลี้ยงปลา



การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตร

วิทยาศาสตร์มหาบัณฑิต

สาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ปีการศึกษา 2568

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



ใบรับรองโครงการค้นคว้าอิสระ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง การศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง สำหรับการวิเคราะห์คุณภาพน้ำ
กรณีศึกษาการเลี้ยงปลา

โดย นางสาวศมรดา รัตนภักดิ์

ได้รับอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชา
การบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ

คณบดีคณะเทคโนโลยีสารสนเทศและ
นวัตกรรมดิจิทัล

(ผู้ช่วยศาสตราจารย์ ดร.สุนันทา สดสี)

คณะกรรมการสอบการค้นคว้าอิสระ

ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.เทอดพงษ์ แดงสี)

อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร.พงศ์ศรัณย์ บุญโญปกรณ์)

กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.นวพร วิสิฐพงศ์พันธ์)

ชื่อ : นางสาวศมลรดา รัตนภักดี
ชื่อการค้นคว้าอิสระ : การศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง
สำหรับการวิเคราะห์คุณภาพน้ำ กรณีศึกษาการเลี้ยงปลา
สาขาวิชา : การบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก : ผู้ช่วยศาสตราจารย์ ดร.พงศ์ศรัณย์ บุญโญปกรณ์
ปีการศึกษา : 2568

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง สำหรับการวิเคราะห์คุณภาพน้ำ กรณีศึกษาการเลี้ยงปลา โดยมีจุดมุ่งหมาย 3 ประการ ได้แก่ 1. เพื่อเปรียบเทียบความแม่นยำของ 5 อัลกอริทึม ได้แก่ Linear Regression, Random Forest, Neural Networks, CNN และ LSTM ในการวิเคราะห์คุณภาพน้ำเพื่อการเลี้ยงปลา 2. เพื่อวิเคราะห์ความสำคัญของปัจจัยด้านคุณภาพน้ำ ที่ส่งผลต่อการตัดสินใจเลี้ยงปลาอย่างเหมาะสม 3. เพื่อเสนอแนวทางเลือกโมเดลที่เหมาะสมสำหรับนำไปใช้งานจริงในระบบ Smart farm การดำเนินงานวิจัยใช้ชุดข้อมูลจำลอง (Synthetic Dataset) ซึ่งสร้างขึ้นจากเกณฑ์คุณภาพน้ำที่เหมาะสมต่อการเลี้ยงปลาในกระชังนิตตามงานวิจัยที่เกี่ยวข้อง โดยกำหนดค่าคุณภาพน้ำที่เหมาะสม จากนั้นแบ่งข้อมูลออกเป็นชุดฝึกและชุดทดสอบในอัตราส่วน 80:20 และทำการฝึกสอนโมเดลด้วยภาษา Python ผู้ศึกษาวิจัยพบว่า Random Forest และ Gradient Boosting ให้ประสิทธิภาพสูงสุด โดยมีความแม่นยำในการจำแนกข้อมูลสูงที่สุด โดยมีค่าความแม่นยำอยู่ร้อยละ 99.98 ขณะที่ CNN และ LSTM ให้ผลลัพธ์ที่ใกล้เคียงกันแต่ใช้เวลาในการฝึกสอนสูงกว่า การวิเคราะห์ Feature Importance แสดงให้เห็นว่าค่า pH และปริมาณออกซิเจนละลายน้ำเป็นปัจจัยที่มีความสำคัญสูงสุดต่อการตัดสินใจของโมเดล

(มีจำนวนทั้งสิ้น 60 หน้า)

คำสำคัญ : Machine Learning, Random Forest, Neural Networks, Gradient Boosting, CNN, LSTM

อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก

Name : Miss Samonrada Rattanapakdee
Independent Study Title : A Study on the Performance of Machine Learning Algorithms for Water Quality Analysis: A Case Study of Aquaculture
Major Field : Network and Information Security Management
King Mongkut's University of Technology North
Bangkok
Independent Study Advisor : Assistant Professor Dr. Pongsarun Boonyopakorn
Academic Year : 2025

ABSTRACT

This research investigates the efficiency of various machine learning algorithms in analyzing water quality specifically for aquaculture applications. The study is structured around three core objectives: comparing the predictive accuracy of five distinct algorithms Linear Regression, Random Forest, Neural Networks, CNN, and LSTM; identifying the critical water quality parameters that influence optimal farming decisions; and proposing a robust model selection framework for integration into Smart Farm systems. The research methodology employed a synthetic dataset modeled after established water quality criteria for tilapia cage farming. This data was processed using Python, with the dataset partitioned into an 80:20 ratio for training and testing purposes, respectively. The empirical results demonstrate that Random Forest and Gradient Boosting outperformed other models, achieving a peak classification accuracy of 99.98%. While deep learning architectures such as CNN and LSTM provided competitive results, they were characterized by significantly higher computational demands and longer training durations. Furthermore, the Feature Importance analysis highlights that pH levels and Dissolved Oxygen (DO) serve as the most influential factors in the models' decision-making processes, marking them as vital indicators for real-time water quality monitoring in smart aquaculture.

(Total 60 Pages)

Keywords: Machine Learning, Neural Networks, Random Forest, Gradient Boosting, CNN, LSTM

Advisor

กิตติกรรมประกาศ

การจัดทำสารนิพนธ์ในครั้งนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี โดยได้รับความร่วมมือและความช่วยเหลือจากบุคคลหลายท่านที่ให้คำปรึกษาและความช่วยเหลือในด้านต่าง ๆ ซึ่งเป็นประโยชน์ต่อการดำเนินการค้นคว้า ข้าพเจ้าขอกราบขอบพระคุณที่ปรึกษาสารนิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.พงศ์ศรัณย์ บุญโญปกรณ์ ที่ได้ให้คำแนะนำและคำปรึกษาอันทรงคุณค่า ตลอดจนคณะกรรมการทุกท่านที่ได้ให้คำแนะนำ ตรวจสอบ แก้ไข และข้อเสนอแนะที่เป็นประโยชน์ในการพัฒนาสารนิพนธ์นี้ให้สำเร็จลุล่วง

ข้าพเจ้าขอขอบคุณอาจารย์ประจำคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัลทุกท่านที่ได้ให้ความรู้ คำแนะนำ และความช่วยเหลือในระหว่างการศึกษา รวมทั้งเพื่อนนักศึกษาสาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ที่ช่วยเหลือกันในทุก ๆ เรื่อง

สุดท้ายนี้ ข้าพเจ้าขอกราบขอบพระคุณครอบครัวและเพื่อน ๆ พี่ ๆ น้อง ๆ ทุกท่าน ที่ให้การสนับสนุนและเป็นกำลังใจให้ข้าพเจ้ามาตลอดจนการค้นคว้าอิสระนี้สำเร็จลุล่วงไปได้ด้วยดี

ศมลรดา รัตนภักดี

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญรูปภาพ.....	ญ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์.....	2
1.3 ขอบเขตการวิจัย.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 เครื่องมือที่ใช้ในงานวิจัย.....	4
1.6 คำจำกัดความที่ใช้ในการวิจัย	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	6
2.1 การเลี้ยงปลาในกระชัง.....	6
2.2 Synthetic Dataset.....	7
2.3 เทคโนโลยี Internet of Things (IoT) ในภาคเกษตรกรรม.....	10
2.4 Machine Learning และการพยากรณ์ข้อมูล.....	11
2.5 Random Forest (RF).....	12
2.5 Neural Network	13
2.6 Gradient Boosting (GB)	15
2.10 งานวิจัยที่เกี่ยวข้อง.....	18
บทที่ 3 วิธีการดำเนินงานวิจัย	20

3.1 กรอบแนวคิดการวิจัย	20
3.2 ขั้นตอนการดำเนินงานวิจัย	22
3.3 การพัฒนาโมเดล Machine Learning	25
3.4 การประเมินประสิทธิภาพโมเดล	30
3.5 การพัฒนาระบบทำนาย.....	32
3.6 การจัดการระบบและการตั้งค่า	33
3.7 การตรวจสอบและการประกันคุณภาพ	34
3.8 ข้อจำกัดและข้อพิจารณา.....	35
บทที่ 4 ผลการดำเนินการวิจัย.....	36
4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น	36
4.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดล.....	39
4.3 ผลการวิเคราะห์ความสำคัญของตัวแปร.....	41
4.4 ผลการประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดทางสถิติ.....	43
4.5 การเปรียบเทียบข้อดีข้อเสียของแต่ละโมเดล	47
4.6 การตรวจสอบความถูกต้องของโมเดล.....	49
4.7 การประยุกต์ใช้งานจริง.....	50
4.8 ข้อจำกัดของการศึกษา	50
บทที่ 5 สรุปและอภิปรายผล	52
5.1 สรุปผลการวิจัย.....	52
5.2 อภิปรายผลการวิจัย.....	53
5.3 ข้อเสนอแนะในการนำผลการวิจัยไปใช้.....	54
5.4 ข้อเสนอแนะสำหรับการวิจัยในอนาคต	54
บรรณานุกรม.....	56
ประวัติผู้เขียน.....	60

สารบัญตาราง

หน้า

ตารางที่ 3.1 ช่วงค่าพารามิเตอร์คุณภาพน้ำที่ใช้ในการสร้างชุดข้อมูลสำหรับการวิเคราะห์.....	24
ตารางที่ 3.2 ตารางสรุปโมเดล.....	27



สารบัญรูปภาพ

หน้า

ภาพที่ 2.1 การจำลองข้อมูลพลังงานและสเปกตรัมของโมเลกุล VO จากแบบจำลองเชิงทฤษฎี.....	8
ภาพที่ 2.2 แสดงกระบวนการทดลองในการใช้ข้อมูลจำลอง (Synthetic Data) เพื่อเพิ่มประสิทธิภาพของแบบจำลอง.....	9
ภาพที่ 2.3 การกระจายตัวของข้อมูลด้วยเทคนิค t-SNE สำหรับแบบ จำลองที่ใช้ข้อมูล Augmentation และข้อมูลจำลอง (Synthetic Data).....	10
ภาพที่ 2.4 กระบวนการทำงานของ Random Forest.....	12
ภาพที่ 2.5 กระบวนการทำงานของโครงข่ายประสาทเทียม Neural Network.....	14
ภาพที่ 2.6 กระบวนการทำงานของอัลกอริธึม Gradient Boosting (GB).....	15
ภาพที่ 2.7 กระบวนการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN).....	16
ภาพที่ 2.8 กระบวนการทำงานของโครงข่ายประสาทเทียมแบบ Long Short-Term Memory (LSTM)	17
ภาพที่ 3.1 กระบวนการวิจัย.....	21
ภาพที่ 4.1 การกระจายตัว (Distribution) และความสัมพันธ์ (Relationship) ของพารามิเตอร์คุณภาพน้ำ.....	38
ภาพที่ 4.2 ผลการประเมินประสิทธิภาพของโมเดล Machine Learning ที่ใช้ในการทดลอง.....	39
ภาพที่ 4.3 ผลการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance).....	41
ภาพที่ 4.4 ผลการวิเคราะห์ Confusion Matrix ของโมเดล Machine Learning.....	43
ภาพที่ 4.5 ค่าพื้นที่ใต้เส้นโค้ง (Area Under the Curve: AUC) ของโมเดลต่าง ๆ.....	45
ภาพที่ 4.6 เส้นโค้ง Precision-Recall ของโมเดลที่ใช้ในการทดลอง.....	46

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การเพาะเลี้ยงสัตว์น้ำ โดยเฉพาะการเลี้ยงปลาในกระชังน้ำจืด เช่น ปลานิลและปลาเศรษฐกิจ ชนิดต่าง ๆ เป็นกิจกรรมทางการเกษตรที่มีความสำคัญต่อเศรษฐกิจและความมั่นคงทางอาหารของประเทศไทย และผู้วิจัยพบว่า เกษตรกรผู้เพาะเลี้ยงสัตว์น้ำในประเทศไทยยังคงเผชิญกับปัญหาหลายประการ เช่น สภาพอากาศที่แปรปรวน ต้นทุนการผลิตที่สูง การจัดการคุณภาพน้ำ และความผันผวนของผลผลิตและราคาในตลาด ส่งผลให้ประสิทธิภาพการผลิตยังไม่สูงเท่าที่ควร [1]

จากการศึกษาของ พรศิลป์ บัวงาม และคณะ พบว่าเกษตรกรผู้เลี้ยงปลาในกระชังส่วนใหญ่ ยังคงอาศัยประสบการณ์ในการจัดการบ่อเลี้ยงมากกว่าการใช้ข้อมูลเชิงระบบ ทำให้การติดตามและวิเคราะห์ปัจจัยที่ส่งผลต่อการเลี้ยงปลาในกระชัง เช่น คุณภาพน้ำ โรค และสภาพแวดล้อม ยังขาดความแม่นยำ ในงานวิจัยนี้ผู้วิจัยเห็นว่าความต้องการที่สำคัญของเกษตรกร ได้แก่ การจัดเก็บข้อมูล การเลี้ยงปลาในกระชัง การป้องกันโรค และการนำเทคโนโลยีมาใช้ในการเลี้ยงปลาในกระชัง ซึ่งสะท้อนถึงความจำเป็นในการพัฒนาระบบเกษตรอัจฉริยะเพื่อช่วยในการติดตามสภาพแวดล้อมและสนับสนุนการตัดสินใจในการบริหารจัดการการเลี้ยงปลาในกระชังอย่างมีประสิทธิภาพ [2]

ปัจจุบัน เทคโนโลยี Internet of Things (IoT) ได้ถูกนำมาประยุกต์ใช้ในการเพาะเลี้ยงสัตว์น้ำ เพื่อช่วยตรวจวัดและติดตามสภาพแวดล้อมของบ่อเลี้ยงปลาในกระชังแบบเรียลไทม์ เช่น ค่าอุณหภูมิ ค่า pH ปริมาณออกซิเจนละลายน้ำ และแอมโมเนีย ซึ่งช่วยให้เกษตรกรสามารถควบคุมสภาพแวดล้อมในการเลี้ยงปลาในกระชังได้อย่างเหมาะสม ส่งผลให้ปลาเจริญเติบโตได้ดีและช่วยเพิ่มประสิทธิภาพการผลิตในการเพาะเลี้ยงสัตว์น้ำ จากการศึกษาผู้วิจัยยังพบว่าเทคโนโลยี IoT มีบทบาทสำคัญในการพัฒนาการเพาะเลี้ยงสัตว์น้ำสมัยใหม่ โดยช่วยให้สามารถติดตามและบริหารจัดการสภาพแวดล้อมของการเลี้ยงปลาในกระชังได้อย่างมีประสิทธิภาพมากยิ่งขึ้น [3]

ผู้วิจัยจึงได้เล็งเห็นว่าเทคโนโลยี IoT จะสามารถรวบรวมข้อมูลสภาพแวดล้อมของบ่อเลี้ยงปลาในกระชังได้อย่างต่อเนื่อง แต่ข้อมูลที่ได้จากอุปกรณ์ IoT มักมีปริมาณมากและมีความซับซ้อน การนำข้อมูลดังกล่าวมาใช้ให้เกิดประโยชน์สูงสุดจึงจำเป็นต้องอาศัยเทคนิคการวิเคราะห์ข้อมูลขั้นสูง เช่น การเรียนรู้ของเครื่อง (Machine Learning) เพื่อค้นหารูปแบบความสัมพันธ์ของข้อมูลและ

นำไปใช้ในการพยากรณ์ผลลัพธ์ของการเลี้ยงปลาในกระชังได้อย่างแม่นยำ ผู้วิจัยพบว่า อัลกอริทึม Machine Learning แต่ละประเภทมีลักษณะการทำงานและประสิทธิภาพที่แตกต่างกัน ดังนั้น ผู้วิจัยจึงต้องการเปรียบเทียบประสิทธิภาพของอัลกอริทึมที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลจากอุปกรณ์ IoT เป็นขั้นตอนสำคัญในการพัฒนาระบบพยากรณ์ผลการเลี้ยงปลาในกระชัง เพื่อเพิ่มความแม่นยำในการวิเคราะห์ข้อมูลและสนับสนุนการบริหารจัดการการเพาะเลี้ยงสัตว์น้ำให้มีประสิทธิภาพมากยิ่งขึ้น

1.2 วัตถุประสงค์

1.2.1 เพื่อเปรียบเทียบความแม่นยำของ 5 อัลกอริทึม ได้แก่ Linear Regression, Random forest, Neural Networks, Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) ในการพยากรณ์คุณภาพน้ำที่เหมาะสมต่อการเลี้ยงปลาในกระชัง

1.2.2 เพื่อวิเคราะห์ความสำคัญของปัจจัยด้านคุณภาพน้ำ (pH, DO, อุณหภูมิ) ที่ส่งผลต่อการตัดสินใจเลี้ยงปลาในกระชังอย่างเหมาะสม

1.2.3 เพื่อเสนอแนวทางเลือกโมเดลที่เหมาะสมสำหรับนำไปใช้งานจริงในระบบ Smart farm

1.3 ขอบเขตการวิจัย

1.3.1 ผู้วิจัยสร้างชุดข้อมูลจำลองขึ้นเพื่อจำลองสถานการณ์คุณภาพน้ำในบ่อเลี้ยงปลาในกระชัง โดยใช้ตัวแปรสำคัญ 3 ตัว ได้แก่ อุณหภูมิ (Temperature), ค่าความเป็นกรด-ด่าง (pH) และปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen) ตามหลักการของ Statistical Simulation เพื่อเพิ่มความแข็งแกร่งของข้อมูล โดยใช้การแจกแจงแบบสุ่ม (Random Sampling) จากการแจกแจงความน่าจะเป็น (Probability Distribution) ได้แก่

1.3.1.1 การแจกแจงแบบ Uniform Distribution สำหรับสร้างข้อมูลที่ครอบคลุมช่วงค่าทั้งที่อยู่ในเกณฑ์เหมาะสมและไม่เหมาะสม

1.3.1.2 การแจกแจงแบบ Normal Distribution (Gaussian) สำหรับเพิ่ม noise เพื่อจำลองความแปรปรวนในธรรมชาติของข้อมูลเซนเซอร์

โดยช่วงค่าของแต่ละตัวแปรกำหนดจาก เกณฑ์มาตรฐานจากงานวิจัยจริง ที่ระบุช่วงคุณภาพ น้ำที่เหมาะสมต่อการเลี้ยงปลาในกระชัง ซึ่งถูกนำมาใช้เป็นเงื่อนไขกำหนดพีเจอร์เป้าหมาย (Target Variable) ของโมเดล

1.3.2 กำหนด กลุ่มเป้าหมาย (target class) จากเกณฑ์ของงานวิจัยที่เกี่ยวข้อง โดยถือว่า คุณภาพน้ำ "เหมาะสม (suitable)" สำหรับการเลี้ยงปลาในกระชังเมื่อตัวแปรอยู่ในช่วง :

Temperature: 25–30 °C

pH: 6.5–9.0

Dissolved Oxygen (DO): 5-8mg/L

1.3.3 เปรียบเทียบความแม่นยำของโมเดล Machine Learning 5 อัลกอริทึม ได้แก่ Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) ในการพยากรณ์ความเหมาะสมของน้ำเพื่อเลี้ยงปลาในกระชัง โดยการใช้เครื่องมือต่อไปนี้

Python 3.12 + VS Code

ไลบรารี: pandas, scikit-learn, matplotlib, seaborn, tensorflow

กราฟ: Confusion Matrix, Accuracy, Precision, Recall, F1-score, ROC-AUC

1.3.4 การประเมินผลจะใช้เกณฑ์การวัดความแม่นยำของแบบจำลอง โดยใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall, F1-score, Confusion Matrix, วิเคราะห์ Feature Importance และ Training Time

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 สร้างชุดข้อมูลจำลอง (Synthetic Dataset) โดยจำลองค่าคุณภาพน้ำ 3 ตัวแปร ได้แก่ อุณหภูมิ (Temperature), ค่าความเป็นกรด-ด่าง (pH) และ ปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen)

1.4.2 การเตรียมข้อมูล

4.2.1.1 สร้างพีเจอร์เป้าหมาย (Target Variable: suitable/unsuitable) ตาม เกณฑ์จากงานวิจัยที่เกี่ยวข้อง

1.4.2.2 ทำการแบ่งชุดข้อมูลเป็น Train/Test (เช่น 80/20)

1.4.3 Model Selection & Training เปรียบเทียบโมเดล 5 อัลกอริทึม ได้แก่ Random Forest, Gradient Boosting, Neural Network CNN, และ LSTM

1.4.4. ทำการประเมินผล

1.4.4.1 ใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall, F1-score และ Confusion Matrix

1.4.4.2 วิเคราะห์ Feature Importance และ Training Time

1.4.5. สรุปผลและนำเสนอข้อเสนอแนะ

1.4.5.1 วิเคราะห์ว่าโมเดลใดมีประสิทธิภาพสูงสุดในการจำแนกข้อมูลคุณภาพน้ำ

1.4.5.2 เสนอแนวทางการนำไปใช้งานร่วมกับ IoT Sensor ในอนาคต

1.5 เครื่องมือที่ใช้ในงานวิจัย

1.5.1 Python 3.12 + VS Code

1.5.2 ไลบรารี: pandas, scikit-learn, matplotlib, seaborn, tensorflow

1.5.3 กราฟ: Confusion Matrix, Accuracy, Precision, Recall, F1-score, ROC-AUC

1.6 คำจำกัดความที่ใช้ในการวิจัย

1.6.1 อินเทอร์เน็ตของสรรพสิ่ง (Internet of Things: IoT) หมายถึง เครือข่ายของอุปกรณ์เซนเซอร์ที่ติดตั้งในกระชังเลี้ยงปลา เพื่อทำหน้าที่ตรวจวัดและส่งข้อมูลสภาพแวดล้อมทางกายภาพแบบเรียลไทม์

1.6.2 Machine Learning หมายถึง วิธีการทางสถิติและอัลกอริธึมคอมพิวเตอร์ที่นำมาใช้ในการวิเคราะห์รูปแบบความสัมพันธ์ของข้อมูลคุณภาพน้ำ เพื่อสร้างแบบจำลองในการพยากรณ์ความเหมาะสมของการเลี้ยงปลาในกระชัง

1.6.3 อัลกอริธึม (Algorithm) หมายถึง ชุดคำสั่งหรือกระบวนการคำนวณที่ใช้ในแบบจำลอง Machine Learning เพื่อจำแนกและพยากรณ์ข้อมูล ซึ่งในงานวิจัยนี้ประกอบด้วย Linear Regression, Random Forest, Neural Networks, CNN และ LSTM

1.6.4 คุณภาพน้ำ (Water Quality) หมายถึง ค่าตัวแปรทางกายภาพและเคมีในบ่อเลี้ยงปลานิลที่ใช้เป็นข้อมูลนำเข้า (Input) ได้แก่ อุณหภูมิ (Temperature), ค่าความเป็นกรด-ด่าง (pH) และปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen: DO)

1.6.5 ชุดข้อมูลจำลอง (Synthetic Dataset) หมายถึง ข้อมูลที่ถูกสร้างขึ้นด้วยวิธีการทางสถิติ (Statistical Simulation) โดยอ้างอิงจากเกณฑ์มาตรฐานในงานวิจัยจริง เพื่อใช้ในการฝึกสอนและทดสอบประสิทธิภาพของโมเดล

1.6.6 ความเหมาะสมในการเลี้ยงปลา (Suitability) หมายถึง ผลลัพธ์ที่เป็นเป้าหมาย (Target Variable) ของการพยากรณ์ โดยแบ่งเป็น 2 สถานะ คือ "เหมาะสม" เมื่อค่าคุณภาพน้ำอยู่ในเกณฑ์ที่กำหนด และ "ไม่เหมาะสม" เมื่อค่าคุณภาพน้ำอยู่นอกเกณฑ์

1.6.7 ประสิทธิภาพของแบบจำลอง (Model Performance) หมายถึง ค่าความสามารถในการพยากรณ์ของอัลกอริทึมที่วัดผ่านตัวชี้วัดต่าง ๆ ได้แก่ ความแม่นยำ (Accuracy), Precision, Recall และ F1-score



บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาค้นคว้าทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบพยากรณ์ปัจจัยที่มีผลต่อการเลี้ยงปลาในกระชังโดยใช้เทคโนโลยี Internet of Things (IoT) และ Machine Learning ซึ่งเป็นพื้นฐานสำคัญสำหรับการพัฒนาแบบจำลองในการศึกษา

2.1 การเลี้ยงปลาในกระชัง

2.2 Synthetic Dataset

2.3 เทคโนโลยี Internet of Things (IoT) ในภาคเกษตรกรรม

2.4 Machine Learning และการพยากรณ์ข้อมูล

2.5 Random Forest (RF)

2.6 Neural Network

2.7 Gradient Boosting (GB)

2.8 Convolutional Neural Network (CNN)

2.9 Long Short-Term Memory (LSTM)

2.10 งานวิจัยที่เกี่ยวข้อง

2.1 การเลี้ยงปลาในกระชัง

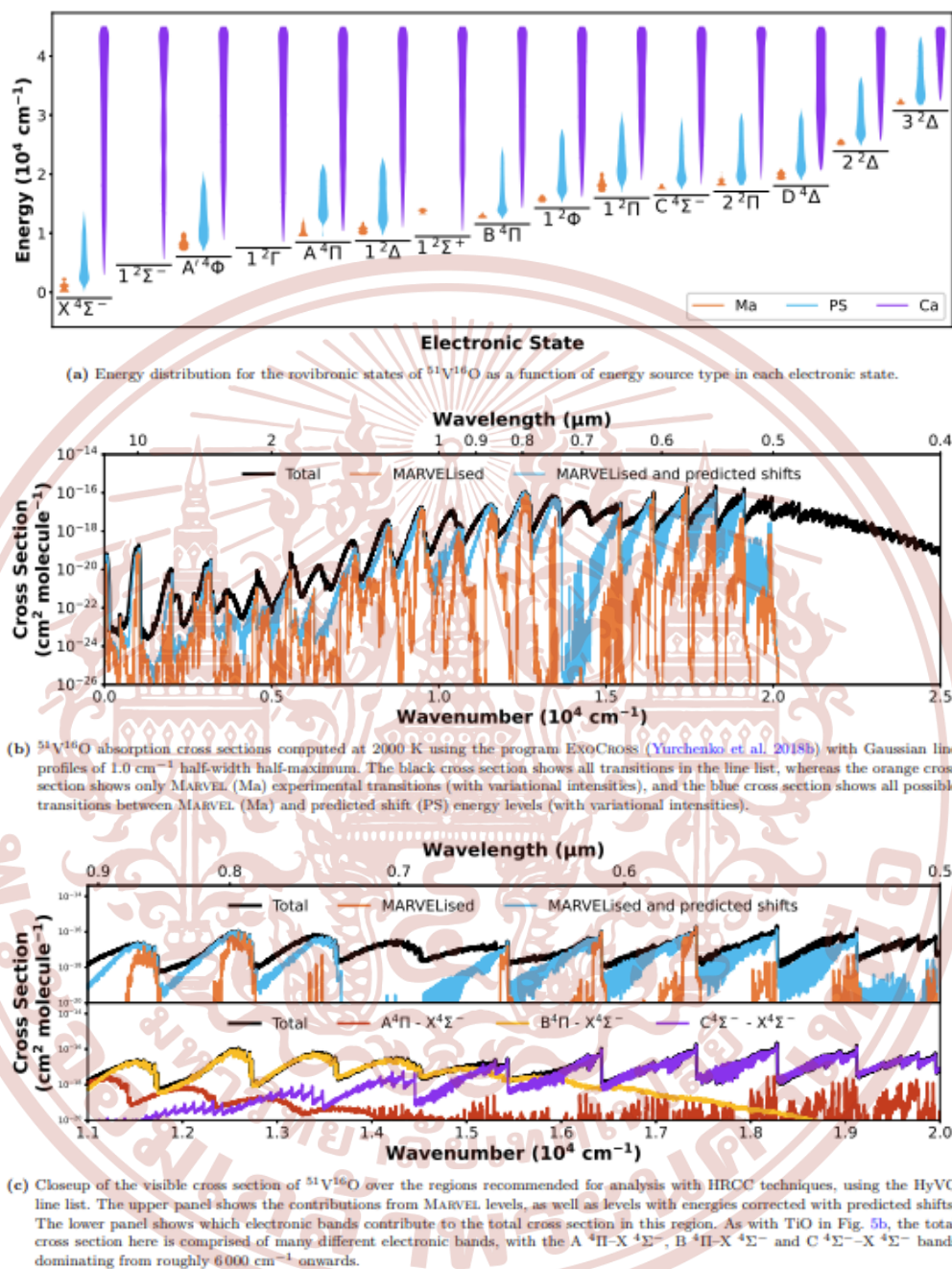
การเลี้ยงปลาในกระชัง ให้ผลผลิตสูงในเชิงเศรษฐศาสตร์ ช่วยลดระยะเวลาในการเลี้ยงให้สั้นลงได้ และมีการลงทุนต่ำกว่ารูปแบบการเลี้ยงอื่น ๆ อย่างไรก็ตามในขณะที่มีผลตอบแทนต่อพื้นที่สูง แต่การเลี้ยงปลาในกระชังในกระชังอาจมีข้อเสียบางอย่างซึ่งไม่สามารถควบคุมได้ เช่น สภาพแวดล้อมที่เปลี่ยนแปลงได้ตลอดเวลา [4]

กรมประมงได้รายงานว่าคุณภาพน้ำเป็นปัจจัยสำคัญที่มีผลต่อการเจริญเติบโต สุขภาพ และอัตราการรอดของปลาในการเพาะเลี้ยงสัตว์น้ำ โดยพารามิเตอร์ที่สำคัญในการประเมินคุณภาพน้ำ ได้แก่ อุณหภูมิของน้ำ (Temperature) ค่าความเป็นกรด-ด่างของน้ำ (pH) และปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen: DO) ซึ่งเป็นปัจจัยที่มีผลต่อกระบวนการหายใจ การเผาผลาญพลังงาน และการดำรงชีวิตของปลา หากค่าพารามิเตอร์ดังกล่าวอยู่ในช่วงที่เหมาะสมจะช่วยส่งเสริมการเจริญเติบโตและเพิ่มประสิทธิภาพการผลิตในการเพาะเลี้ยงปลาในกระชัง โดยอุณหภูมิของน้ำที่เหมาะสมต่อการเลี้ยงปลาในกระชังควรอยู่ในช่วงประมาณ 25–30 องศาเซลเซียส ซึ่งเป็นช่วงที่เอื้อ

ต่อการเจริญเติบโตและการกินอาหารของปลา ขณะที่ค่าความเป็นกรด-ด่างของน้ำควรอยู่ในช่วงประมาณ pH 7.0–8.0 ซึ่งเป็นช่วงที่เหมาะสมต่อการดำรงชีวิตของสัตว์น้ำ และปริมาณออกซิเจนละลายน้ำควรมีค่ามากกว่า 4 มิลลิกรัมต่อลิตร (mg/L) เพื่อให้ปลาได้รับออกซิเจนเพียงพอต่อการหายใจและลดความเสี่ยงต่อการเกิดความเครียดหรือการตายของปลา ดังนั้น การติดตามและควบคุมค่าพารามิเตอร์คุณภาพน้ำให้อยู่ในช่วงที่เหมาะสมจึงเป็นปัจจัยสำคัญที่ช่วยเพิ่มประสิทธิภาพและความสำเร็จในการเพาะเลี้ยงสัตว์น้ำ [5]

2.2 Synthetic Dataset

L. K. McKemmish และคณะ ได้ศึกษาแนวทางการสร้างข้อมูลเชิงทฤษฎีโดยใช้การคำนวณแบบ *ab initio* เพื่อสนับสนุนการสร้างแบบจำลองเชิงสเปกโทรสโกปีของโมเลกุล vanadium monoxide (VO) โดยใช้วิธีการคำนวณทางเคมีควอนตัมเพื่อสร้างข้อมูลที่มีความแม่นยำสูงสำหรับการสร้างแบบจำลองพลังงานและคุณสมบัติของโมเลกุล ในกระบวนการสร้างข้อมูลจำลองดังกล่าว นักวิจัยได้ใช้การคำนวณเชิงทฤษฎีเพื่อสร้าง potential energy curves, ค่า dipole moment, และค่า spin-orbit coupling ของโมเลกุล โดยใช้เทคนิค Complete Active Space Self-Consistent Field (CASSCF) และ Multi-Reference Configuration Interaction (MRCI) ซึ่งเป็นวิธีการคำนวณขั้นสูงในเคมีควอนตัม เพื่อจำลองพฤติกรรมของอิเล็กตรอนในโมเลกุลและสร้างข้อมูลพลังงานของสถานะอิเล็กตรอนิกส์ต่าง ๆ ของโมเลกุล VO จากนั้นข้อมูลที่ได้จากการคำนวณจะถูกนำมาใช้สร้างแบบจำลองสเปกโทรสโกปีและสร้างชุดข้อมูลเชิงสังเคราะห์ที่มีจำนวนมาก เช่น ระดับพลังงานและการเปลี่ยนผ่านของโมเลกุลจำนวนหลายล้านรายการ [7]



ภาพที่ 2.1 การจำลองข้อมูลพลังงานและสเปกตรัมของโมเลกุล VO จากแบบจำลองเชิงทฤษฎี

จากรูปจะเห็นได้ว่า การใช้แบบจำลองเชิงทฤษฎีสามารถสร้างข้อมูลที่ครอบคลุมช่วงค่าที่หลากหลาย และสามารถอธิบายพฤติกรรมของระบบได้อย่างต่อเนื่อง แม้ในกรณีที่ไม่มีข้อมูลจากการทดลองจริงในช่วง ซึ่งแสดงให้เห็นว่าการสร้างข้อมูลจำลองสามารถช่วยสนับสนุนการวิเคราะห์และการพัฒนาแบบจำลองเชิงคำนวณได้อย่างมีประสิทธิภาพ [7]

M. Frid-Adar และคณะ ได้ศึกษาการสร้างข้อมูลจำลอง (Synthetic Data) สำหรับการเพิ่มประสิทธิภาพของโมเดล Deep Learning ในการจำแนกภาพทางการแพทย์ โดยใช้เทคนิค Generative Adversarial Networks (GANs) เพื่อสร้างภาพทางการแพทย์จำลองจากชุดข้อมูลที่มีอยู่ ซึ่งสามารถช่วยแก้ปัญหาการมีข้อมูลฝึกสอนที่มีจำนวนจำกัดได้ งานวิจัยพบว่าการใช้ Synthetic Data เพื่อเพิ่มจำนวนข้อมูลฝึกสอนสามารถช่วยเพิ่มประสิทธิภาพของโมเดล Convolutional Neural Network (CNN) ในการจำแนกภาพรอยโรคตับ

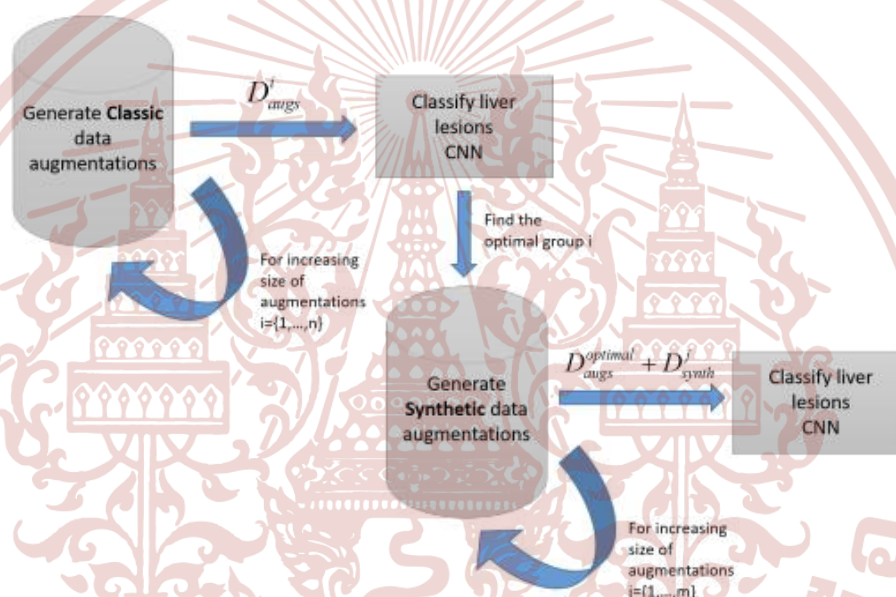


Fig. 7. Experiment flowchart for evaluating synthetic data augmentation in the task of classifying liver lesion ROIs.

ภาพที่ 2.2 แสดงกระบวนการทดลองในการใช้ข้อมูลจำลอง (Synthetic Data) เพื่อเพิ่มประสิทธิภาพของแบบจำลอง

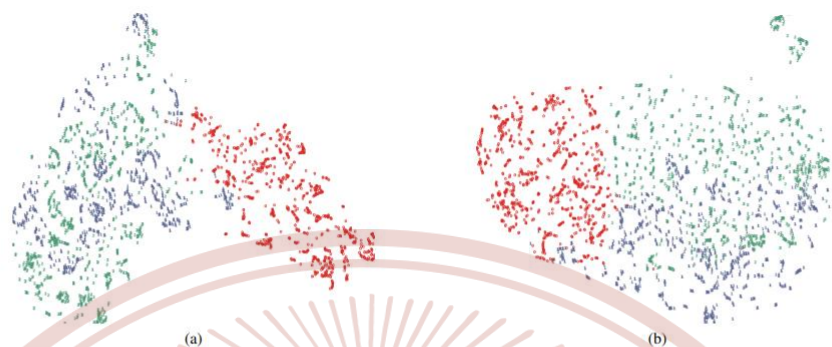


Fig. 9. T-SNE embedding of Cysts (red), Metastases (blue) and Hemangiomas (green) real lesion ROIs. (a) Features extracted from CNN-AUG (b) Features extracted from CNN-AUG-GAN.

TABLE III
SUMMARY OF EXPERTS' ASSESSMENT OF LESION ROI

	Classification Accuracy			Is ROI Real?	
	Real	Synthetic	Total Score	Total Score	
Expert 1	78%	77.5%	235\302=77.8%	189\302=62.5%	
Expert 2	69.2%	69.2%	209\302=69.2%	177\302=58.6%	

TABLE IV
PERFORMANCE COMPARISON FOR LIVER LESION CLASSIFICATION BETWEEN GENERATIVE MODELS

Method	Sensitivity	Specificity
CNN-AUG-GAN (DCGAN)	85.7%	92.4%
CNN-AUG-GAN (ACGAN)	81.3%	90.0%
ACGAN discriminator	79.1%	88.8%

ภาพที่ 2.3 การกระจายตัวของข้อมูลด้วยเทคนิค t-SNE สำหรับแบบจำลองที่ใช้ข้อมูล Augmentation และข้อมูลจำลอง (Synthetic Data)

พบว่าเมื่อใช้ข้อมูลแบบ augmentation เพียงอย่างเดียว (CNN-AUG) ข้อมูลยังมีการกระจายตัวที่ทับซ้อนกันระหว่างแต่ละกลุ่ม อย่างไรก็ตาม เมื่อมีการใช้ข้อมูลจำลอง (Synthetic Data) ร่วมกับข้อมูลจริง (CNN-AUG-GAN) พบว่ากลุ่มข้อมูลสามารถแยกออกจากกันได้อย่างชัดเจนมากยิ่งขึ้น ซึ่งสะท้อนให้เห็นถึงความสามารถของแบบจำลองในการเรียนรู้คุณลักษณะของข้อมูลได้ดีขึ้น

นอกจากนี้ ผลการประเมินประสิทธิภาพของแบบจำลองยังแสดงให้เห็นว่าค่าความไว (sensitivity) ของโมเดลเพิ่มขึ้นจาก 78.6% เป็น 85.7% และค่าความจำเพาะ (specificity) เพิ่มขึ้นจาก 88.4% เป็น 92.4% เมื่อมีการใช้ข้อมูลจำลองร่วมกับข้อมูลจริงในการฝึกแบบจำลอง ซึ่งแสดงให้เห็นถึงประสิทธิภาพของ Synthetic Dataset ในการช่วยเพิ่มความแม่นยำของแบบจำลองอย่างมีนัยสำคัญ [8]

2.3 เทคโนโลยี Internet of Things (IoT) ในภาคเกษตรกรรม

เทคโนโลยี Internet of Things (IoT) ได้เข้ามามีบทบาทสำคัญในการพัฒนาภาคเกษตรกรรมสมัยใหม่ โดยเฉพาะในบริบทของการเกษตรอัจฉริยะ (Smart Agriculture) ซึ่งมุ่งเน้นการใช้ข้อมูลจากอุปกรณ์เซนเซอร์เพื่อเพิ่มประสิทธิภาพในการบริหารจัดการการผลิตทางการเกษตร

IoT เป็นระบบที่เชื่อมโยงอุปกรณ์ตรวจวัด เช่น เซนเซอร์วัดอุณหภูมิ ความชื้น และคุณภาพน้ำ เข้ากับเครือข่ายอินเทอร์เน็ต เพื่อให้สามารถเก็บรวบรวม วิเคราะห์ และควบคุมสภาพแวดล้อมได้แบบเรียลไทม์

ในภาคเกษตรกรรม IoT ถูกนำมาใช้ในการตรวจวัดและติดตามสภาพแวดล้อมของฟาร์ม เช่น อุณหภูมิ ความชื้น และคุณภาพน้ำ ซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อผลผลิตและประสิทธิภาพการเลี้ยงสัตว์น้ำ โดยงานวิจัยของ Karle et al. ได้นำเสนอระบบโรงเรือนอัจฉริยะที่ใช้ Raspberry Pi ร่วมกับเซนเซอร์ IoT สำหรับตรวจวัดสภาพแวดล้อม ซึ่งช่วยให้สามารถบริหารจัดการการเพาะปลูกได้อย่างมีประสิทธิภาพมากขึ้น [9]

ผลการศึกษพบว่าระบบสามารถทำงานได้อย่างมีประสิทธิภาพ และช่วยให้การจัดการด้านการเกษตรมีความแม่นยำมากยิ่งขึ้น ลดการใช้แรงงาน และเพิ่มประสิทธิภาพในการใช้น้ำและทรัพยากรต่าง ๆ ได้อย่างเหมาะสม ซึ่งแสดงให้เห็นถึงศักยภาพของการนำเทคโนโลยี IoT มาประยุกต์ใช้ในงานด้านการเกษตรอัจฉริยะได้อย่างมีประสิทธิภาพ [10]

2.4 Machine Learning และการพยากรณ์ข้อมูล

Machine Learning เป็นเทคนิคที่มีบทบาทสำคัญในการพยากรณ์ข้อมูลในยุคดิจิทัล โดยเฉพาะอย่างยิ่งในบริบทของการวิเคราะห์แนวโน้ม ความต้องการสินค้า และพฤติกรรมผู้บริโภค งานวิจัยหลากหลายฉบับได้นำอัลกอริธึม เช่น LSTM, Random Forest, Gradient Boosted Trees และ Support Vector Machine มาประยุกต์ใช้ในการพยากรณ์ในหลากหลายบริบท ทั้งภาคเกษตรกรรม ค่าปลี และระบบขนส่ง

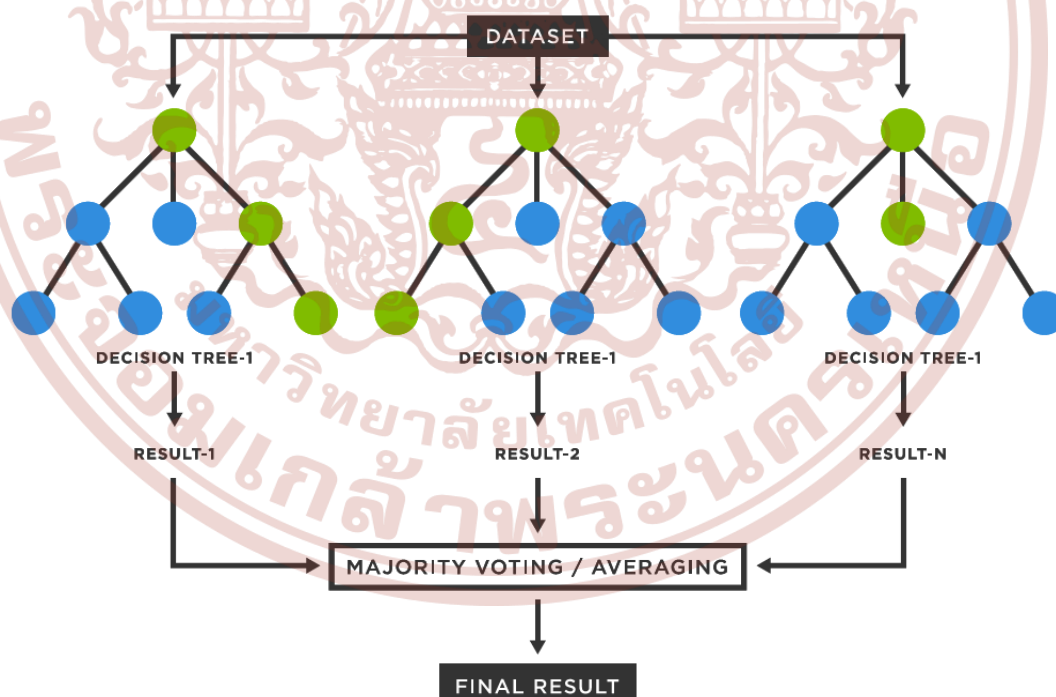
A. Kantasard และคณะ ได้ศึกษาการใช้ LSTM (Long Short-Term Memory) เพื่อพยากรณ์ความต้องการสินค้าเกษตรในระบบ Physical Internet โดยเปรียบเทียบกับ ARIMAX, SVR และ MLR พร้อมการจูนไฮเปอร์พารามิเตอร์ด้วย Genetic Algorithm และ Scatter Search พบว่า LSTM เหมาะสมที่สุดสำหรับข้อมูลที่มีความผันผวนสูง และให้ความแม่นยำสูงในการพยากรณ์ [11]

W. Kusonkhum และคณะ ได้ศึกษาโมเดล Machine Learning สำหรับการพยากรณ์จำนวนผู้โดยสารรถไฟฟ้าใต้ดิน MRT โดยใช้ข้อมูลจากปัจจัยต่าง ๆ เช่น วัน เวลา สภาพอากาศ และจำนวนผู้โดยสารในอดีต โดยใช้อัลกอริธึมเช่น Artificial Neural Network, Random Forest และ Decision Tree ผลการวิจัยพบว่าโมเดล Artificial Neural Network มีความแม่นยำสูงที่สุดในการทำนายผู้โดยสารแบบเรียลไทม์ (Accuracy > 0.85) [12]

C. Phakdeephrom และ A. Phuangchaktada ได้ศึกษาการเปรียบเทียบเทคนิคการพยากรณ์ข้อมูลด้วย 6 เทคนิค Machine Learning ได้แก่ Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees และ Support Vector Machine เทียบกับวิธีแบบดั้งเดิม เช่น ARIMA และ Holt-Winters ผลการวิจัยนี้ พบว่า Gradient Boosted Trees มีประสิทธิภาพดีที่สุดในการพยากรณ์ โดยให้ค่า MAE, RMSE และ MAPE ต่ำที่สุด [13]

2.5 Random Forest (RF)

L. Breiman ได้เสนออัลกอริธึม Random Forest ซึ่งเป็นเทคนิคการเรียนรู้ของเครื่องแบบ ensemble learning ที่รวมผลลัพธ์จากต้นไม้ตัดสินใจหลายต้นเพื่อเพิ่มความแม่นยำในการจำแนกและพยากรณ์ข้อมูล โดยอัลกอริธึมดังกล่าวใช้วิธีการสุ่มตัวอย่างข้อมูล (bootstrap sampling) และการเลือกตัวแปรแบบสุ่มในแต่ละโหนดของต้นไม้ตัดสินใจเพื่อลดปัญหา overfitting และเพิ่มความเสถียรของแบบจำลอง [14]



ภาพที่ 2.4 กระบวนการทำงานของ Random Forest

E. Asamoah และคณะ ได้ประยุกต์ใช้ Random Forest ในการพยากรณ์ผลผลิตทางการเกษตร โดยใช้ข้อมูลด้านสภาพดิน ภูมิอากาศ และปัจจัยการจัดการการเพาะปลูกเป็นตัวแปรในการ

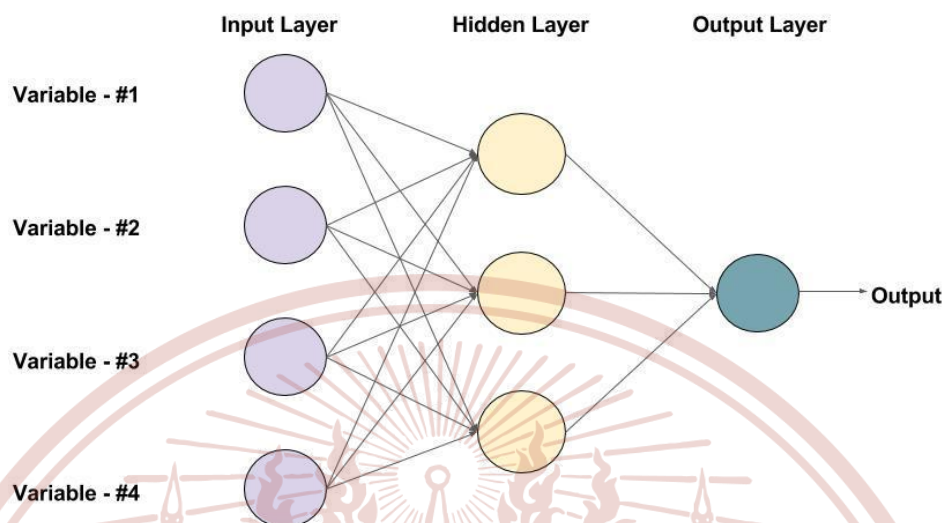
สร้างแบบจำลอง ผลการศึกษาพบว่า Random Forest สามารถพยากรณ์ผลผลิตได้อย่างมีประสิทธิภาพและให้ค่าความแม่นยำสูงเมื่อเปรียบเทียบกับวิธีการวิเคราะห์แบบดั้งเดิม

ในเชิงเทคนิค RF ทำงานโดยใช้การตัดสินใจแบบเสียงข้างมาก (majority vote) หรือการหาค่าเฉลี่ย (average) จากผลลัพธ์ของต้นไม้หลายต้นเพื่อคาดการณ์ผลลัพธ์สุดท้าย และสามารถแสดงความสำคัญของตัวแปร (feature importance) ได้ ซึ่งมีประโยชน์อย่างยิ่งในการระบุปัจจัยสำคัญที่มีผลต่อผลลัพธ์ เช่น ในกรณีศึกษาดังกล่าวพบว่า ปัจจัยที่มีผลต่อผลการดำเนินงานมากที่สุดในช่วงปีแรกหลังเข้าตลาด คือ ความสามารถในการใช้สินทรัพย์ถาวร (Tangibility) ในขณะที่ปัจจัยที่สำคัญในช่วงปีที่สองและสาม คือ เงินทุนที่ระดมได้จาก IPO และอัตรากำไรสุทธิ (Net Profit Margin) ด้วยคุณสมบัติของ RF ที่สามารถรับมือกับข้อมูลจำนวนมาก มีความสามารถในการวิเคราะห์ปัจจัยหลายมิติ พร้อมทั้งสามารถจัดการกับความไม่สมดุลของข้อมูล (imbalanced data) ได้ดี RF จึงเป็นอัลกอริธึมที่เหมาะสมอย่างยิ่งในการนำมาใช้ในการวิจัยด้านการวิเคราะห์และพยากรณ์ในหลากหลายสาขา [15]

2.5 Neural Network

โครงข่ายประสาทเทียม Neural Network เป็นเทคนิคการเรียนรู้ของเครื่องที่ได้รับแรงบันดาลใจจากระบบประสาทชีวภาพในมนุษย์ โดยสามารถเรียนรู้จากข้อมูลที่ซับซ้อน จัดการกับความไม่เป็นเชิงเส้นของข้อมูล (Nonlinear Data) และประยุกต์ใช้ได้หลากหลายสาขา ทั้งด้านเศรษฐกิจ การเงิน การเกษตร และสิ่งแวดล้อม

G. P. Zhang ได้ศึกษาการใช้ Neural Network สำหรับการพยากรณ์ข้อมูลอนุกรมเวลา โดยเปรียบเทียบประสิทธิภาพของแบบจำลอง Neural Network กับแบบจำลองทางสถิติแบบ Autoregressive Integrated Moving Average (ARIMA) และเสนอการผสมผสานระหว่างสองวิธีในรูปแบบ Hybrid ARIMA-ANN model เพื่อใช้ประโยชน์จากจุดเด่นของทั้งสองแบบจำลอง เมื่อได้ผลสรุปการวิจัยจากชุดข้อมูลจริงหลายชุดพบว่าแบบจำลอง Hybrid สามารถให้ค่าความแม่นยำในการพยากรณ์สูงกว่าแบบจำลองเดี่ยว โดยมีค่าความคลาดเคลื่อนลดลงอย่างมีนัยสำคัญ เช่น ค่า Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) ต่ำกว่าโมเดล ARIMA และ ANN ที่ใช้เพียงอย่างเดียว ซึ่งแสดงให้เห็นว่า ANN สามารถเรียนรู้ความสัมพันธ์แบบไม่เชิงเส้นของข้อมูลได้ดี และเมื่อใช้ร่วมกับแบบจำลองเชิงเส้นจะช่วยเพิ่มประสิทธิภาพของการพยากรณ์ได้มากขึ้น [16]



An example of a Feed-forward Neural Network with one hidden layer (with 3 neurons)

ภาพที่ 2.5 กระบวนการทำงานของโครงข่ายประสาทเทียม Neural Network

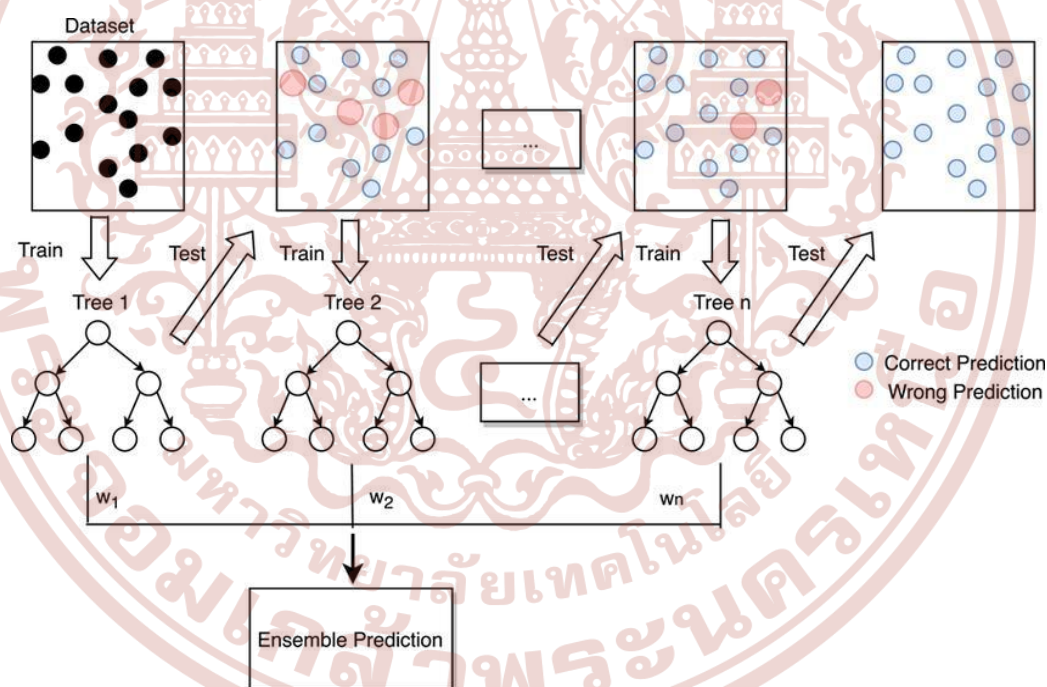
S. Pereira และคณะ ได้พัฒนาแบบจำลอง Neural Network สำหรับการพยากรณ์ข้อมูลพลังงานแสงอาทิตย์ โดยใช้ตัวแปรด้านสภาพอากาศและข้อมูลรังสีดวงอาทิตย์ในการฝึกโมเดล สรุปผลการวิจัยได้ ดังต่อไปนี้ แบบจำลอง ANN สามารถให้ความแม่นยำในการพยากรณ์สูง โดยมีค่า $R^2 = 0.0646$, ค่า $MAE = 21.1 \text{ W/m}^2$ และค่า $RMSE = 27.9 \text{ W/m}^2$ สำหรับการพยากรณ์ในช่วงเวลาหนึ่งวันข้างหน้า ซึ่งแสดงให้เห็นว่าโครงข่ายประสาทเทียมมีประสิทธิภาพในการวิเคราะห์ข้อมูลที่มีความสัมพันธ์ซับซ้อนและไม่เชิงเส้น และสามารถนำไปประยุกต์ใช้ในงานพยากรณ์ข้อมูลด้านพลังงานและสิ่งแวดล้อมได้อย่างมีประสิทธิภาพ [17]

M. S. A. Bakar และคณะ ได้ศึกษาการใช้ Neural Network ในการพยากรณ์ผลผลิตทางการเกษตร โดยใช้ข้อมูลสภาพภูมิอากาศ เช่น อุณหภูมิ ปริมาณฝน และความชื้น เป็นตัวแปรในการฝึกแบบจำลอง ANN สรุปผลการวิจัยได้ ดังต่อไปนี้ แบบจำลอง ANN สามารถพยากรณ์ผลผลิตพืชได้อย่างมีประสิทธิภาพ โดยมีค่าความแม่นยำสูงเมื่อประเมินด้วยตัวชี้วัด Mean Absolute Error (MAE) = 0.045, Root Mean Square Error (RMSE) = 0.067 และ Coefficient of Determination (R^2) = 0.91 ซึ่งแสดงให้เห็นว่า ANN สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนของข้อมูลด้านการเกษตรได้อย่างมีประสิทธิภาพ และเหมาะสมสำหรับการพยากรณ์ข้อมูลที่มีความไม่แน่นอนสูง [18]

2.6 Gradient Boosting (GB)

Gradient Boosting เป็นเทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) ซึ่งถูกนำมาใช้ในการแก้ปัญหการพยากรณ์ (Prediction) โดยเฉพาะอย่างยิ่งในการวิเคราะห์ข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (Non-linear) และซับซ้อน

J. H. Friedman ได้เสนออัลกอริธึม Gradient Boosting Machine (GBM) ซึ่งเป็นเทคนิคการเรียนรู้แบบ ensemble learning ที่ใช้หลักการสร้างแบบจำลองหลายตัวในลักษณะลำดับ โดยแต่ละแบบจำลองจะถูกสร้างขึ้นเพื่อแก้ไขข้อผิดพลาดของแบบจำลองก่อนหน้า อัลกอริธึม Gradient Boosting ใช้แนวคิดของ gradient descent optimization ในการปรับค่าพารามิเตอร์ของแบบจำลองเพื่อให้ค่าความคลาดเคลื่อนลดลงอย่างต่อเนื่อง สรุปผลการวิจัยได้ ดังต่อไปนี้ Gradient Boosting สามารถให้ความแม่นยำในการพยากรณ์สูงและสามารถจัดการข้อมูลที่มีความสัมพันธ์ซับซ้อนได้อย่างมีประสิทธิภาพ [19]



ภาพที่ 2.6 กระบวนการทำงานของอัลกอริธึม Gradient Boosting (GB)

P. Mahesh and R. Soundrapandiyan ได้ศึกษาการใช้เทคนิค Gradient Boosting โดยเฉพาะอัลกอริธึม Extreme Gradient Boosting (XGBoost) และ Gradient Boosting Regressor (GBR) ในการพยากรณ์ผลผลิตข้าวโพดพันธุ์ Zhengdan 958 โดยใช้ข้อมูลด้านสภาพอากาศและข้อมูลทางการเกษตรเป็นตัวแปรในการฝึกแบบจำลอง สรุปผลการวิจัยได้ ดังต่อไปนี้ แบบจำลอง Gradient Boosting สามารถพยากรณ์ผลผลิตพืชได้อย่างมีประสิทธิภาพ โดยแบบจำลอง

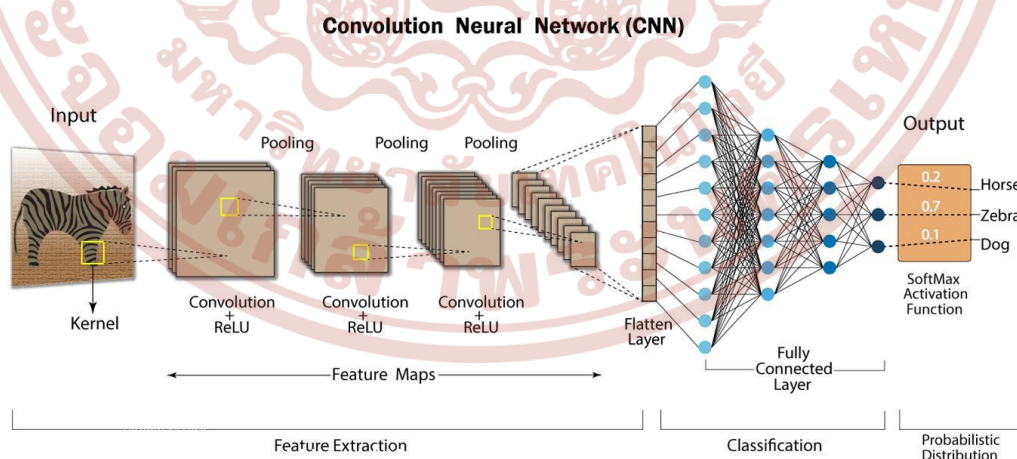
XGBoost ให้ค่า RMSE = 0.37, ค่า MAE = 0.26 และค่า $R^2 = 0.87$ ขณะที่แบบจำลอง Gradient Boosting Regressor (GBR) ให้ค่า RMSE = 0.39, ค่า MAE = 0.27 และค่า $R^2 = 0.86$ ซึ่งแสดงให้เห็นว่าอัลกอริธึม Gradient Boosting สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนระหว่างตัวแปรด้านสภาพแวดล้อมและผลผลิตทางการเกษตรได้อย่างมีประสิทธิภาพ และมีศักยภาพในการนำมาใช้พยากรณ์ผลผลิตทางการเกษตรในสถานการณ์จริงได้อย่างแม่นยำ [20]

2.8 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) เป็นสถาปัตยกรรมของโครงข่ายประสาทเทียมที่ได้รับความนิยมอย่างแพร่หลายในงานที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลเชิงโครงสร้าง เช่น ภาพ เสียง และสัญญาณ โดยมีจุดเด่นในการสกัดคุณลักษณะ (Feature Extraction) จากข้อมูลโดยอัตโนมัติผ่านชั้น Convolution และ Pooling

Y. LeCun และคณะกล่าวว่า CNN สามารถเรียนรู้รูปแบบเชิงพื้นที่ (spatial patterns) ได้อย่างมีประสิทธิภาพ โดยลดจำนวนพารามิเตอร์ของโมเดลเมื่อเทียบกับโครงข่ายแบบ Fully Connected ทำให้สามารถฝึกสอนโมเดลได้เร็วขึ้นและลดปัญหา overfitting [28]

T. Li และคณะกล่าวว่า CNN จะถูกใช้อย่างแพร่หลายในงานด้านภาพ แต่สามารถนำมาประยุกต์ใช้กับข้อมูลจากเซนเซอร์ IoT ได้เช่นกัน โดยเฉพาะข้อมูลที่มีความสัมพันธ์ระหว่างตัวแปรหลายมิติ ซึ่งสอดคล้องกับงานวิจัยของ Li et al. ที่แสดงให้เห็นว่าโมเดล Deep Learning สามารถนำมาใช้ในการวิเคราะห์และพยากรณ์คุณภาพน้ำได้อย่างมีประสิทธิภาพ [23]



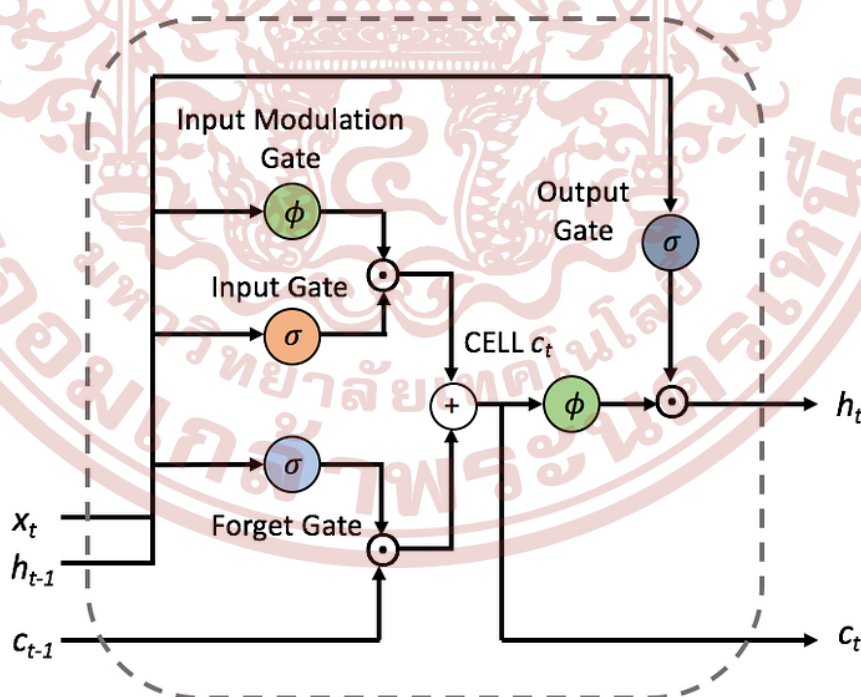
ภาพที่ 2.7 กระบวนการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชัน
(Convolutional Neural Network: CNN)

2.9 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) เป็นโครงข่ายประสาทเทียมในกลุ่ม Recurrent Neural Network (RNN) ที่ถูกพัฒนาขึ้นเพื่อแก้ปัญหา vanishing gradient ซึ่งมักเกิดในโครงข่าย RNN แบบดั้งเดิม โดย LSTM มีความสามารถในการเรียนรู้ข้อมูลลำดับเวลา (time series data) และจดจำความสัมพันธ์ระยะยาวของข้อมูลได้อย่างมีประสิทธิภาพ

S. Hochreiter และ J. Schmidhuber ได้เสนอ LSTM โดยใช้โครงสร้างของหน่วยความจำ (memory cell) และกลไกประตู (gates) ได้แก่ input gate, forget gate และ output gate ซึ่งช่วยควบคุมการไหลของข้อมูลภายในเครือข่าย ทำให้โมเดลสามารถเลือกจดจำหรือหลงลืมข้อมูลในช่วงเวลาต่าง ๆ ได้อย่างเหมาะสม [29]

T. Li และคณะกล่าวว่า ในบริบทของการเพาะเลี้ยงสัตว์น้ำ ข้อมูลคุณภาพน้ำจากเซนเซอร์ IoT มักมีลักษณะเป็นข้อมูลต่อเนื่องตามเวลา เช่น การเปลี่ยนแปลงของอุณหภูมิ ค่า pH และปริมาณออกซิเจนละลายน้ำ ซึ่งเหมาะสมอย่างยิ่งสำหรับการนำ LSTM มาประยุกต์ใช้ โดย LSTM สามารถพยากรณ์คุณภาพน้ำได้อย่างแม่นยำสูง เนื่องจากสามารถเรียนรู้แนวโน้มและความสัมพันธ์ของข้อมูลในระยะยาวได้ดีกว่าโมเดลทั่วไป [23]



ภาพที่ 2.8 กระบวนการทำงานของโครงข่ายประสาทเทียมแบบ Long Short-Term Memory (LSTM)

2.10 งานวิจัยที่เกี่ยวข้อง

M. A. M. M. Hridoy และคณะได้พัฒนาแบบจำลอง Machine Learning สำหรับการพยากรณ์คุณภาพน้ำในระบบเพาะเลี้ยงปลาในกระชัง โดยใช้ข้อมูลจากเซ็นเซอร์ IoT ที่ตรวจวัดค่า DO, pH และอุณหภูมิ ในบ่อเลี้ยงปลาในกระชัง จากนั้นนำข้อมูลมาวิเคราะห์ด้วยเทคนิค ensemble machine learning เพื่อเพิ่มความแม่นยำของการพยากรณ์ สรุปผลการวิจัยได้ ดังต่อไปนี้ ระบบสามารถพยากรณ์ค่าดัชนีคุณภาพน้ำได้อย่างแม่นยำและช่วยเพิ่มประสิทธิภาพในการจัดการฟาร์มเลี้ยงปลาในกระชัง [21]

R. P. Shete และคณะได้พัฒนาแบบจำลอง IoT-driven ensemble machine learning สำหรับการพยากรณ์คุณภาพน้ำในระบบเพาะเลี้ยงปลาในกระชัง โดยใช้ข้อมูลจากเซ็นเซอร์ตรวจวัดค่าพารามิเตอร์ของน้ำ เช่น ค่า Dissolved Oxygen (DO) อุณหภูมิ และค่า pH ในการฝึกแบบจำลอง Machine Learning พบว่าแบบจำลอง ensemble สามารถพยากรณ์ค่าดัชนีคุณภาพน้ำได้อย่างแม่นยำและช่วยเพิ่มประสิทธิภาพในการบริหารจัดการฟาร์มเพาะเลี้ยงสัตว์น้ำ [22]

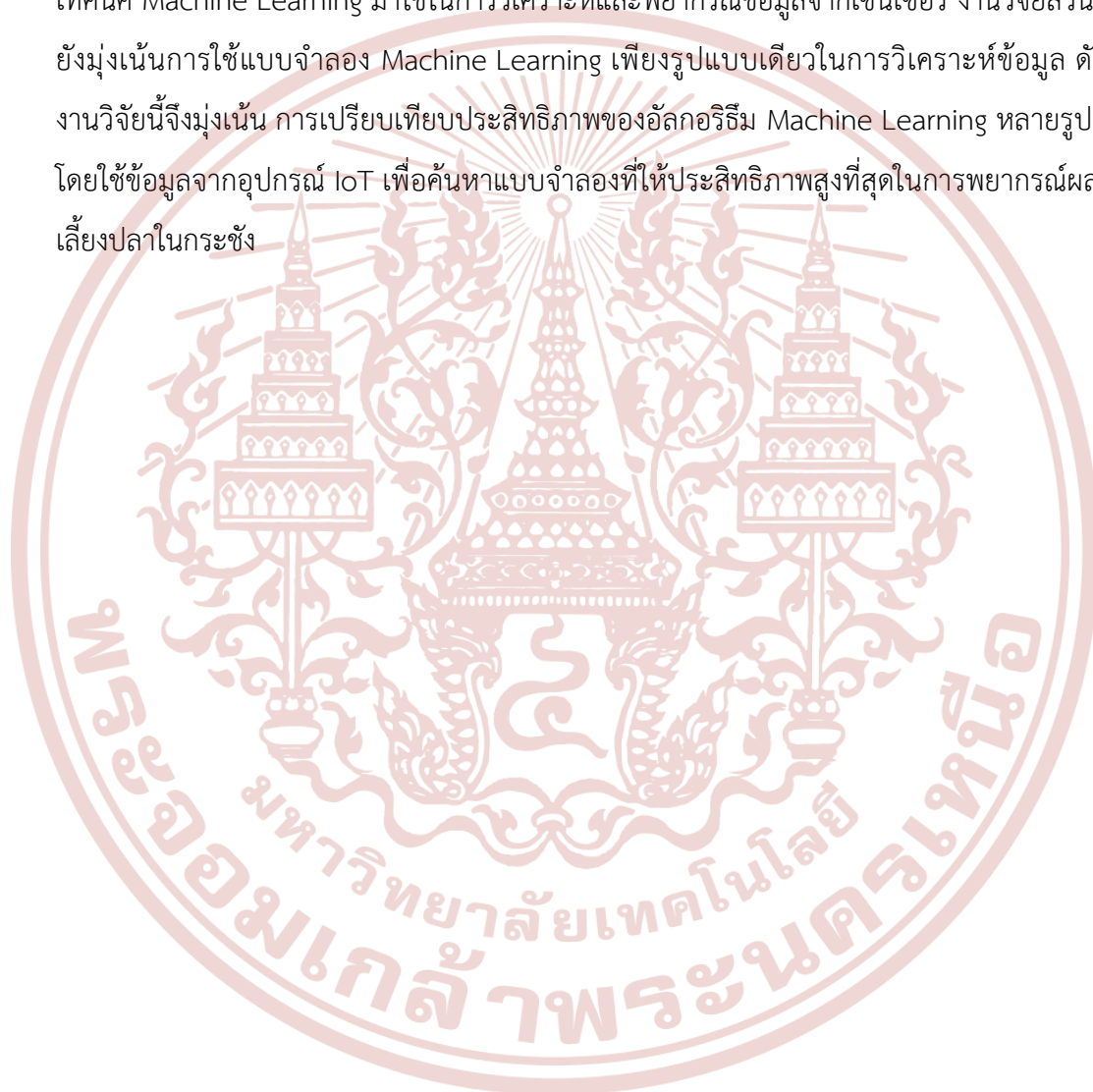
T. Li และคณะได้ศึกษาการใช้เทคนิค Machine Learning สำหรับการพยากรณ์คุณภาพน้ำในระบบเพาะเลี้ยงสัตว์น้ำ โดยเปรียบเทียบแบบจำลองหลายรูปแบบ เช่น Support Vector Machine (SVM), Backpropagation Neural Network (BPNN) และ LSSVM สรุปผลการวิจัยได้ ดังต่อไปนี้ แบบจำลอง SVM ให้ค่าความแม่นยำในการพยากรณ์สูงที่สุด โดยมีค่า Accuracy ประมาณ 99% ซึ่งแสดงให้เห็นถึงศักยภาพของเทคนิค Machine Learning ในการวิเคราะห์ข้อมูลคุณภาพน้ำที่มีความสัมพันธ์ซับซ้อน [23]

K. Anupama และคณะได้พัฒนาระบบตรวจสอบคุณภาพน้ำในระบบเพาะเลี้ยงสัตว์น้ำโดยใช้เทคโนโลยี IoT ร่วมกับ Machine Learning โดยใช้เซ็นเซอร์ตรวจวัดค่า pH อุณหภูมิ และความขุ่นของน้ำ จากนั้นส่งข้อมูลไปยังระบบคลาวด์เพื่อวิเคราะห์แบบเรียลไทม์ จากการวิจัยนี้ทำให้ระบบสามารถช่วยตรวจสอบคุณภาพน้ำได้อย่างมีประสิทธิภาพและช่วยให้เกษตรกรสามารถตัดสินใจในการจัดการฟาร์มได้อย่างเหมาะสม [24]

S. M. Naim และคณะได้ศึกษาการใช้เทคนิค Extreme Gradient Boosting (XGBoost) ในการจำแนกคุณภาพน้ำสำหรับการเพาะเลี้ยงปลาในกระชัง โดยใช้ข้อมูลพารามิเตอร์ของน้ำ เช่น Ammonia, Nitrite, Dissolved Oxygen, Turbidity, Temperature และ pH จากการวิจัยนี้ทำให้แบบจำลอง XGBoost สามารถให้ค่าความแม่นยำสูงที่สุดถึง 98.45% และมีค่าเฉลี่ยประมาณ 98.05%

ซึ่งแสดงให้เห็นถึงประสิทธิภาพของอัลกอริธึม Gradient Boosting ในการวิเคราะห์ข้อมูลคุณภาพน้ำจากระบบ IoT [25]

จากงานวิจัยที่ผ่านมา พบว่ามีการนำเทคโนโลยี Internet of Things (IoT) มาใช้ในการเก็บรวบรวมข้อมูลคุณภาพน้ำและสภาพแวดล้อมในการเพาะเลี้ยงสัตว์น้ำอย่างแพร่หลาย และมีการนำเทคนิค Machine Learning มาใช้ในการวิเคราะห์และพยากรณ์ข้อมูลจากเซ็นเซอร์ งานวิจัยส่วนใหญ่ยังมุ่งเน้นการใช้แบบจำลอง Machine Learning เพียงรูปแบบเดียวในการวิเคราะห์ข้อมูล ดังนั้นงานวิจัยนี้จึงมุ่งเน้น การเปรียบเทียบประสิทธิภาพของอัลกอริธึม Machine Learning หลายรูปแบบ โดยใช้ข้อมูลจากอุปกรณ์ IoT เพื่อค้นหาแบบจำลองที่ให้ประสิทธิภาพสูงที่สุดในการพยากรณ์ผลการเลี้ยงปลาในกระชัง



บทที่ 3

วิธีการดำเนินงานวิจัย

การวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพของอัลกอริธึม Machine Learning ในการพยากรณ์ผลการเลี้ยงปลาในกระชังด้วยข้อมูลจากอุปกรณ์ Internet of Things (IoT) มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของอัลกอริธึม Machine Learning หลายรูปแบบในการพยากรณ์ผลการเลี้ยงปลาในกระชัง โดยใช้ข้อมูลจากเซ็นเซอร์ที่ติดตั้งในระบบ IoT เพื่อตรวจวัดสภาพแวดล้อมของบ่อเลี้ยงปลาในกระชัง ซึ่งผู้วิจัยได้กำหนดขั้นตอนการดำเนินงานวิจัยดังต่อไปนี้

3.1 กรอบแนวคิดการวิจัย

3.2 ขั้นตอนการดำเนินงานวิจัย

3.3 สรุปขั้นตอนการดำเนินการวิจัย

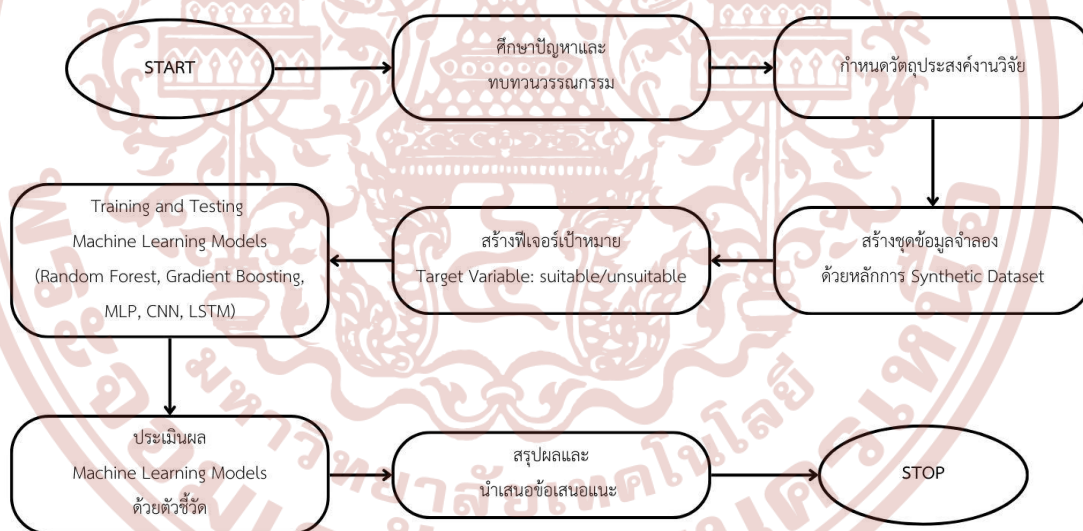
3.1 กรอบแนวคิดการวิจัย

งานวิจัยนี้มีแนวคิดในการนำเทคโนโลยี Machine Learning มาประยุกต์ใช้ในการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการเพาะเลี้ยงปลาในกระชัง เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์ความเหมาะสมของสภาพแวดล้อมในการเลี้ยงปลาในกระชัง โดยอาศัยข้อมูลที่ได้จากพารามิเตอร์คุณภาพน้ำซึ่งมักถูกตรวจวัดด้วยอุปกรณ์ Internet of Things (IoT) เช่น อุณหภูมิของน้ำ ค่า pH และปริมาณออกซิเจนละลายน้ำ ซึ่งเป็นปัจจัยสำคัญที่มีผลต่อการเจริญเติบโตของปลา

ในการวิจัยครั้งนี้ ผู้วิจัยได้ใช้แนวคิดในการสร้างชุดข้อมูลสำหรับการวิเคราะห์ด้วยวิธี Synthetic Dataset เพื่อจำลองข้อมูลที่มีความสัมพันธ์กับคุณภาพน้ำในการเลี้ยงปลาในกระชัง จากนั้นกำหนดตัวแปรเป้าหมาย (Target Variable) เพื่อจำแนกความเหมาะสมของสภาพแวดล้อมในการเลี้ยงปลาในกระชังเป็น 2 กลุ่ม ได้แก่ Suitable และ Unsuitable โดยอาศัยเกณฑ์ค่าพารามิเตอร์ของคุณภาพน้ำที่เหมาะสมต่อการเลี้ยงปลาในกระชัง หลังจากได้ชุดข้อมูลสำหรับการวิเคราะห์แล้ว ผู้วิจัยได้นำข้อมูลเข้าสู่กระบวนการสร้างแบบจำลองด้วยอัลกอริธึม Machine Learning จำนวน 5 รูปแบบ ได้แก่ Random Forest, Gradient Boosting, Neural Network, Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) เพื่อใช้ในการเรียนรู้รูปแบบของข้อมูลและสร้างแบบจำลองสำหรับการพยากรณ์ จากนั้นทำการฝึกแบบจำลอง (Training) และทดสอบแบบจำลอง (Testing) เพื่อประเมินประสิทธิภาพของอัลกอริธึมแต่ละรูปแบบ การวิจัยครั้งนี้ผู้วิจัยได้เลือกใช้ Library สองระบบหลัก ได้แก่ scikit-learn สำหรับอัลกอริธึม

Machine Learning พื้นฐาน (เช่น Random Forest, Gradient Boosting) เนื่องจากมีประสิทธิภาพสูงในการจัดการกับข้อมูลในรูปแบบตาราง (Tabular Data) และใช้ทรัพยากรระบบน้อยกว่า และ TensorFlow สำหรับโมเดลโครงข่ายประสาทเทียมเชิงลึก (Deep Learning) เช่น CNN และ LSTM ซึ่งถูกออกแบบมาเพื่อรองรับโครงสร้างโมเดลที่มีความซับซ้อนและการประมวลผลแบบขนาน (Parallel Computing) เมื่อทำงานเปรียบเทียบประสิทธิภาพโมเดลแล้ว ผู้วิจัยจะใช้ตัวชี้วัดทางสถิติสำหรับการประเมินผลแบบจำลอง เช่น Accuracy, Precision, Recall, F1-score และ Area Under the ROC Curve (AUC) เพื่อนำผลลัพธ์มาเปรียบเทียบประสิทธิภาพของอัลกอริธึม Machine Learning แต่ละชนิด ใน

ผลการวิจัยที่ได้จะช่วยให้สามารถระบุอัลกอริธึม Machine Learning ที่มีประสิทธิภาพสูงที่สุดสำหรับการพยากรณ์ความเหมาะสมของสภาพแวดล้อมในการเลี้ยงปลาในกระชัง ซึ่งสามารถนำไปประยุกต์ใช้ในการพัฒนาระบบวิเคราะห์ข้อมูลจากอุปกรณ์ IoT เพื่อสนับสนุนการตัดสินใจในการบริหารจัดการฟาร์มเพาะเลี้ยงสัตว์น้ำได้อย่างมีประสิทธิภาพ



ภาพที่ 3.1 กระบวนการวิจัย

3.2 ขั้นตอนการดำเนินงานวิจัย

3.2.1 การสร้างชุดข้อมูลจำลอง (Synthetic Dataset)

ผู้วิจัยดำเนินการสร้างชุดข้อมูลจำลอง (Synthetic Dataset) เพื่อใช้เป็นข้อมูลสำหรับการวิเคราะห์และการสร้างแบบจำลอง Machine Learning โดยชุดข้อมูลดังกล่าวถูกออกแบบให้จำลองลักษณะของข้อมูลคุณภาพน้ำที่มักถูกตรวจวัดด้วยอุปกรณ์ Internet of Things (IoT) ในระบบเพาะเลี้ยงปลาในกระชัง ตัวแปรที่ใช้ในการสร้างชุดข้อมูลประกอบด้วยพารามิเตอร์คุณภาพน้ำจำนวน 3 ตัวแปร ได้แก่ อุณหภูมิของน้ำ (Temperature) ซึ่งมีผลต่อการเจริญเติบโตและการดำรงชีวิตของปลา ช่วงที่เหมาะสมอยู่ที่ 25–30 องศาเซลเซียส ค่าความเป็นกรด-ด่างของน้ำ (pH) ซึ่งมีผลต่อระบบสรีรวิทยาของปลาและความสมดุลของระบบนิเวศในบ่อเลี้ยง ช่วงที่เหมาะสมอยู่ที่ 7.0–8.0 และปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen: DO) ซึ่งเป็นปัจจัยสำคัญต่อการหายใจของปลา มากกว่า 4 มิลลิกรัมต่อลิตร

ช่วงค่าของตัวแปรดังกล่าวถูกกำหนดตามเกณฑ์คุณภาพน้ำที่เหมาะสมสำหรับการเลี้ยงปลา จากนั้นทำการสุ่มสร้างข้อมูลเพื่อจำลองสถานการณ์คุณภาพน้ำในบ่อเลี้ยงปลา เพื่อใช้เป็นข้อมูลสำหรับการวิเคราะห์ด้วยเทคนิค Machine Learning โดยมีกระบวนการสร้างข้อมูลอาศัยหลักการทางสถิติ (Statistical Simulation) ดังนี้

1) การกำหนดขอบเขตและรูปแบบการแจกแจงข้อมูล ผู้วิจัยได้กำหนดช่วงค่าของตัวแปรตามเกณฑ์มาตรฐานคุณภาพน้ำที่เหมาะสมสำหรับการเลี้ยงปลา จากนั้นทำการสุ่มค่าข้อมูลโดยใช้วิธี Random Sampling จากการแจกแจงแบบสม่ำเสมอ (Uniform Distribution) เพื่อให้ค่าที่สุ่มได้กระจายตัวอย่างเท่าเทียมกันภายในช่วงที่กำหนด

ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function: PDF) ของการแจกแจงแบบสม่ำเสมอสามารถเขียนได้ดังสมการ

$$f(x) = \frac{1}{b - a}$$

กำหนดให้

x คือค่าของตัวแปรที่ถูกสุ่มขึ้น

a คือค่าต่ำสุดของช่วงข้อมูล

b คือค่าสูงสุดของช่วงข้อมูล

การแจกแจงแบบสม่ำเสมอมีคุณสมบัติที่ค่าทุกค่าภายในช่วง a , b มีโอกาสเกิดขึ้นเท่ากัน จึงเหมาะสำหรับการสร้างข้อมูลจำลองในกรณีที่ต้องการให้ข้อมูลครอบคลุมช่วงค่าที่กำหนด [26]

ในการวิจัยนี้ ผู้วิจัยได้กำหนดช่วงค่าของตัวแปรคุณภาพน้ำเพื่อใช้ในการจำลองข้อมูล ได้แก่

- Temperature : 25 – 30 °C
- pH : 6.5 – 9.0
- Dissolved Oxygen : 5 – 8 mg/L

2) การเพิ่มสัญญาณรบกวนของข้อมูล (Noise Injection) เพื่อให้ข้อมูลจำลองมีลักษณะใกล้เคียงกับข้อมูลที่ได้จากเซนเซอร์ในสภาพแวดล้อมจริง ผู้วิจัยได้เพิ่มสัญญาณรบกวน (Noise) ลงในข้อมูลที่จำลองขึ้น โดยใช้การแจกแจงแบบปกติ (Normal Distribution) หรือ Gaussian Noise ซึ่งมักใช้ในการจำลองความคลาดเคลื่อนของเซนเซอร์ในระบบ Internet of Things (IoT) การเพิ่มสัญญาณรบกวนสามารถเขียนได้ดังสมการ

$$X_{final} = X_{simulated} + \epsilon$$

กำหนดให้

$X_{simulated}$ คือค่าข้อมูลที่ถูกสร้างขึ้นจากการจำลอง

X_{final} คือค่าข้อมูลหลังจากเพิ่มสัญญาณรบกวน

ϵ คือค่าความแปรปรวนสุ่มจากการแจกแจงแบบปกติ

และสามารถกำหนดได้ว่า

$$\epsilon \sim N(0, \sigma^2)$$

กำหนดให้

N คือการแจกแจงแบบปกติ (Normal Distribution)

0 คือค่าเฉลี่ยของสัญญาณรบกวน

σ^2 คือค่าความแปรปรวนของข้อมูล

การเพิ่ม Gaussian Noise ช่วยให้ข้อมูลจำลองมีความสมจริงมากขึ้นและสามารถสะท้อนลักษณะของข้อมูลที่ได้จากเซนเซอร์ในระบบ IoT ได้ดียิ่งขึ้น [27]

จากกระบวนการข้างต้น ผู้วิจัยจึงได้สร้างชุดข้อมูลจำลองที่มีลักษณะใกล้เคียงกับข้อมูลคุณภาพน้ำจริง เพื่อนำไปใช้ในการฝึกและทดสอบแบบจำลอง Machine Learning สำหรับการพยากรณ์ความเหมาะสมของสภาพแวดล้อมในการเลี้ยงปลา

Parameter	Range
Temperature	25–30 °C
pH	6.5–9.0
Dissolved Oxygen	5–8 mg/L

ตารางที่ 3.1 ช่วงค่าพารามิเตอร์คุณภาพน้ำที่ใช้ในการสร้างชุดข้อมูลสำหรับการวิเคราะห์

3.2.2 การเตรียมข้อมูล (Data Preparation) หลังจากสร้างชุดข้อมูลจำลอง (Synthetic Dataset) แล้ว ผู้วิจัยดำเนินการเตรียมข้อมูลก่อนนำไปใช้ในการสร้างแบบจำลอง Machine Learning เพื่อให้ข้อมูลมีความเหมาะสมสำหรับการวิเคราะห์และลดความคลาดเคลื่อนที่อาจเกิดขึ้นในกระบวนการเรียนรู้ของแบบจำลอง โดยมีขั้นตอนดังต่อไปนี้

3.2.2.1 การสร้างตัวแปรเป้าหมาย (Target Variable) ผู้วิจัยได้กำหนดตัวแปรเป้าหมาย (Target Variable) สำหรับการจำแนกความเหมาะสมของคุณภาพน้ำในการเลี้ยงปลา โดยใช้รูปแบบการจำแนกข้อมูลแบบสองกลุ่ม (Binary Classification) ซึ่งแบ่งออกเป็น 2 กลุ่ม ได้แก่

- Suitable ($Y = 1$) หมายถึง คุณภาพน้ำที่อยู่ในช่วงค่าที่เหมาะสมต่อการเลี้ยงปลา
- Unsuitable ($Y = 0$) หมายถึง คุณภาพน้ำที่อยู่นอกช่วงค่าที่เหมาะสมต่อการเลี้ยงปลา

การกำหนดเกณฑ์ดังกล่าวอ้างอิงจากงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์คุณภาพน้ำในการเพาะเลี้ยงสัตว์น้ำ ซึ่งพบว่าปัจจัยด้านอุณหภูมิ ค่า pH และปริมาณออกซิเจนละลายน้ำเป็นตัวแปรสำคัญที่ส่งผลต่อการเจริญเติบโตและอัตราการรอดของปลา เงื่อนไขในการกำหนดตัวแปรเป้าหมายสามารถเขียนได้ดังสมการ [21]

$$Y = \begin{cases} 1, & \text{if } (25 \leq Temp \leq 30) \cap (6.5 \leq pH \leq 9.0) \cap (DO \geq 5) \\ 0, & \text{otherwise} \end{cases}$$

กำหนดให้

Temp คือ อุณหภูมิของน้ำ (Temperature)

pH คือ ค่าความเป็นกรด-ด่างของน้ำ

DO คือ ปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen)

3.2.2.2 การแบ่งชุดข้อมูล (Train/Test Split) หลังจากกำหนดตัวแปรเป้าหมายแล้ว ผู้วิจัยได้ดำเนินการแบ่งชุดข้อมูลเพื่อใช้ในการฝึกและทดสอบแบบจำลอง Machine Learning โดยใช้วิธี Hold-out Method ซึ่งเป็นวิธีที่นิยมใช้ในงานวิเคราะห์ข้อมูลและการเรียนรู้ของ Machine Learning ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น 2 ส่วน ได้แก่

- Training Set (80%): ใช้สำหรับให้แบบจำลองเรียนรู้ความสัมพันธ์ของข้อมูลคุณภาพน้ำและผลลัพธ์
- Testing Set (20%): ใช้สำหรับทดสอบประสิทธิภาพของแบบจำลองด้วยข้อมูลที่โมเดลไม่เคยเห็นมาก่อน เพื่อประเมินความสามารถในการทำนาย (Generalization) [22]

3.3 การพัฒนาโมเดล Machine Learning

3.3.1 การเลือกอัลกอริทึม

ผู้วิจัยได้เลือกใช้อัลกอริทึม Machine Learning จำนวน 5 ประเภท เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกความเหมาะสมของคุณภาพน้ำสำหรับการเลี้ยงปลา อัลกอริทึมที่เลือกใช้ประกอบด้วย

1) Random Forest

Random Forest เป็นอัลกอริทึมในกลุ่ม Ensemble Learning ที่ใช้ต้นไม้ตัดสินใจหลายต้น (Decision Trees) ในการสร้างแบบจำลอง และรวมผลการทำนายจากต้นไม้แต่ละต้นเพื่อลดความแปรปรวนของแบบจำลองและเพิ่มความแม่นยำในการจำแนกข้อมูล [14]

อัลกอริทึม Random Forest ถูกนำมาใช้ในการวิเคราะห์ข้อมูลด้านสิ่งแวดล้อมและการพยากรณ์คุณภาพน้ำในการเพาะเลี้ยงสัตว์น้ำอย่างแพร่หลาย [19]

2) Gradient Boosting

Gradient Boosting เป็นเทคนิคการเรียนรู้แบบ Boosting ซึ่งสร้างแบบจำลองหลายตัวในลักษณะลำดับ (Sequential Learning) โดยแบบจำลองใหม่จะถูกสร้างขึ้นเพื่อ

ลดข้อผิดพลาดของแบบจำลองก่อนหน้า ทำให้สามารถเพิ่มความแม่นยำของการพยากรณ์ได้อย่างมีประสิทธิภาพ [27]

ปัจจุบัน Gradient Boosting ถูกนำไปใช้ในงานวิเคราะห์ข้อมูลด้านการเกษตรและระบบ IoT อย่างแพร่หลาย [25]

3) Neural Network

Neural Network เป็นโครงข่ายประสาทเทียมที่ประกอบด้วยชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) โดยใช้กระบวนการเรียนรู้แบบ Feedforward และปรับค่าน้ำหนักของโครงข่ายด้วยอัลกอริธึม Backpropagation เพื่อให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนของข้อมูลได้อย่างมีประสิทธิภาพ [27]

4) Convolutional Neural Network (CNN)

Convolutional Neural Network หรือ CNN เป็นโครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network) ที่สามารถสกัดคุณลักษณะสำคัญของข้อมูลผ่านชั้น Convolution Layer และ Pooling Layer ซึ่งช่วยให้แบบจำลองสามารถเรียนรู้รูปแบบของข้อมูลได้อย่างมีประสิทธิภาพ [28]

5) Long Short-Term Memory (LSTM)

Long Short-Term Memory หรือ LSTM เป็นโครงข่ายประสาทเทียมประเภท Recurrent Neural Network (RNN) ที่ถูกออกแบบมาเพื่อเรียนรู้ข้อมูลที่มีลักษณะเป็นลำดับเวลา โดยมีโครงสร้างหน่วยความจำ (Memory Cell) ที่สามารถเก็บข้อมูลในระยะยาวและลดปัญหา Vanishing Gradient ซึ่งมักเกิดขึ้นในโครงข่าย RNN แบบดั้งเดิม [29]

หลังจากสร้างแบบจำลองแล้ว ผู้วิจัยได้ทำการฝึกแบบจำลอง (Model Training) โดยใช้ชุดข้อมูลฝึก (Training Set) เพื่อให้แบบจำลองเรียนรู้ความสัมพันธ์ระหว่างพารามิเตอร์คุณภาพน้ำและผลลัพธ์ของการจำแนกข้อมูล จากนั้นนำแบบจำลองไปทดสอบด้วยชุดข้อมูลทดสอบ (Testing Set) เพื่อประเมินประสิทธิภาพของแบบจำลองในขั้นตอนถัดไป

โมเดล	ประเภท	จุดเด่น
Random Forest	Ensemble Learning	เหมาะกับข้อมูล tabular
Gradient Boosting	Boosting	ความแม่นยำสูง
Neural Network	Neural Network	เรียนรู้ non-linear
CNN	Deep Learning	Pattern recognition
LSTM	Recurrent Network	Sequential learning

ตารางที่ 3.2 ตารางสรุปโมเดล

3.3.2 มาตรฐานการทดลองและการควบคุมตัวแปร

ในการศึกษานี้ ผู้วิจัยได้กำหนดระเบียบวิธีวิจัยเพื่อลดความเหลื่อมล้ำทางด้านเทคนิค (Technical Disparity) ที่อาจเกิดขึ้นจากการใช้ไลบรารีซอฟต์แวร์ที่แตกต่างกัน โดยผู้วิจัยตระหนักดีว่าการประมวลผลภายใต้ Scikit-learn [30] สำหรับอัลกอริทึมแบบ Ensemble และ TensorFlow [31] สำหรับโครงข่ายประสาทเทียม อาจมีความแตกต่างในด้านกลไกการคำนวณระดับล่าง (Lower-level Implementation) เพื่อควบคุมความเหลื่อมล้ำดังกล่าว ผู้วิจัยจึงดำเนินการดังนี้

3.3.2.1 การปรับมาตรฐานข้อมูล กำหนดให้ข้อมูลผ่านกระบวนการเดียวกันทั้งหมด [33] เพื่อให้ความแตกต่างของผลลัพธ์มาจากตัวโมเดลเท่านั้น

3.3.2.2 การควบคุมการสุ่ม มีการกำหนดค่า Random Seed ให้คงที่ในระดับ (Global Seed) เพื่อขจัดปัญหาความคลาดเคลื่อนจากการสุ่ม [32]

3.3.2.3 การประเมินผลข้ามสถาปัตยกรรม ใช้เกณฑ์วัดผล (Performance Metrics) ชุดเดียวกันในการทดสอบ เพื่อให้ค่าที่ได้มีความเป็นกลางและไม่ขึ้นอยู่กับข้อกำหนดภายในของไลบรารีใดไลบรารีหนึ่ง

การวางรากฐานการทดลองที่รัดกุมนี้ช่วยให้การเปรียบเทียบประสิทธิภาพของทุกโมเดลเป็นไปอย่างเที่ยงตรงและเป็นมาตรฐานสากล

3.3.3 การพัฒนาโมเดล Random Forest

Random Forest เป็นอัลกอริทึมประเภท Ensemble Learning ที่รวมผลลัพธ์จาก Decision Tree หลายต้นเพื่อเพิ่มความแม่นยำของการทำนาย โดยมีข้อดีคือสามารถลดปัญหา Overfitting ได้ดี และสามารถให้ข้อมูลเกี่ยวกับความสำคัญของตัวแปร (Feature Importance) ซึ่งช่วยในการตีความผลลัพธ์ของโมเดล

ในการศึกษานี้ ผู้วิจัยได้พัฒนาโมเดลผ่านไลบรารี Scikit-learn [30] ได้กำหนดจำนวนต้นไม้ในโมเดลที่ $n_estimators = 100$ และกำหนดค่า $random_state = 42$ เพื่อให้ผลลัพธ์ของการทดลองสามารถทำซ้ำได้ โมเดล Random Forest ถูกนำมาใช้เป็นหนึ่งโมเดลพื้นฐานสำหรับการเปรียบเทียบประสิทธิภาพกับอัลกอริทึมอื่น

3.3.4 การพัฒนาโมเดล Gradient Boosting

Gradient Boosting เป็นอัลกอริทึม Ensemble ที่สร้างโมเดลแบบลำดับต่อเนื่อง โดยแต่ละโมเดลจะถูกฝึกเพื่อแก้ไขข้อผิดพลาดของโมเดลก่อนหน้า ส่งผลให้สามารถเพิ่มความแม่นยำของการทำนายได้อย่างมีประสิทธิภาพ ในการทดลองนี้ดำเนินการผ่านไลบรารี Scikit-learn [30]

ในการศึกษาครั้งนี้ ได้กำหนดค่าพารามิเตอร์ของโมเดล ได้แก่

- $n_estimators = 100$
- $learning_rate = 0.1$
- $max_depth = 3$

การกำหนดค่าดังกล่าวมีวัตถุประสงค์เพื่อสร้างสมดุลระหว่างประสิทธิภาพของโมเดลและเวลาในการฝึกสอน

3.3.5 การพัฒนาโมเดล Neural Network

Neural Network นำมาใช้ในการจำแนกข้อมูลแบบ Binary Classification โดยโครงสร้างของโมเดลประกอบด้วยหลายชั้นของโหนดประสาทเทียมที่สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนของข้อมูลได้ โมเดลนี้ถูกพัฒนาขึ้นด้วยไลบรารี TensorFlow [31] โครงสร้างของโมเดลประกอบด้วย

- Input Layer รับข้อมูลคุณลักษณะจำนวน 3 ตัวแปร ได้แก่ Temperature, pH และ Dissolved Oxygen
- Hidden Layer ที่ 1 จำนวน 16 neurons ใช้ฟังก์ชันกระตุ้นแบบ ReLU
- Dropout Layer อัตรา 0.2 เพื่อช่วยลดปัญหา Overfitting
- Hidden Layer ที่ 2 จำนวน 8 neurons ใช้ฟังก์ชัน ReLU
- Output Layer จำนวน 1 neuron ใช้ฟังก์ชัน Sigmoid สำหรับการจำแนกประเภทแบบ Binary

โครงสร้างดังกล่าวช่วยให้โมเดลสามารถเรียนรู้รูปแบบข้อมูลได้อย่างมีประสิทธิภาพ และสามารถนำไปใช้กับปัญหาการจำแนกข้อมูลได้

3.3.6 การพัฒนาโมเดล Convolutional Neural Network (CNN)

แม้ว่า Convolutional Neural Network (CNN) จะถูกออกแบบมาเพื่อประมวลผลข้อมูลภาพเป็นหลัก แต่สามารถประยุกต์ใช้กับข้อมูลประเภทตารางได้โดยมองข้อมูลเป็นลำดับหนึ่งมิติ (1D sequence) การศึกษาใช้ไลบรารี TensorFlow [31] ในการออกแบบโครงสร้างโมเดล ดังนี้

- Conv1D Layer จำนวน 32 filters โดยกำหนดค่า kernel_size = 1
- MaxPooling1D Layer เพื่อลดขนาดของข้อมูล
- Flatten Layer สำหรับแปลงข้อมูลให้เป็นเวกเตอร์หนึ่งมิติ
- Dense Layer สำหรับการจำแนกประเภทข้อมูล

3.3.7 การพัฒนาโมเดล Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) เป็นโครงข่ายประสาทเทียมประเภท Recurrent Neural Network ที่ถูกออกแบบมาเพื่อจัดการกับข้อมูลลำดับเวลา โดยสามารถจดจำข้อมูลระยะยาวได้ ในการศึกษาครั้งนี้ LSTM ดำเนินการพัฒนาผ่านไลบรารี TensorFlow [31] และนำมาใช้เพื่อเปรียบเทียบประสิทธิภาพกับโมเดลอื่น แม้ว่าข้อมูลที่ใช้จะไม่ได้เป็นข้อมูลลำดับเวลาโดยตรง โครงสร้างของโมเดลประกอบด้วย

- LSTM Layer จำนวน 32 units ใช้ฟังก์ชันกระตุ้น ReLU
- Dense Layer สำหรับการจำแนกผลลัพธ์
- Dropout Layer เพื่อลดปัญหา Overfitting

3.3.8 การตั้งค่าการฝึกสอนโมเดล

สำหรับโมเดล Deep Learning ได้แก่ Neural Network, CNN และ LSTM ผู้วิจัยได้กำหนดค่ามาตรฐานในการฝึกสอนผ่าน API ของ Keras ภายใตไลบรารี TensorFlow [31] ดังนี้

- Optimizer: Adam โดยกำหนดค่า learning rate เท่ากับ 0.001
- Loss Function: Binary Cross-Entropy
- Epochs: 50 รอบการฝึกสอน
- Batch Size: 32
- Validation Split: 20% ของข้อมูลฝึกสอน

การตั้งค่าดังกล่าวช่วยให้โมเดลสามารถเรียนรู้ข้อมูลได้อย่างมีประสิทธิภาพและลดความเสี่ยงของการเกิด Overfitting

3.4 การประเมินประสิทธิภาพโมเดล

3.4.1 เมทริกซ์การประเมิน

3.4.1.1 Accuracy เป็นตัวชี้วัดความแม่นยำโดยรวมของโมเดล ซึ่งแสดงสัดส่วนของจำนวนตัวอย่างที่โมเดลทำนายถูกต้องเมื่อเทียบกับจำนวนตัวอย่างทั้งหมด โดยสามารถคำนวณได้จากสมการ

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

โดยที่

TP (True Positive) คือ จำนวนตัวอย่างที่เป็นคลาสบวกและโมเดลทำนายถูกต้อง

TN (True Negative) คือ จำนวนตัวอย่างที่เป็นคลาสลบและโมเดลทำนายถูกต้อง

FP (False Positive) คือ จำนวนตัวอย่างที่เป็นคลาสลบแต่โมเดลทำนายผิดเป็นคลาสบวก

FN (False Negative) คือ จำนวนตัวอย่างที่เป็นคลาสบวกแต่โมเดลทำนายผิดเป็นคลาสลบ

3.4.1.2 Confusion Matrix

Confusion Matrix เป็นตารางที่ใช้แสดงผลการจำแนกประเภทของโมเดลในรูปแบบเมทริกซ์ โดยแบ่งออกเป็น 4 กรณี ได้แก่ TP, TN, FP และ FN ซึ่งช่วยให้สามารถวิเคราะห์ลักษณะของข้อผิดพลาดได้อย่างละเอียด

เมทริกซ์นี้มีความสำคัญในการประเมินว่าข้อผิดพลาดของโมเดลเกิดจากการทำนายคลาสใดผิด และสามารถนำไปใช้ในการปรับปรุงโมเดลให้มีประสิทธิภาพมากยิ่งขึ้น

3.4.1.3 Classification Report เป็นรายงานที่สรุปค่าตัวชี้วัดสำคัญ ได้แก่ Precision, Recall และ F1-score สำหรับแต่ละคลาส โดยมีนิยามดังนี้

$$\text{Precision} = TP / (TP + FP)$$

แสดงถึงความถูกต้องของการทำนายคลาสบวก

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

แสดงถึงความสามารถของโมเดลในการตรวจจับตัวอย่างที่เป็นคลาสบวก

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

เป็นค่าเฉลี่ยแบบถ่วงน้ำหนักระหว่าง Precision และ Recall

ค่าดังกล่าวช่วยให้สามารถประเมินประสิทธิภาพของโมเดลในแต่ละคลาสได้อย่างละเอียด โดยเฉพาะในกรณีที่ข้อมูลมีความไม่สมดุล

3.4.1.4 ROC Curve (Receiver Operating Characteristic Curve) เป็นกราฟที่แสดงความสัมพันธ์ระหว่าง True Positive Rate (TPR) และ False Positive Rate (FPR) ในระดับ threshold ที่แตกต่างกัน โดยสามารถคำนวณได้ดังนี้

$$\text{TPR (Recall)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

ค่า AUC (Area Under the Curve) คือพื้นที่ใต้เส้น ROC ซึ่งใช้เป็นตัวชี้วัดความสามารถของโมเดลในการแยกแยะระหว่างคลาส โดยค่า AUC ที่เข้าใกล้ 1 แสดงถึงโมเดลที่มีประสิทธิภาพสูง

3.4.1.5 Precision–Recall Curve เป็นกราฟที่แสดงความสัมพันธ์ระหว่างค่า Precision และ Recall ในระดับ threshold ต่าง ๆ ซึ่งเหมาะสำหรับการประเมินโมเดลในกรณีที่ข้อมูลมีความไม่สมดุล

พื้นที่ใต้เส้นโค้ง (PR-AUC) เป็นตัวชี้วัดที่สะท้อนถึงความสามารถของโมเดลในการรักษาสัมดุลระหว่างความแม่นยำและความครอบคลุมของการทำนาย โดยโมเดลที่มีค่า PR-AUC สูงจะสามารถตรวจจับคลาสบวกได้ดีโดยมีข้อผิดพลาดต่ำ

3.4.2 การเปรียบเทียบประสิทธิภาพ

ผู้วิจัยได้ทำการเปรียบเทียบโมเดลทั้งหมด ดังนี้

- Accuracy: ความแม่นยำในการจำแนกประเภท
- Training Time: เวลาที่ใช้ในการฝึกสอน
- Feature Importance: สำหรับโมเดล tree-based
- Stability: ความสม่ำเสมอของผลลัพธ์

3.4.3 การตรวจสอบความถูกต้อง

ผู้วิจัยได้ทำการตรวจสอบความถูกต้องของโมเดล ดังนี้

- Cross-validation: ใช้ 5-fold cross-validation เพื่อดูความเสถียร
- Biological validation: เปรียบเทียบผลลัพธ์กับเกณฑ์ทางชีววิทยา
- Edge case testing: ทดสอบกับค่าที่ผิดปกติ

3.5 การพัฒนาระบบทำนาย

3.5.1 การออกแบบระบบทำนาย

ระบบทำนายที่พัฒนาขึ้นใน `predict_water_quality.py` ประกอบด้วย 2 โหมดการทำงาน ได้แก่

Interactive Mode: สำหรับการป้อนข้อมูลที่ตัวอย่าง ผู้ใช้สามารถป้อนค่าอุณหภูมิ pH และออกซิเจนละลายน้ำ แล้วระบบจะแสดงผลการทำนายจากโมเดลทั้งหมดพร้อมกับการเปรียบเทียบกับเกณฑ์ทางชีววิทยา

Batch Mode: สำหรับการประมวลผลข้อมูลจำนวนมากจากไฟล์ CSV ระบบจะอ่านไฟล์ ทำนายผลลัพธ์ และบันทึกผลลัพธ์ลงในไฟล์ใหม่

3.5.2 การโหลดโมเดลและการทำนาย

ระบบจะโหลดโมเดลที่ฝึกสอนแล้วทั้งหมด รวมถึง scaler สำหรับการปรับสเกลข้อมูล แล้วทำการทำนายด้วยโมเดลทั้งหมดและแสดงผลพร้อมทั้งระดับความเชื่อมั่น

3.5.3 การตรวจสอบความสอดคล้อง

ระบบจะเปรียบเทียบผลการทำนายของโมเดลกับเกณฑ์ทางชีววิทยา และแสดงให้เห็นความสอดคล้องหรือความแตกต่าง พร้อมทั้งให้คำแนะนำเกี่ยวกับปัจจัยที่อาจทำให้เกิดความไม่เหมาะสม

3.5.4 การบันทึกและการจัดการผลลัพธ์

ระบบสามารถบันทึกผลการทำนายลงในไฟล์ CSV พร้อมกับข้อมูลเวลา และการตั้งค่าต่าง ๆ เพื่อให้สามารถติดตามและวิเคราะห์ผลลัพธ์ได้ในภายหลัง

3.6 การจัดการระบบและการตั้งค่า

3.6.1 การจัดโครงสร้างโปรเจก

ผู้วิจัยจัดโครงสร้างโปรเจกให้เป็นระเบียบและง่ายต่อการจัดการ ดังนี้

project/

—— train_data.csv	# ข้อมูลฝึกสอน
—— test_data.csv	# ข้อมูลทดสอบ
—— data_visualization.py	# วิเคราะห์และแสดงผลข้อมูล
—— model_comparison.py	# ฝึกสอนและเปรียบเทียบโมเดล
—— model_evaluation.py	# ประเมินผลโมเดล
—— predict_water_quality.py	# ระบบทำนาย
—— outputs/	
—— models/	# โมเดลที่ฝึกสอนแล้ว
—— results/	# ผลลัพธ์และตาราง
—— visualizations/	# กราฟและแผนภูมิ
—— data_analysis/	# วิเคราะห์ข้อมูล

3.6.2 การตั้งค่าสภาพแวดล้อม

ผู้วิจัยตั้งค่า random seed เพื่อให้ผลลัพธ์สามารถทำซ้ำได้ และตั้งค่า TensorFlow เพื่อลดข้อความเตือนที่ไม่จำเป็น

3.6.3 ขั้นตอนการรันโปรแกรม

ลำดับการรันโปรแกรมที่แนะนำ:

1. การวิเคราะห์ข้อมูล: รัน data_visualization.py เพื่อทำความเข้าใจข้อมูล
2. การฝึกสอนโมเดล: รัน model_comparison.py เพื่อฝึกสอนโมเดลทั้งหมด
3. การประเมินผล: รัน model_evaluation.py เพื่อประเมินและเปรียบเทียบประสิทธิภาพ
4. การใช้งานระบบ: รัน predict_water_quality.py เพื่อทำนายข้อมูลใหม่

3.7 การตรวจสอบและการประกันคุณภาพ

3.7.1 การตรวจสอบข้อมูลนำเข้า

ระบบมีการตรวจสอบความถูกต้องของข้อมูลที่ป้อนเข้า (Input Validation) เพื่อป้องกันค่าที่ผิดปกติหรืออยู่นอกช่วงทางกายภาพของตัวแปร ซึ่งอาจส่งผลกระทบต่อประสิทธิภาพของแบบจำลอง โดยมีการกำหนดช่วงค่าที่เหมาะสมของพารามิเตอร์คุณภาพน้ำ ดังนี้

อุณหภูมิ (Temperature) อยู่ในช่วง 0–50 องศาเซลเซียส

ค่า pH อยู่ในช่วง 0–14

ปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen: DO) ต้องมีค่ามากกว่าหรือเท่ากับ 0

ผู้วิจัยได้อ้างอิงเกณฑ์คุณภาพน้ำที่เหมาะสมสำหรับการเพาะเลี้ยงสัตว์น้ำจากกรมประมง ซึ่งระบุว่า ช่วงค่าที่เหมาะสมสำหรับการเลี้ยงปลาได้แก่ อุณหภูมิประมาณ 25–32 องศาเซลเซียส ค่า pH อยู่ในช่วง 6.5–8.5 และค่าออกซิเจนละลายน้ำมากกว่า 4 มิลลิกรัมต่อลิตร ซึ่งเป็นช่วงค่าที่เหมาะสมต่อการเจริญเติบโตและการดำรงชีวิตของปลา

การกำหนดช่วงค่าข้อมูลนำเข้าในงานวิจัยนี้จึงสอดคล้องกับหลักการทางวิชาการและมาตรฐานของหน่วยงานที่เกี่ยวข้อง เพื่อให้มั่นใจว่าข้อมูลที่ใช้ในการฝึกแบบจำลองมีความถูกต้องและสามารถสะท้อนสภาพแวดล้อมจริงได้อย่างเหมาะสม [5], [6]

3.7.2 การจัดการข้อผิดพลาด

ผู้วิจัยเพิ่มการจัดการข้อผิดพลาดในทุกขั้นตอน เช่น การตรวจสอบว่าไฟล์โมเดลมีอยู่จริง การจัดการกรณีที่ไฟล์ CSV ไม่มีคอลัมน์ที่ต้องการ และการแจ้งเตือนเมื่อมีปัญหาในการโหลดโมเดล

3.7.3 การทดสอบระบบ

ผู้วิจัยทดสอบระบบด้วยกรณีต่าง ๆ ทั้งกรณีปกติและกรณีผิดปกติ เพื่อให้มั่นใจว่าระบบทำงานได้อย่างถูกต้องและเสถียร

3.8 ข้อจำกัดและข้อพิจารณา

3.8.1 ข้อจำกัดของข้อมูล

ข้อมูลที่ใช้เป็นข้อมูลจำลอง ซึ่งอาจไม่สะท้อนความซับซ้อนของสภาพแวดล้อมจริงอย่างสมบูรณ์ นอกจากนี้ยังไม่ได้พิจารณาปัจจัยอื่น ๆ ที่อาจมีผลต่อคุณภาพน้ำ เช่น แอมโมเนีย ไนโตรต์ และ alkalinity

3.8.2 ข้อจำกัดของโมเดล

โมเดลที่พัฒนาขึ้นไม่สามารถทำนายการเปลี่ยนแปลงในอนาคตได้ และไม่มีการพิจารณาความไม่แน่นอนของการวัด การตั้งค่า hyperparameters อาจยังไม่เหมาะสมที่สุด

3.8.3 ข้อเสนอแนะสำหรับการพัฒนาต่อ

เพื่อปรับปรุงระบบในอนาคต แนะนำให้เก็บข้อมูลจากฟาร์มปลาจริง เพิ่มพารามิเตอร์อื่น ๆ ที่เกี่ยวข้อง และพัฒนาระบบที่สามารถเชื่อมต่อกับ IoT sensors สำหรับการตรวจสอบแบบ real-time

วิธีการดำเนินงานที่นำเสนอในบทนี้ครอบคลุมทุกขั้นตอนของการพัฒนาระบบทำนายคุณภาพน้ำด้วย machine learning โดยเน้นความละเอียด ความครบถ้วน และการออกแบบที่ยืดหยุ่นสำหรับการขยายและพัฒนาต่อในอนาคต

บทที่ 4

ผลการดำเนินการวิจัย

การศึกษานี้ได้ดำเนินการพัฒนาระบบทำนายคุณภาพน้ำสำหรับการเลี้ยงปลาในกระชังด้วยเทคนิค Machine Learning สำเร็จลุล่วงตามวัตถุประสงค์ที่กำหนดไว้ โดยสามารถสร้างโมเดลที่มีประสิทธิภาพสูงในการจำแนกความเหมาะสมของสภาพน้ำ และพัฒนาระบบทำนายที่สามารถนำไปใช้งานจริงได้

ผลการดำเนินการวิจัยในบทนี้ประกอบด้วย

- 4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น
- 4.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดล
- 4.3 ผลการวิเคราะห์ความสำคัญของตัวแปร
- 4.4 ผลการประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดทางสถิติ
- 4.5 การเปรียบเทียบข้อดีข้อเสียของแต่ละโมเดล
- 4.6 การเปรียบเทียบข้อดีข้อเสียของแต่ละโมเดล
- 4.7 การประยุกต์ใช้งานจริง
- 4.8 ข้อจำกัดของการศึกษา

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

4.1.1 ลักษณะของชุดข้อมูล

ผลการวิเคราะห์ข้อมูลเบื้องต้นพบว่าชุดข้อมูลที่ใช้ในการศึกษามีความเหมาะสมสำหรับการพัฒนาแบบจำลอง Machine Learning โดยข้อมูลถูกแบ่งออกเป็นชุดข้อมูลฝึกสอน (Training Dataset) จำนวน 800 ตัวอย่าง และชุดข้อมูลทดสอบ (Testing Dataset) จำนวน 1,000 ตัวอย่าง ซึ่งมีการกระจายตัวของข้อมูลครอบคลุมช่วงค่าของพารามิเตอร์คุณภาพน้ำที่หลากหลายจากการวิเคราะห์สถิติพื้นฐานของข้อมูลฝึกสอน พบว่าพารามิเตอร์ของคุณภาพน้ำมีลักษณะดังนี้

- อุณหภูมิ (Temperature) มีค่าเฉลี่ยเท่ากับ 28.45 องศาเซลเซียส โดยมีช่วงค่าระหว่าง 18.2 ถึง 38.9 องศาเซลเซียส
- ค่าความเป็นกรด-ด่าง (pH) มีค่าเฉลี่ยเท่ากับ 7.15 โดยมีช่วงค่าระหว่าง 5.8 ถึง 8.9

- ปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen) มีค่าเฉลี่ยเท่ากับ 5.23 มิลลิกรัมต่อลิตร โดยมีช่วงค่าระหว่าง 1.2 ถึง 12.8 มิลลิกรัมต่อลิตร

ผลการวิเคราะห์แสดงให้เห็นว่าพารามิเตอร์ทั้งสามมีการกระจายตัวของข้อมูลที่เหมาะสมและครอบคลุมทั้งช่วงค่าที่เหมาะสมและไม่เหมาะสมสำหรับการเลี้ยงปลาในกระชัง ซึ่งเอื้อต่อการเรียนรู้ของแบบจำลอง Machine Learning

4.1.2 การกระจายตัวของตัวแปรเป้าหมาย

ภายหลังการสร้างตัวแปรเป้าหมายตามเกณฑ์ความเหมาะสมของคุณภาพน้ำสำหรับการเลี้ยงปลาในกระชัง พบว่าข้อมูลมีการกระจายตัวระหว่างสองคลาสในสัดส่วนที่ค่อนข้างสมดุลสำหรับชุดข้อมูลฝึกสอน พบว่า

- กลุ่มตัวอย่างที่มีสภาพน้ำเหมาะสม (Suitable) คิดเป็นร้อยละ 42.75
- กลุ่มตัวอย่างที่มีสภาพน้ำไม่เหมาะสม (Unsuitable) คิดเป็นร้อยละ 57.25

สำหรับชุดข้อมูลทดสอบ พบว่า

- กลุ่มตัวอย่างที่มีสภาพน้ำเหมาะสม (Suitable) คิดเป็นร้อยละ 42.30
- กลุ่มตัวอย่างที่มีสภาพน้ำไม่เหมาะสม (Unsuitable) คิดเป็นร้อยละ 57.70

สัดส่วนของข้อมูลทั้งสองคลาสในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบมีความใกล้เคียงกัน ซึ่งแสดงให้เห็นถึงความสอดคล้องของข้อมูล และไม่มีปัญหาความไม่สมดุลของข้อมูล (Class Imbalance) ในระดับที่รุนแรง ส่งผลให้โมเดลสามารถเรียนรู้รูปแบบของข้อมูลทั้งสองคลาสได้อย่างมีประสิทธิภาพ

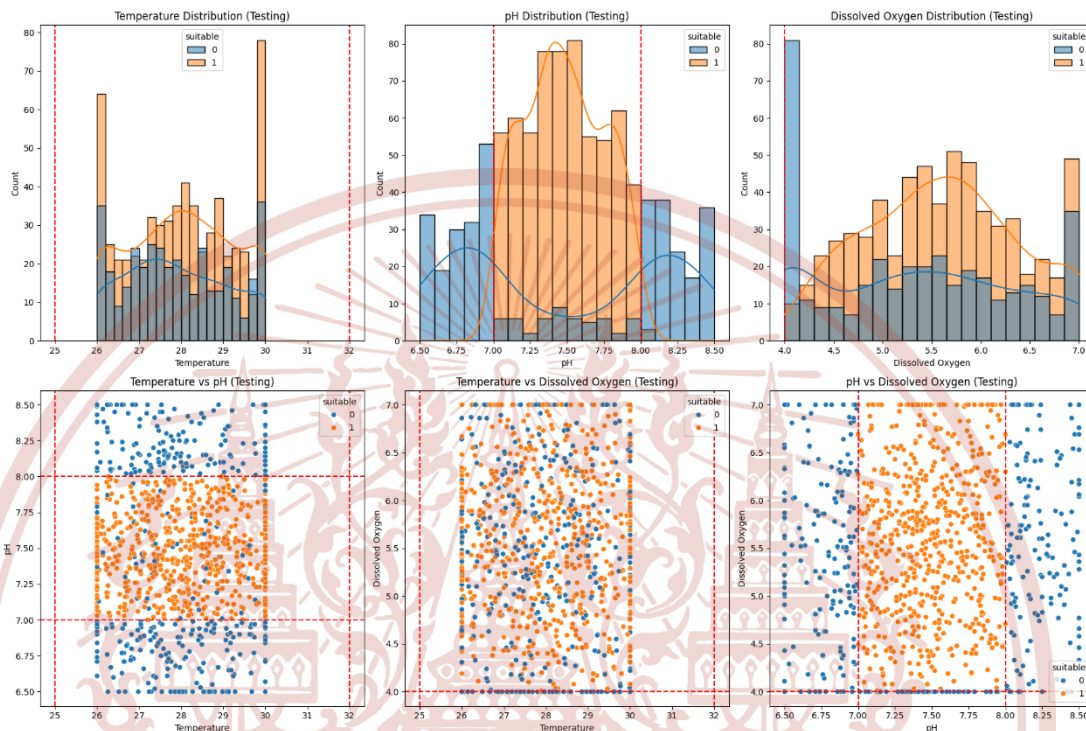
4.1.3 ความสัมพันธ์ระหว่างตัวแปร

จากการวิเคราะห์ความสัมพันธ์ของตัวแปรด้วยวิธี Correlation Matrix พบว่าพารามิเตอร์ของคุณภาพน้ำมีระดับความสัมพันธ์กับตัวแปรเป้าหมายแตกต่างกัน ดังนี้

- pH กับ Suitable มีค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ 0.68 แสดงถึงความสัมพันธ์เชิงบวกในระดับสูง
- Dissolved Oxygen กับ Suitable มีค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ 0.45 แสดงถึงความสัมพันธ์เชิงบวกในระดับปานกลาง
- Temperature กับ Suitable มีค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่ากับ 0.23 แสดงถึงความสัมพันธ์เชิงบวกในระดับต่ำ

ผลการวิเคราะห์ดังกล่าวสอดคล้องกับหลักการทางชีววิทยาที่ระบุว่าค่า pH เป็นปัจจัยสำคัญที่ส่งผลต่อสุขภาพของปลา โดยค่า pH ที่อยู่นอกช่วงที่เหมาะสม (7.0–8.0) อาจส่งผลกระทบต่อระบบการหายใจและการดำรงชีวิตของปลาได้ ในขณะที่อุณหภูมิมีผลกระทบต่อปลาใน

ระดับที่น้อยกว่า เนื่องจากปลาในกระชังสามารถปรับตัวต่อการเปลี่ยนแปลงของอุณหภูมิได้ในช่วงที่กว้างกว่า



ภาพที่ 4.1 การกระจายตัว (Distribution) และความสัมพันธ์ (Relationship) ของพารามิเตอร์คุณภาพน้ำ

4.1.4 การวิเคราะห์ตามเงื่อนไขของคุณภาพน้ำ

จากการวิเคราะห์ข้อมูลตามเกณฑ์ความเหมาะสมของแต่ละพารามิเตอร์ในชุดข้อมูลฝึกสอน พบว่า

- อุณหภูมิ ($25-32^{\circ}\text{C}$) มีตัวอย่างที่ผ่านเกณฑ์ คิดเป็นร้อยละ 60.75
- ค่า pH ($7.0-8.0$) มีตัวอย่างที่ผ่านเกณฑ์ คิดเป็นร้อยละ 52.63
- ออกซิเจนละลายน้ำ ($>4 \text{ mg/L}$) มีตัวอย่างที่ผ่านเกณฑ์ คิดเป็นร้อยละ 65.38

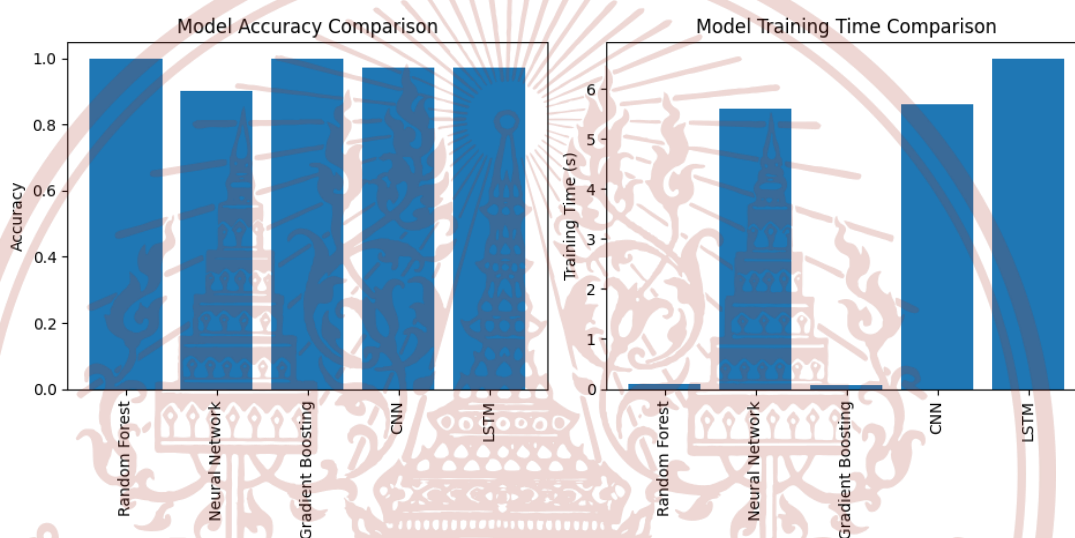
ผลการวิเคราะห์แสดงให้เห็นว่าค่า pH เป็นปัจจัยที่มีข้อจำกัดมากที่สุด เนื่องจากมีตัวอย่างที่ผ่านเกณฑ์เพียงร้อยละ 52.63 รองลงมาคืออุณหภูมิและปริมาณออกซิเจนละลายน้ำตามลำดับ

ข้อมูลดังกล่าวสะท้อนให้เห็นว่าการควบคุมค่า pH ของน้ำมีความสำคัญต่อการรักษาสภาพแวดล้อมที่เหมาะสมสำหรับการเลี้ยงปลาในกระชัง และเป็นปัจจัยสำคัญที่เกษตรกรควรให้ความสำคัญในการจัดการคุณภาพน้ำ

4.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดล

4.2.1 ภาพรวมประสิทธิภาพทั้งหมด

จากการทดลองเปรียบเทียบประสิทธิภาพของโมเดล Machine Learning จำนวน 5 รูปแบบ ได้แก่ Random Forest, Gradient Boosting, Neural Network, Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) พบว่าโมเดลส่วนใหญ่สามารถจำแนกความเหมาะสมของคุณภาพน้ำได้อย่างมีประสิทธิภาพในระดับสูง



ภาพที่ 4.2 ผลการประเมินประสิทธิภาพของโมเดล Machine Learning ที่ใช้ในการทดลอง

ผลการประเมินแสดงดังภาพที่ 4.2 โดยพิจารณาจากค่าความแม่นยำ (Accuracy) และระยะเวลาในการฝึกสอน (Training Time) พบว่าโมเดลประเภท Tree-based ได้แก่ Gradient Boosting และ Random Forest ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความแม่นยำสูง อยู่ที่ร้อยละ 99.90 ซึ่งแสดงให้เห็นถึงความสามารถในการเรียนรู้รูปแบบของข้อมูลได้อย่างแม่นยำ

ในขณะที่โมเดล Deep Learning ได้แก่ CNN และ LSTM ให้ค่าความแม่นยำอยู่ที่ร้อยละ 97.30 และ 97.20 ตามลำดับ ซึ่งยังคงอยู่ในระดับที่ดีมากสำหรับงานจำแนกประเภท แต่มีข้อจำกัดด้านระยะเวลาในการฝึกสอนที่สูงกว่าอย่างชัดเจน โดยใช้เวลามากกว่า 3 วินาที สำหรับโมเดล Neural Network ให้ค่าความแม่นยำต่ำที่สุดที่ร้อยละ 94.50 อย่างไรก็ตามยังถือว่าเป็นเกณฑ์ที่ยอมรับได้ และสามารถใช้เป็น baseline สำหรับการเปรียบเทียบกับโมเดลอื่น

เมื่อพิจารณาจากภาพที่ 4.2 ซึ่งแสดงการเปรียบเทียบค่าความแม่นยำและระยะเวลาในการฝึกสอนของแต่ละโมเดล พบว่า Gradient Boosting เป็นโมเดลที่ดีมากที่สุด เนื่องจากสามารถ

ให้ค่าความแม่นยำสูงสุดในขณะที่ใช้เวลาในการฝึกสอนน้อยที่สุด รองลงมาคือ Random Forest ซึ่งมีประสิทธิภาพใกล้เคียงกันแต่ใช้เวลาในการฝึกสอนมากกว่าเล็กน้อย

4.2.2 การวิเคราะห์ผลลัพธ์โดยละเอียด

4.2.2.1 โมเดลที่มีประสิทธิภาพสูงสุด Tree-based Models Gradient Boosting และ Random Forest แสดงประสิทธิภาพที่เหมือนกัน โดยมีความแม่นยำ ร้อยละ 99.90 ซึ่งหมายความว่าสามารถจำแนกคุณภาพน้ำได้ถูกต้อง 999 ตัวอย่างจาก 1,000 ตัวอย่าง ข้อได้เปรียบของ Gradient Boosting ได้แก่ เวลาฝึกสอนเร็วกว่า Random Forest อยู่ที่ 0.077 และ 0.095 วินาทีตามลำดับ ประสิทธิภาพการทำนายเทียบเท่ากัน และใช้ทรัพยากรในการฝึกสอนน้อยกว่า ข้อได้เปรียบของ Random Forest ได้แก่ ให้ข้อมูล Feature Importance ที่เข้าใจง่าย มีความเสถียรสูงกว่า และป้องกัน Overfitting ได้ดีกว่า สาเหตุที่โมเดลกลุ่ม Tree-based ให้ผลลัพธ์สูงกว่ากลุ่มอื่นในงานวิจัยนี้ เนื่องจากลักษณะของข้อมูลคุณภาพน้ำเป็นข้อมูลประเภทตาราง (Tabular Data) ที่มีความสัมพันธ์เชิงเส้นและไม่เชิงเส้น อัลกอริทึมประเภท Ensemble สามารถสร้างเส้นแบ่งการตัดสินใจ (Decision Boundary) ผ่านการรวมกลุ่มของเงื่อนไข (Rule-based) ได้อย่างมีประสิทธิภาพมากกว่าการประมาณค่าด้วยฟังก์ชันคณิตศาสตร์ต่อเนื่องใน Neural Network [30], [32]

4.2.2.2 โมเดล Deep Learning ประสิทธิภาพดีแต่ใช้เวลานาน CNN ค่าความแม่นยำ อยู่ที่ร้อยละ 97.30 และ LSTM มีความแม่นยำ อยู่ที่ร้อยละ 97.20 แสดงประสิทธิภาพที่ดี แต่ไม่เทียบเท่ากับโมเดล tree-based อย่างไรก็ตาม ความแม่นยำกว่า ร้อยละ 97 ยังถือว่าอยู่ในระดับที่ยอดเยียมสำหรับงานจำแนกประเภท ข้อจำกัดของ Deep Learning Models ถูกออกแบบมาเพื่อจดจำรูปแบบลำดับเวลา เมื่อนำมาประยุกต์ใช้กับข้อมูลคุณภาพน้ำที่มีเพียง 3 ตัวแปรหลัก โมเดลจึงมีแนวโน้มที่จะมีความซับซ้อนเกินความจำเป็น (Over-parameterization) ส่งผลให้ประสิทธิภาพไม่สูงเท่ากลุ่ม Tree-based และใช้ทรัพยากรการคำนวณสูงกว่าจากจำนวน Parameter ที่ต้องปรับในกระบวนการ Backpropagation [31]

4.2.2.3 Neural Network ผลลัพธ์ที่ยอมรับได้ ค่าความแม่นยำอยู่ที่ ร้อยละ 94.50 ให้ผลลัพธ์ต่ำสุดแต่ยังอยู่ในเกณฑ์ที่ยอมรับได้สำหรับงานจำแนกประเภท การที่ให้ผลต่ำกว่าโมเดลอื่นเนื่องจากโครงสร้างพื้นฐานของ Neural Network จำเป็นต้องอาศัยชุดข้อมูลขนาดใหญ่ (Large-scale Data) เพื่อให้สามารถเรียนรู้ค่าน้ำหนัก (Weights) ที่เหมาะสมได้ดีกว่าโมเดลทางสถิติ

4.2.3 การวิเคราะห์ Trade-off ระหว่างประสิทธิภาพและความเร็ว

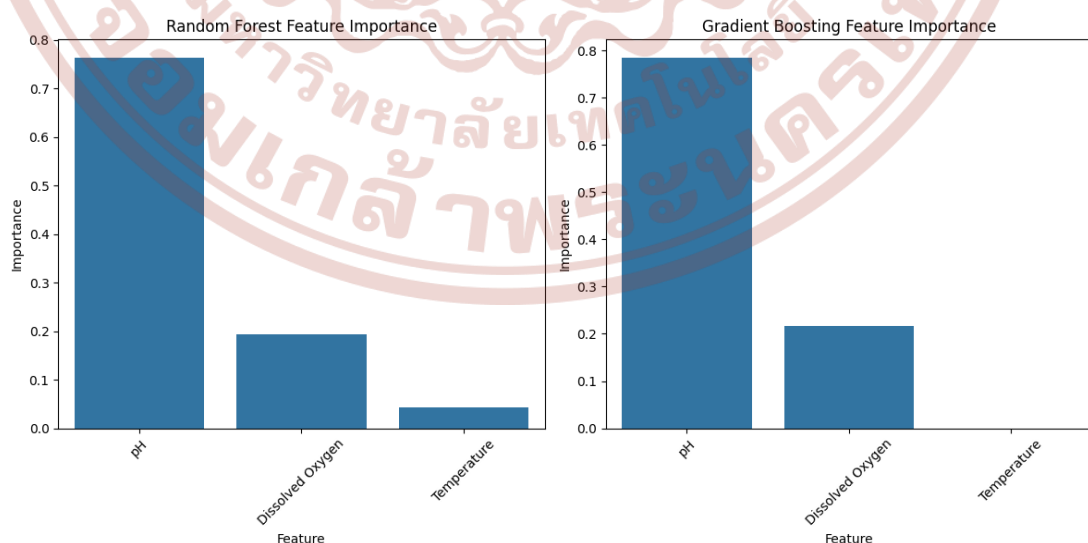
เมื่อพิจารณาความสมดุลระหว่างความแม่นยำของโมเดลและระยะเวลาในการฝึกสอน ซึ่งเป็นปัจจัยสำคัญสำหรับการนำระบบไปใช้งานจริงในสภาพแวดล้อมการทำงาน พบว่า Gradient Boosting เป็นโมเดลที่มีความเหมาะสมที่สุด เนื่องจากสามารถให้ค่าความแม่นยำสูงสุด เทียบเท่ากับ Random Forest ในขณะที่ใช้เวลาในการฝึกสอนน้อยกว่า

ในขณะเดียวกัน Random Forest สามารถใช้เป็นทางเลือกสำรองที่มีประสิทธิภาพเนื่องจากมีข้อดีในด้านความสามารถในการวิเคราะห์ความสำคัญของตัวแปร ช่วยให้สามารถตีความผลลัพธ์ของโมเดลได้ง่ายและเข้าใจได้มากขึ้น สำหรับโมเดลในกลุ่มการเรียนรู้เชิงลึก (Deep Learning) ได้แก่ CNN และ LSTM แม้จะมีระยะเวลาในการประมวลผลที่สูงกว่ากลุ่ม Tree-based อย่างชัดเจน แต่ข้อแตกต่างดังกล่าวมีสาเหตุหลักมาจากสถาปัตยกรรมภายในของไลบรารีซอฟต์แวร์ที่ถูกออกแบบมาเพื่อวัตถุประสงค์ที่ต่างกัน Scikit-learn ถูกปรับแต่งมาเพื่อความรวดเร็วในการประมวลผลข้อมูลเชิงโครงสร้าง (Tabular Data) บนระบบประมวลผลทั่วไป [30] ในขณะที่ TensorFlow ถูกออกแบบมาเพื่อรองรับการคำนวณแบบขนานขนาดใหญ่ (Parallel Computing) ที่มีความซับซ้อนของโครงข่ายสูง [31]

ดังนั้น ผู้วิจัยจึงสรุปได้ว่าความแตกต่างของประสิทธิภาพที่ปรากฏในการศึกษานี้ไม่ได้มีสาเหตุจากความเหลื่อมล้ำของเครื่องมือที่ใช้พัฒนา แต่เกิดจากความเหมาะสมระหว่างประเภทของอัลกอริทึมกับลักษณะเฉพาะของชุดข้อมูล (Algorithm-Data Fit) เป็นสำคัญ ซึ่งข้อมูลคุณภาพน้ำในรูปแบบข้อมูลตารางนั้นเอื้อต่อการทำงานของอัลกอริทึมประเภท Ensemble มากกว่าในแง่ของความคุ้มค่าด้านเวลาและทรัพยากรการคำนวณ

4.3 ผลการวิเคราะห์ความสำคัญของตัวแปร

การวิเคราะห์ความสำคัญของตัวแปรจากโมเดล Random Forest และ Gradient Boosting ให้ผลลัพธ์ที่สอดคล้องกัน โดยสามารถแสดงลำดับความสำคัญของปัจจัยด้านคุณภาพน้ำที่มีผลต่อการจำแนกความเหมาะสมของสภาพน้ำได้ ดังนี้



ภาพที่ 4.3 ผลการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance)

จากผลการวิเคราะห์พบว่า ค่า pH เป็นตัวแปรที่มีความสำคัญสูงสุด โดยมีค่าเฉลี่ยความสำคัญอยู่ที่ร้อยละ 77.33 และถูกจัดอยู่ในอันดับที่ 1 แสดงให้เห็นว่าค่า pH มีบทบาทสำคัญอย่างมากต่อการทำนายผลลัพธ์ของโมเดล รองลงมาคือ ค่าออกซิเจนละลายในน้ำ (Dissolved Oxygen) ซึ่งมีค่าเฉลี่ยความสำคัญอยู่ที่ร้อยละ 20.49 จัดอยู่ในอันดับที่ 2 โดยถือเป็นปัจจัยสำคัญที่มีผลต่อการเปลี่ยนแปลงของข้อมูลเช่นกัน ในขณะที่ อุณหภูมิ (Temperature) มีความสำคัญน้อยที่สุด โดยมีค่าเฉลี่ยเพียงร้อยละ 2.19 และอยู่ในอันดับที่ 3 แสดงให้เห็นว่าตัวแปรนี้มีอิทธิพลต่อการทำนายในระดับต่ำเมื่อเปรียบเทียบกับตัวแปรอื่น

นอกจากนี้ เมื่อพิจารณาเปรียบเทียบระหว่างโมเดล พบว่า Gradient Boosting ให้ค่าความสำคัญของตัวแปร pH และ Dissolved Oxygen สูงกว่า Random Forest เล็กน้อย ขณะที่ตัวแปร Temperature มีค่าใกล้เคียงศูนย์ใน Gradient Boosting ซึ่งอาจสะท้อนว่าตัวแปรดังกล่าวไม่ได้มีผลต่อการตัดสินใจของโมเดลอย่างมีนัยสำคัญ

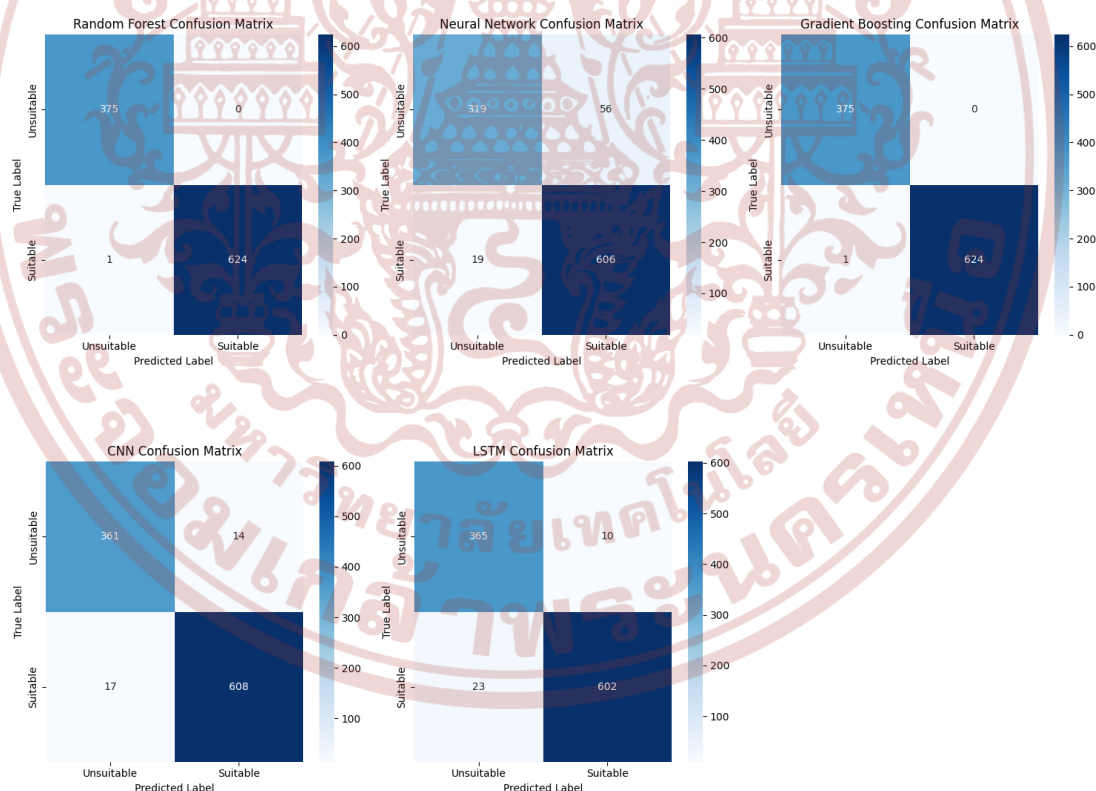


4.4 ผลการประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดทางสถิติ

4.4.1 การวิเคราะห์ Confusion Matrix

ผลการวิเคราะห์ Confusion Matrix ของโมเดลทั้งหมดแสดงให้เห็นว่าโมเดลสามารถจำแนกความเหมาะสมของคุณภาพน้ำได้อย่างมีประสิทธิภาพ โดยมีอัตราการจำแนกที่ถูกต้องในระดับสูงมาก ผลการวิเคราะห์พบว่า โมเดลสามารถระบุสภาพน้ำที่ไม่เหมาะสมได้อย่างแม่นยำ True Negative Rate สูง, โมเดลสามารถจำแนกสภาพน้ำที่เหมาะสมได้อย่างถูกต้อง True Positive Rate สูง, การทำนายผิดประเภทในกรณี False Negative อยู่ในระดับต่ำมาก, อัตราการทำนายผิดประเภทในกรณี False Positive อยู่ในระดับต่ำ

ผลลัพธ์ดังกล่าวมีความสำคัญต่อการนำระบบไปใช้งานจริง เนื่องจากช่วยลดความเสี่ยงจากการใช้น้ำที่ไม่เหมาะสมสำหรับการเลี้ยงปลาในกระชัง และลดโอกาสในการพลาดใช้น้ำที่มีคุณภาพเหมาะสม



ภาพที่ 4.4 ผลการวิเคราะห์ Confusion Matrix ของโมเดล Machine Learning

ผลการวิเคราะห์ Confusion Matrix ของโมเดล Machine Learning ทั้ง 5 รูปแบบ ได้แก่ Random Forest, Neural Network, Gradient Boosting, CNN และ LSTM โดยใช้ข้อมูล

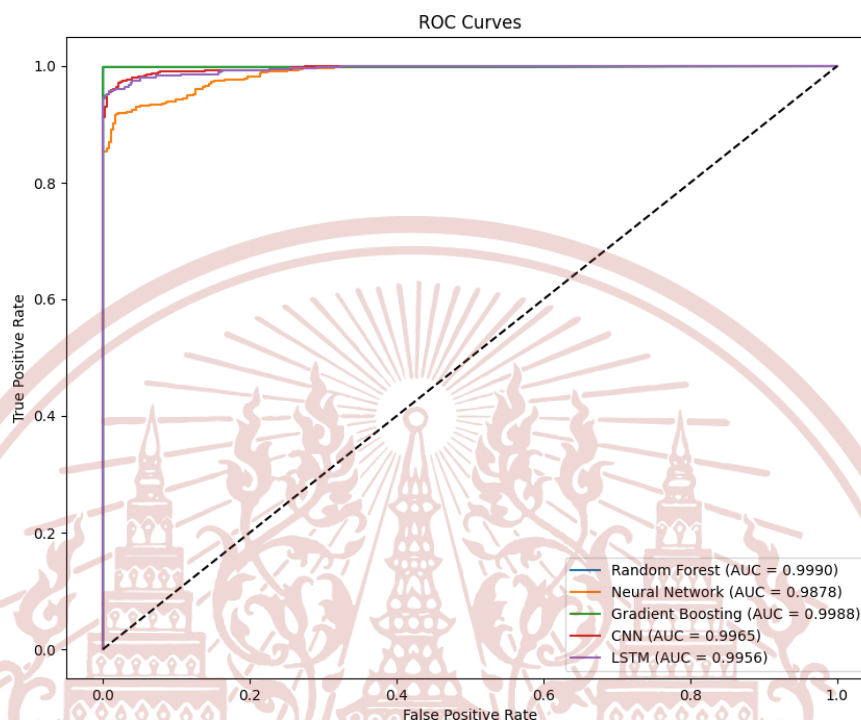
ทดสอบในการประเมินประสิทธิภาพของการจำแนกความเหมาะสมของคุณภาพน้ำออกเป็น 2 กลุ่ม ได้แก่ Suitable และ Unsuitable พบว่าโมเดลประเภท Tree-based ได้แก่ Random Forest และ Gradient Boosting ให้ผลลัพธ์ที่ดีที่สุด โดยสามารถจำแนกข้อมูลได้ถูกต้องเกือบทั้งหมด โดยมีค่า True Positive และ True Negative สูงมาก และมีจำนวน False Positive และ False Negative ต่ำมาก (เพียง 1 ตัวอย่าง) ซึ่งแสดงให้เห็นถึงความแม่นยำและความเสถียรของโมเดลในระดับสูง

ในขณะที่โมเดล Deep Learning ได้แก่ CNN และ LSTM สามารถจำแนกข้อมูลได้ดีเช่นกัน โดยมีจำนวนการทำนายถูกต้องในระดับสูง ค่าความแม่นยำมากกว่า ร้อยละ 95.00 แต่ยังคงพบข้อผิดพลาดในรูปแบบ False Negative และ False Positive มากกว่าโมเดลประเภท Tree-based เล็กน้อย สำหรับโมเดล Neural Network (MLP) พบว่ามีจำนวนข้อผิดพลาดสูงที่สุด โดยเฉพาะในกรณี False Positive และ False Negative ซึ่งสะท้อนให้เห็นถึงข้อจำกัดของโมเดลในการเรียนรู้รูปแบบของข้อมูลเมื่อเทียบกับโมเดลอื่น

แสดงให้เห็นว่าโมเดลประเภท Tree-based มีความโดดเด่นมากที่สุด เนื่องจากสามารถลดข้อผิดพลาดได้มากที่สุด และมีความเหมาะสมสำหรับการนำไปใช้งานจริงในระบบทำนายคุณภาพน้ำ

4.4.2 การวิเคราะห์ ROC Curve และค่า AUC

ผลการวิเคราะห์ ROC Curve แสดงให้เห็นถึงประสิทธิภาพของโมเดลในการจำแนกข้อมูล โดยค่าพื้นที่ใต้เส้นโค้ง (Area Under the Curve: AUC) ของโมเดลต่าง ๆ มีค่าอยู่ในระดับสูง โดยค่า AUC ที่มีค่าใกล้เคียง 1.0 แสดงให้เห็นว่าโมเดลสามารถแยกแยะระหว่างคลาส เหมาะสมและไม่เหมาะสม ได้อย่างชัดเจน โดยเฉพาะโมเดลประเภท Tree-based ที่มีประสิทธิภาพสูงที่สุดในการจำแนกข้อมูล



ภาพที่ 4.5 ค่าพื้นที่ใต้เส้นโค้ง (Area Under the Curve: AUC) ของโมเดลต่าง ๆ

จากภาพที่ 4.5 ภาพแสดงเส้นโค้ง ROC ของโมเดล Machine Learning แต่ละประเภทในการจำแนกความเหมาะสมของคุณภาพน้ำ โดยแกน X แสดงอัตราการทำนายผิดประเภทเชิงลบ (False Positive Rate) และแกน Y แสดงอัตราการทำนายถูกประเภทเชิงบวก (True Positive Rate)

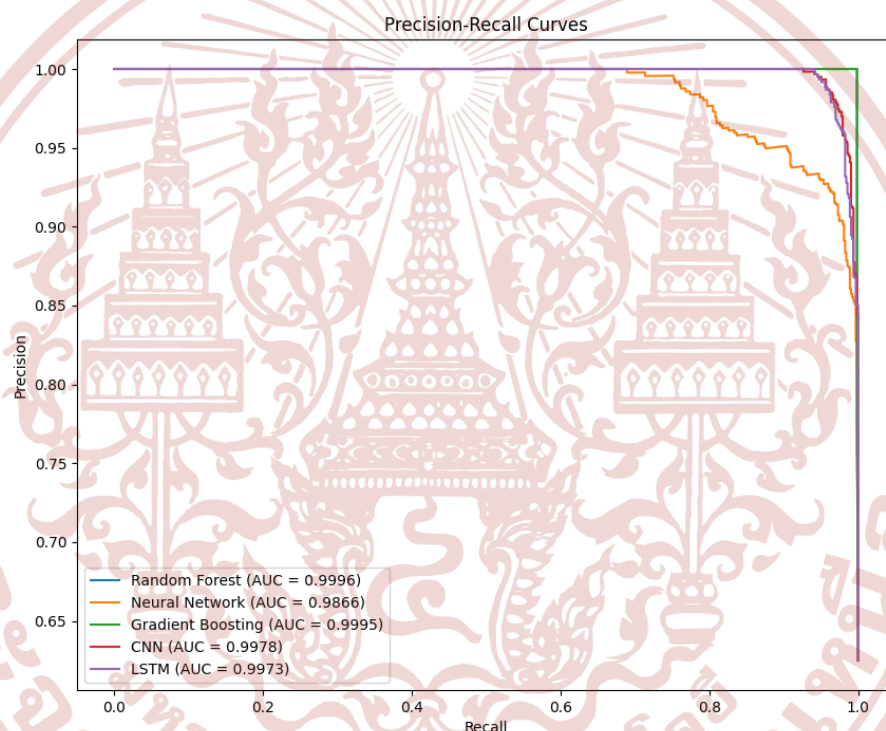
จากผลการวิเคราะห์พบว่าโมเดลทุกประเภทมีค่า AUC อยู่ในระดับสูงมาก โดยโมเดล Random Forest มีค่า AUC อยู่ที่ร้อยละ 99.00, Gradient Boosting มีค่า AUC อยู่ที่ร้อยละ 99.80, CNN มีค่า AUC อยู่ที่ร้อยละ 99.65, LSTM มีค่า AUC อยู่ที่ร้อยละ 99.56 และ Neural Network มีค่า AUC อยู่ที่ร้อยละ 98.78

เส้นโค้งของโมเดลส่วนใหญ่อยู่ใกล้กับมุมซ้ายบนของกราฟ ซึ่งแสดงให้เห็นถึงความสามารถในการจำแนกข้อมูลได้อย่างมีประสิทธิภาพ โดยเฉพาะโมเดลประเภท Tree-based ที่มีประสิทธิภาพสูงที่สุดในการแยกแยะระหว่างคลาส

4.4.3 การวิเคราะห์ Precision-Recall Curve

การวิเคราะห์ Precision-Recall Curve ถูกนำมาใช้เพื่อประเมินประสิทธิภาพของโมเดลในการจำแนกข้อมูลเพิ่มเติมจากการวิเคราะห์ ROC Curve โดยเฉพาะในกรณีที่ต้องการพิจารณาความสามารถของโมเดลในการรักษาสัมดุลระหว่างค่า Precision และ Recall

ผลการวิเคราะห์ค่า Precision-Recall AUC ของแต่ละโมเดลแสดงให้เห็นว่าโมเดลทุกประเภทมีประสิทธิภาพอยู่ในระดับสูง โดยมีค่าดังนี้



ภาพที่ 4.6 เส้นโค้ง Precision-Recall ของโมเดลที่ใช้ในการทดลอง

ผลการวิเคราะห์พบว่าโมเดลทุกประเภทมีค่า Precision-Recall AUC อยู่ในระดับสูง โดย Gradient Boosting มีค่า AUC สูงสุด อยู่ที่ร้อยละ 99.95 รองลงมาคือ Random Forest มีค่า AUC อยู่ที่ร้อยละ 99.96, CNN มีค่า AUC อยู่ที่ร้อยละ 99.81, LSTM มีค่า AUC อยู่ที่ร้อยละ 99.76 และ Neural Network มีค่า AUC อยู่ที่ร้อยละ 99.31

เส้นโค้งของโมเดลส่วนใหญ่อยู่ใกล้กับค่าบนสุดของกราฟ แสดงให้เห็นว่าโมเดลสามารถรักษาค่า Precision และ Recall ได้ในระดับสูง โดยเฉพาะโมเดลประเภท Tree-based ที่มีความสามารถในการทำนายได้อย่างแม่นยำและครอบคลุม

4.5 การเปรียบเทียบข้อดีข้อเสียของแต่ละโมเดล

4.5.1 Random Forest

โมเดล Random Forest เป็นเทคนิคการเรียนรู้แบบ Ensemble Learning ที่รวมผลลัพธ์จาก Decision Tree หลายต้นเพื่อเพิ่มความแม่นยำในการทำนาย จากผลการวิจัยของผู้วิจัยพบว่า โมเดลนี้มีความแม่นยำสูง อยู่ที่ร้อยละ 99.90 ซึ่งอยู่ในระดับเดียวกับโมเดลที่มีประสิทธิภาพดีที่สุด อีกทั้งยังมีเวลาฝึกสอนที่รวดเร็ว โดยใช้เวลาเพียงประมาณ 0.095 วินาที ทำให้เหมาะสมสำหรับการประยุกต์ใช้งานจริงในระบบที่ต้องการการประมวลผลอย่างรวดเร็ว และสามารถแสดงค่าความสำคัญของตัวแปร (Feature Importance) ได้อย่างชัดเจน ซึ่งช่วยให้สามารถตีความผลลัพธ์ของโมเดลได้ง่าย และช่วยให้เข้าใจปัจจัยที่มีผลต่อการทำนายได้อย่างมีประสิทธิภาพ อีกทั้งยังสามารถลดปัญหา overfitting ได้ดี เนื่องจากใช้วิธีการรวมผลลัพธ์จากหลายต้นไม่ตัดสินใจ และสามารถทำงานได้ดีกับข้อมูลประเภทตาราง (tabular data) โดยไม่จำเป็นต้องมีขั้นตอนการเตรียมข้อมูลที่ซับซ้อน

โมเดล Random Forest ยังมีข้อจำกัดบางประการ เช่น ขนาดของโมเดลที่ค่อนข้างใหญ่เนื่องจากต้องจัดเก็บต้นไม้จำนวนมาก และการตีความโครงสร้างภายในของโมเดลอาจซับซ้อนกว่าโมเดลเชิงเส้น นอกจากนี้ โมเดลอาจมีประสิทธิผลลดลงในกรณีที่ข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็วหรือมีรูปแบบที่ซับซ้อนมาก Random Forest จึงเหมาะสำหรับการใช้งานจริงที่ต้องการความแม่นยำสูง ความเสถียรของโมเดล และความสามารถในการตีความผลลัพธ์ได้อย่างชัดเจน

4.5.2 Gradient Boosting

Gradient Boosting เป็นเทคนิคการเรียนรู้แบบ Ensemble Learning ที่พัฒนาโมเดลอย่างต่อเนื่อง โดยแต่ละโมเดลจะเรียนรู้จากข้อผิดพลาดของโมเดลก่อนหน้า จากผลการวิจัยของผู้วิจัยพบว่า โมเดล Gradient Boosting มีความแม่นยำสูงอยู่ที่ร้อยละ 99.90 ซึ่งเทียบเท่ากับ Random Forest และยังมีเวลาฝึกสอนที่เร็วที่สุดในบรรดาโมเดลทั้งหมด โดยใช้เวลาเพียงประมาณ 0.077 วินาที Gradient Boosting มีความสามารถในการเรียนรู้จากข้อผิดพลาดของโมเดล Random Forest ทำให้สามารถปรับปรุงประสิทธิภาพการทำนายได้อย่างต่อเนื่อง อีกทั้งยังสามารถให้ค่าความสำคัญของตัวแปรที่มีความหมายต่อการวิเคราะห์ข้อมูล และมีประสิทธิภาพในการใช้หน่วยความจำที่ดี Gradient Boosting มีความอ่อนไหวต่อการตั้งค่า hyperparameters มากกว่า Random Forest หากตั้งค่าไม่เหมาะสมอาจทำให้เกิดปัญหา overfitting ได้ง่าย นอกจากนี้ กระบวนการฝึกสอนของ Gradient Boosting เป็นแบบลำดับขั้น (sequential learning) ซึ่งทำให้ไม่สามารถประมวลผลแบบขนานได้เต็มประสิทธิภาพ

ผู้วิจัยพบว่า ลักษณะของโมเดล Gradient Boosting จึงเหมาะสำหรับการใช้งานใน production environment ที่ต้องการทั้งความแม่นยำสูงและความรวดเร็วในการประมวลผล

4.5.3 Convolutional Neural Network (CNN)

Convolutional Neural Network หรือ CNN เป็นโมเดลในกลุ่ม Deep Learning ที่ได้รับการออกแบบมาเพื่อเรียนรู้รูปแบบข้อมูลที่ซับซ้อน โดยเฉพาะข้อมูลที่มีโครงสร้างเชิงพื้นที่ เช่น ภาพหรือสัญญาณ จากผลการวิจัยของผู้วิจัยพบว่า โมเดล CNN มีความแม่นยำอยู่ที่ร้อยละ 97.30 ซึ่งอยู่ในระดับที่ดีและสามารถยอมรับได้สำหรับการทำนายข้อมูล CNN มีความสามารถในการเรียนรู้รูปแบบข้อมูลที่ซับซ้อน และสามารถปรับโครงสร้างสถาปัตยกรรมของโมเดลให้เหมาะสมกับลักษณะของข้อมูลได้ นอกจากนี้ โมเดลยังสามารถนำไปประยุกต์ใช้กับข้อมูลที่มีมิติสูง เช่น ข้อมูลภาพหรือข้อมูลจากเซนเซอร์ในระบบ IoT โมเดล CNN มีข้อจำกัดสำคัญคือใช้เวลาฝึกสอนค่อนข้างนาน โดยในการทดลองนี้ใช้เวลาประมาณ 3.811 วินาที ซึ่งมากกว่าโมเดลประเภท tree-based อย่างมีนัยสำคัญ อีกทั้งยังต้องการข้อมูลจำนวนมากเพื่อให้ได้ประสิทธิภาพที่ดี และการตีความผลลัพธ์ของโมเดลทำได้ค่อนข้างยาก เนื่องจากเป็นโมเดลแบบ black-box

ดังนั้น CNN จึงเหมาะสำหรับการศึกษาเชิงวิจัยหรือการประยุกต์ใช้กับข้อมูลที่มีความซับซ้อนสูงมากกว่าการใช้งานกับข้อมูล tabular ขนาดเล็ก

4.5.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory หรือ LSTM เป็นโมเดลในกลุ่ม Recurrent Neural Network ที่ออกแบบมาเพื่อจัดการกับข้อมูลลำดับเวลา (time series data) โดยสามารถจดจำความสัมพันธ์ของข้อมูลในช่วงเวลาต่าง ๆ ได้อย่างมีประสิทธิภาพ จากผลการวิจัยของผู้วิจัยพบว่า โมเดล LSTM มีความแม่นยำอยู่ที่ร้อยละ 97.20 ซึ่งใกล้เคียงกับโมเดล CNN ข้อดีของ LSTM คือสามารถเรียนรู้ความสัมพันธ์ของข้อมูลตามลำดับเวลาได้ดี จึงเหมาะสำหรับการนำไปประยุกต์ใช้กับข้อมูลที่มีการเปลี่ยนแปลงตามเวลา เช่น ข้อมูลจากเซนเซอร์ที่เก็บอย่างต่อเนื่อง หรือข้อมูลการเปลี่ยนแปลงของคุณภาพน้ำในระยะยาว

ดังนั้น โมเดล LSTM มีความซับซ้อนค่อนข้างสูง และใช้เวลาฝึกสอนมากที่สุดในบรรดาโมเดลทั้งหมด โดยใช้เวลาประมาณ 3.907 วินาที อีกทั้งยังต้องการการปรับแต่งพารามิเตอร์ที่ละเอียด และต้องการข้อมูลจำนวนมากเพื่อให้สามารถเรียนรู้ได้อย่างมีประสิทธิภาพ LSTM จึงเหมาะสำหรับการประยุกต์ใช้กับข้อมูลลำดับเวลา หรือการศึกษาการเปลี่ยนแปลงของข้อมูลในระยะยาว มากกว่าการใช้งานกับข้อมูลที่ไม่มีลักษณะเชิงเวลา

4.5.5 Neural Network

Neural Network เป็นโมเดลพื้นฐานของ Deep Learning ที่สามารถเรียนรู้ความสัมพันธ์แบบไม่เชิงเส้นระหว่างตัวแปรได้ จากผลการวิจัยของผู้วิจัยพบว่า โมเดลนี้มีความแม่นยำอยู่ที่ร้อยละ 94.50 ซึ่งต่ำกว่าโมเดลอื่นในการทดลองครั้งนี้ข้อดีของ Neural Network คือมีโครงสร้างที่ยืดหยุ่นและสามารถปรับแต่งสถาปัตยกรรมของโมเดลได้หลากหลาย รวมทั้งสามารถเรียนรู้

ความสัมพันธ์ที่ซับซ้อนของข้อมูลได้ โมเดล Neural Network มีข้อจำกัดหลายประการ เช่น มีความอ่อนไหวต่อการตั้งค่าพารามิเตอร์เริ่มต้นและ hyperparameters และอาจเกิดปัญหา overfitting ได้ง่ายหากไม่มีการใช้เทคนิค regularization ที่เหมาะสม

ดังนั้น Neural Network จึงเหมาะสำหรับการศึกษาเชิงทดลองหรือการพัฒนาโมเดล Deep Learning ในขั้นต้น มากกว่าการนำไปใช้เป็นโมเดลหลักในระบบทำนายจริง

4.6 การตรวจสอบความถูกต้องของโมเดล

เพื่อประเมินความน่าเชื่อถือและความถูกต้องของโมเดลที่พัฒนาขึ้น ผู้วิจัยได้ดำเนินการตรวจสอบประสิทธิภาพของโมเดลด้วยวิธีการหลายรูปแบบ ได้แก่ การทดสอบด้วยเทคนิค cross-validation การทดสอบกับกรณีข้อมูลที่มีค่าผิดปกติ และการเปรียบเทียบผลลัพธ์กับหลักการทางชีววิทยาที่เกี่ยวข้องกับการเพาะเลี้ยงปลาในกระชัง การตรวจสอบดังกล่าวมีวัตถุประสงค์เพื่อประเมินความสามารถของโมเดลในการเรียนรู้รูปแบบข้อมูลอย่างแท้จริง และลดความเสี่ยงของการเกิดปัญหา overfitting

การทดสอบด้วยวิธี cross-validation เป็นกระบวนการแบ่งชุดข้อมูลออกเป็นหลายส่วนเพื่อนำมาใช้ในการฝึกสอนและทดสอบโมเดลสลับกันในแต่ละรอบ ซึ่งช่วยให้สามารถประเมินประสิทธิภาพของโมเดลได้อย่างมีความน่าเชื่อถือมากขึ้น ผลการทดสอบพบว่าโมเดลประเภท tree-based เช่น Random Forest และ Gradient Boosting มีค่าความแม่นยำเฉลี่ยสูงและมีค่าความแปรปรวนต่ำ แสดงให้เห็นถึงความเสถียรของโมเดลในการทำนายข้อมูล

ผู้วิจัยได้ทดสอบโมเดลกับกรณีข้อมูลที่มีค่าผิดปกติ เช่น ค่า pH ที่ต่ำหรือสูงเกินช่วงมาตรฐานระดับออกซิเจนละลายน้ำที่ต่ำมาก และอุณหภูมิที่อยู่นอกช่วงที่เหมาะสมสำหรับการเลี้ยงปลาในกระชัง ผลการทดสอบพบว่าโมเดลสามารถจำแนกสภาพน้ำที่ไม่เหมาะสมได้อย่างถูกต้อง ซึ่งแสดงให้เห็นว่าโมเดลมีความสามารถในการตอบสนองต่อสถานการณ์ที่มีความเสี่ยงได้อย่างเหมาะสม ผลการวิเคราะห์ความสำคัญของตัวแปรยังสอดคล้องกับหลักการทางชีววิทยาในการเพาะเลี้ยงสัตว์น้ำ โดยพบว่าค่า pH เป็นปัจจัยที่มีความสำคัญมากที่สุด รองลงมาคือค่าออกซิเจนละลายน้ำ และอุณหภูมิ ซึ่งสอดคล้องกับแนวทางการจัดการคุณภาพน้ำในการเพาะเลี้ยงปลาในกระชัง ดังนั้น ผลการตรวจสอบดังกล่าวจึงยืนยันว่าโมเดลที่พัฒนาขึ้นมีความถูกต้องและสามารถเรียนรู้รูปแบบของข้อมูลได้อย่างเหมาะสม

4.7 การประยุกต์ใช้งานจริง

ผลการวิจัยในครั้งนี้แสดงให้เห็นว่าการประยุกต์ใช้เทคนิค Machine Learning สามารถนำมาใช้ในการพัฒนาระบบช่วยตัดสินใจด้านการจัดการคุณภาพน้ำในการเพาะเลี้ยงปลาในกระชังได้อย่างมีประสิทธิภาพ โดยระบบที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ร่วมกับเทคโนโลยี Internet of Things (IoT) เพื่อสนับสนุนการดำเนินงานของฟาร์มอัจฉริยะ (Smart Farm)

ในระบบการเลี้ยงปลาแบบดั้งเดิม การประเมินคุณภาพน้ำมักอาศัยประสบการณ์ของเกษตรกรหรือการตรวจวัดด้วยอุปกรณ์พื้นฐาน ซึ่งอาจทำให้การตัดสินใจมีความคลาดเคลื่อน อย่างไรก็ตาม การนำเทคโนโลยี Machine Learning มาประยุกต์ใช้ร่วมกับเซนเซอร์ IoT สามารถช่วยเก็บข้อมูลคุณภาพน้ำแบบอัตโนมัติและวิเคราะห์ข้อมูลได้อย่างรวดเร็ว ส่งผลให้เกษตรกรสามารถตรวจสอบความเหมาะสมของสภาพน้ำได้อย่างแม่นยำมากขึ้น

ระบบทำนายที่พัฒนาขึ้นยังสามารถใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการจัดการฟาร์ม เช่น การตรวจสอบสภาพน้ำก่อนปล่อยปลา การติดตามคุณภาพน้ำในระหว่างการเลี้ยง และการแจ้งเตือนเมื่อคุณภาพน้ำอยู่ในระดับที่อาจส่งผลกระทบต่ออาการเจริญเติบโตของปลา การประยุกต์ใช้ระบบดังกล่าวจึงสามารถช่วยเพิ่มประสิทธิภาพการผลิต ลดความเสี่ยงในการสูญเสียจากสภาพน้ำที่ไม่เหมาะสม และสนับสนุนการพัฒนาการเพาะเลี้ยงสัตว์น้ำอย่างยั่งยืน

4.8 ข้อจำกัดของการศึกษา

การศึกษาครั้งนี้จะแสดงให้เห็นถึงศักยภาพของการใช้เทคนิค Machine Learning ในการทำนายความเหมาะสมของคุณภาพน้ำสำหรับการเลี้ยงปลาในกระชัง อย่างไรก็ตาม ยังมีข้อจำกัดบางประการที่ควรนำมาพิจารณาในการตีความผลการวิจัย

4.8.1 ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลสังเคราะห์ (synthetic data) ซึ่งถูกสร้างขึ้นจากช่วงค่าที่อ้างอิงตามหลักการทางชีววิทยา แม้ว่าข้อมูลดังกล่าวจะช่วยให้สามารถควบคุมโครงสร้างของข้อมูลได้ แต่ก็อาจไม่สะท้อนความซับซ้อนของสภาพแวดล้อมจริงในฟาร์มเพาะเลี้ยงสัตว์น้ำได้อย่างสมบูรณ์

4.8.2 จำนวนตัวแปรที่ใช้ในการวิเคราะห์ยังมีจำกัด โดยการศึกษาพิจารณาเพียงสามพารามิเตอร์หลัก ได้แก่ ค่า pH อุณหภูมิ และปริมาณออกซิเจนละลายน้ำ ในขณะที่การจัดการคุณภาพน้ำในสภาพแวดล้อมจริงอาจได้รับอิทธิพลจากปัจจัยอื่น ๆ เช่น แอมโมเนีย ไนไตรต์ ความขุ่นของน้ำ หรือสภาพอากาศ

4.8.3 ชุดข้อมูลที่ใช้ในการศึกษาครั้งนี้ไม่ได้รวมมิติของเวลา (time series) ซึ่งในความเป็นจริงคุณภาพน้ำในกระชังเลี้ยงปลามีการเปลี่ยนแปลงอย่างต่อเนื่องตามช่วงเวลา การนำ

ข้อมูลลักษณะ time series มาใช้ในการพัฒนาโมเดลอาจช่วยเพิ่มความสามารถในการทำนายแนวโน้มของคุณภาพน้ำในอนาคตได้ดียิ่งขึ้น

ดังนั้น ในการศึกษาครั้งต่อไปควรมีการรวบรวมข้อมูลจากฟาร์มเพาะเลี้ยงจริง เพิ่มจำนวนพารามิเตอร์ที่ใช้ในการวิเคราะห์ และพัฒนาโมเดลที่สามารถวิเคราะห์ข้อมูลเชิงเวลา เพื่อเพิ่มความแม่นยำและความสามารถในการประยุกต์ใช้งานจริงของระบบทำนายในอนาคต



บทที่ 5

สรุปและอภิปรายผล

บทนี้เป็นการสรุปผลการวิจัยจากการศึกษาการเปรียบเทียบประสิทธิภาพของอัลกอริทึม Machine Learning ในการทำนายความเหมาะสมของคุณภาพน้ำสำหรับการเลี้ยงปลาในกระชัง โดยใช้ข้อมูลจากอุปกรณ์ Internet of Things (IoT) ซึ่งประกอบด้วยพารามิเตอร์สำคัญ ได้แก่ อุณหภูมิ ค่า pH และปริมาณออกซิเจนละลายน้ำ การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบประสิทธิภาพของโมเดล Machine Learning หลายประเภท ได้แก่ Random Forest, Gradient Boosting, Neural Network, Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM) เพื่อหาโมเดลที่มีความเหมาะสมและมีประสิทธิภาพสูงสุดสำหรับการนำไปประยุกต์ใช้ในการประเมินคุณภาพน้ำสำหรับการเพาะเลี้ยงสัตว์น้ำ จากผลการวิจัยของผู้วิจัยในบทที่ 4 ได้มีการนำเสนอผลการวิเคราะห์ข้อมูลเบื้องต้น ผลการเปรียบเทียบประสิทธิภาพของโมเดล Machine Learning การวิเคราะห์ความสำคัญของตัวแปร รวมถึงการประเมินประสิทธิภาพของโมเดลด้วยตัวชี้วัดทางสถิติต่าง ๆ เช่น Accuracy, Precision, Recall, F1-score และค่า Area Under the Curve (AUC) เพื่อนำมาวิเคราะห์และสรุปผลการวิจัย ตลอดจนอภิปรายผลเชิงลึกเกี่ยวกับความเหมาะสมของโมเดลแต่ละประเภทในการนำไปประยุกต์ใช้กับระบบการจัดการคุณภาพน้ำในการเพาะเลี้ยงปลาในกระชัง โดยแบ่งเนื้อหาออกเป็น 4 ส่วน ดังนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

5.4 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

5.1 สรุปผลการวิจัย

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึม Machine Learning ในการทำนายความเหมาะสมของคุณภาพน้ำสำหรับการเลี้ยงปลาในกระชัง โดยใช้ข้อมูลจากอุปกรณ์ Internet of Things (IoT) ซึ่งประกอบด้วยพารามิเตอร์สำคัญ ได้แก่ ค่า pH อุณหภูมิ และปริมาณออกซิเจนละลายน้ำ (Dissolved Oxygen: DO) โดยมีการนำอัลกอริทึม Machine Learning จำนวน 5 ประเภทมาทดสอบและเปรียบเทียบ ได้แก่ Random Forest, Gradient Boosting, Neural Network, Convolutional Neural Network (CNN) และ Long Short-Term Memory (LSTM)

การดำเนินการวิจัย ผู้วิจัยได้ทำการเตรียมชุดข้อมูลและแบ่งข้อมูลออกเป็นชุดฝึกสอนและชุดทดสอบ เพื่อใช้ในการพัฒนาและประเมินประสิทธิภาพของโมเดล จากนั้นจึงทำการฝึกสอนโมเดลแต่ละประเภทและประเมินผลด้วยตัวชี้วัดด้านประสิทธิภาพ เช่น Accuracy, Precision, Recall, F1-score และค่า Area Under the Curve (AUC)

ผู้ศึกษาวิจัยพบว่าโมเดลประเภท tree-based ได้แก่ Random Forest และ Gradient Boosting มีประสิทธิภาพสูงที่สุด โดยสามารถให้ค่าความแม่นยำสูงมีค่าอยู่ที่ ร้อยละ 99.90 และมีค่า AUC อยู่ในระดับใกล้เคียง 1.0 ซึ่งแสดงให้เห็นถึงความสามารถในการจำแนกความเหมาะสมของคุณภาพน้ำได้อย่างแม่นยำ ในขณะที่โมเดล Deep Learning ได้แก่ CNN และ LSTM ให้ค่าความแม่นยำมีค่าอยู่ที่ ร้อยละ 97.00 ส่วนโมเดล Neural Network มีค่าความแม่นยำมีค่าอยู่ที่ ร้อยละ 94.00

ผลการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance) จากโมเดล Random Forest และ Gradient Boosting พบว่าค่า pH เป็นปัจจัยที่มีความสำคัญมากที่สุดในการทำนายความเหมาะสมของคุณภาพน้ำ รองลงมาคือปริมาณออกซิเจนละลายน้ำ และอุณหภูมิ ตามลำดับ ผลลัพธ์ดังกล่าวสอดคล้องกับหลักการทางชีววิทยาในการเพาะเลี้ยงปลาในกระชัง ซึ่งระบุว่าค่า pH และระดับออกซิเจนละลายน้ำมีผลโดยตรงต่อสุขภาพและการเจริญเติบโตของปลา

จากการวิเคราะห์ทั้งหมด ผู้วิจัยสามารถสรุปได้ว่าโมเดล Gradient Boosting เป็นโมเดลที่เหมาะสมที่สุดสำหรับการนำไปประยุกต์ใช้ในระบบทำนายคุณภาพน้ำ เนื่องจากมีทั้งความแม่นยำสูงและใช้เวลาในการฝึกสอนต่ำ จึงเหมาะสำหรับการนำไปใช้งานในระบบฟาร์มอัจฉริยะ (Smart Farm) ที่ต้องการการประมวลผลข้อมูลอย่างรวดเร็วและมีประสิทธิภาพ

5.2 อภิปรายผลการวิจัย

ผลการศึกษาพบว่าอัลกอริทึมประเภท tree-based ได้แก่ Random Forest และ Gradient Boosting มีประสิทธิภาพสูงกว่าโมเดล Deep Learning ในการจำแนกความเหมาะสมของคุณภาพน้ำ ซึ่งอาจเนื่องมาจากลักษณะของข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลแบบตาราง (tabular data) และมีจำนวนตัวแปรไม่มาก ทำให้อัลกอริทึมประเภท tree-based สามารถเรียนรู้รูปแบบของข้อมูลได้อย่างมีประสิทธิภาพมากกว่า

นอกจากนี้ โมเดล tree-based ยังมีข้อได้เปรียบในด้านความสามารถในการตีความผลลัพธ์ เนื่องจากสามารถแสดงค่าความสำคัญของตัวแปรได้อย่างชัดเจน ทำให้สามารถวิเคราะห์ปัจจัยที่มีผลต่อการตัดสินใจของโมเดลได้ ซึ่งเป็นประโยชน์อย่างมากในการนำผลการวิเคราะห์ไปประยุกต์ใช้ด้านการจัดการคุณภาพน้ำในการเพาะเลี้ยงสัตว์น้ำ ในขณะที่โมเดล Deep Learning เช่น CNN และ LSTM แม้จะสามารถเรียนรู้รูปแบบข้อมูลที่ซับซ้อนได้ดี แต่เนื่องจากข้อมูลที่ใช้ในการวิจัยมีขนาดไม่

ใหญ่มากและไม่มีโครงสร้างเชิงพื้นที่หรือเชิงเวลาอย่างชัดเจน จึงทำให้ประสิทธิภาพของโมเดล Deep Learning ไม่ได้สูงกว่าโมเดลประเภท tree-based อย่างมีนัยสำคัญ

ผลการวิเคราะห์ความสำคัญของตัวแปรยังแสดงให้เห็นว่าค่า pH เป็นปัจจัยที่มีผลต่อความเหมาะสมของคุณภาพน้ำมากที่สุด ซึ่งสอดคล้องกับหลักการทางชีววิทยาในการเพาะเลี้ยงปลาในกระชัง เนื่องจากค่า pH ที่อยู่นอกช่วงที่เหมาะสมสามารถส่งผลกระทบต่อกระบวนการทางสรีรวิทยาของปลา รวมถึงการดูดซึมออกซิเจนผ่านเหงือก และความสมดุลของเกลือแร่ในร่างกายปลา

นอกจากนี้ ระดับออกซิเจนละลายน้ำยังเป็นปัจจัยสำคัญที่มีผลต่อการอยู่รอดและการเจริญเติบโตของปลา โดยระดับออกซิเจนที่ต่ำเกินไปอาจทำให้ปลาเกิดความเครียดและเพิ่มความเสี่ยงต่อการเกิดโรค ดังนั้น การตรวจสอบและควบคุมระดับออกซิเจนในน้ำจึงเป็นปัจจัยสำคัญในการจัดการฟาร์มเพาะเลี้ยงปลา

5.3 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

จากผลการวิจัยที่ได้ ระบบทำนายคุณภาพน้ำที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในระบบฟาร์มอัจฉริยะ โดยเชื่อมต่อกับอุปกรณ์ IoT เพื่อเก็บข้อมูลคุณภาพน้ำแบบอัตโนมัติ และใช้โมเดล Machine Learning ในการวิเคราะห์ข้อมูลแบบเรียลไทม์ ซึ่งจะช่วยให้เกษตรกรสามารถตรวจสอบความเหมาะสมของสภาพน้ำได้อย่างรวดเร็วและแม่นยำ

ระบบดังกล่าวยังสามารถใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการจัดการฟาร์ม เช่น การตรวจสอบคุณภาพน้ำก่อนปล่อยปลา การติดตามคุณภาพน้ำในระหว่างการเลี้ยง และการแจ้งเตือนเมื่อคุณภาพน้ำอยู่ในระดับที่อาจส่งผลกระทบต่อ การเจริญเติบโตของปลา

5.4 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

ในการศึกษาครั้งต่อไปควรมีการรวบรวมข้อมูลจากฟาร์มเพาะเลี้ยงจริงในหลายพื้นที่ เพื่อให้โมเดลสามารถเรียนรู้รูปแบบของข้อมูลที่มีความหลากหลายมากขึ้น นอกจากนี้ ควรเพิ่มจำนวนพารามิเตอร์ที่ใช้ในการวิเคราะห์ เช่น แอมโมเนีย ไนไตรต์ ความขุ่นของน้ำ และค่าความเป็นด่างของน้ำ ซึ่งอาจมีผลต่อคุณภาพน้ำในการเพาะเลี้ยงสัตว์น้ำ

ควรมีการพัฒนาโมเดลที่สามารถวิเคราะห์ข้อมูลเชิงเวลา (time series) เพื่อทำนายแนวโน้มการเปลี่ยนแปลงของคุณภาพน้ำในอนาคต รวมถึงการขยายขอบเขตชุดข้อมูลให้มีความหลากหลายและมีปริมาณที่มากขึ้น เพื่อให้โมเดลในกลุ่ม Deep Learning สามารถแสดงศักยภาพในการเรียนรู้รูปแบบความสัมพันธ์ที่ซับซ้อนได้ดียิ่งขึ้น และผู้วิจัยเล็งเห็นว่าในอนาคตสามารถพัฒนาระบบแจ้ง

เตือนล่วงหน้าเมื่อมีแนวโน้มว่าคุณภาพน้ำจะอยู่ในระดับที่ไม่เหมาะสม ซึ่งจะช่วยให้เพิ่มประสิทธิภาพในการจัดการฟาร์มและลดความเสี่ยงในการสูญเสียจากสภาพแวดล้อมที่ไม่เหมาะสม



บรรณานุกรม

- S. Khunthongchan, T. Athipanyakul, and S. Sairuamyaat, "Driving agriculture toward modern agriculture in Thailand: A case study of freshwater fish farming in the central region," *Local Administration Journal*, vol. 15, no. 4, 2022.
- P. Buangam, P. Jaiman, and S. Jankaew, "Problems, needs, and components for developing a smart agriculture system for precision tilapia farming in coastal areas," *Journal of Science and Technology to Community*, vol. 3, no. 4, Jul.-Aug. 2025.
- A. Yusoff and H. F. Hanafi, "A technology trend on IoT based fish aquaculture sensor: A comprehensive review," *ASEAN Artificial Intelligence Journal*, 2025.
- A. F. Abdullah, H. C. Man, M. M. A. Karim, A. Mohammed, S. U. Yunusa, and N. A. B. M. Jais, "Charting the aquaculture Internet of Things impact: Key applications, challenges, and future trends," *Aquaculture Reports*.
- Department of Fisheries, "Water quality for aquaculture," Bangkok, Thailand: Department of Fisheries, 2014.
- Department of Fisheries, "Water quality management in aquaculture," Bangkok, Thailand: Department of Fisheries, 2020.
- L. K. McKemmish, S. N. Yurchenko, and J. Tennyson, "Ab initio calculations to support accurate modelling of the rovibronic spectroscopy of vanadium monoxide (VO)," 2016.
- M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," 2018.
- Department of Fisheries, *A study to determine guidelines and policies for cage fish farming in public water resources*, Bangkok, Thailand: Department of

Fisheries, 2012. [Online]. Available:

https://www.fisheries.go.th/research_cage_fish

- S. Karle, V. Bansode, P. Tambet, and R. Bhambare, "IoT based greenhouse monitoring system using Raspberry Pi," *International Journal of Scientific Research in Science and Technology (IJSRST)*, vol. 8, no. 4, pp. 360–367, Jul.–Aug. 2021, doi: 10.32628/IJSRST218437.
- A. Kantasa-ard, M. Nouiri, A. Bekrar, A. Ait El Cadi, and Y. Saliez, "Machine learning for demand forecasting in the Physical Internet: A case study of agricultural products in Thailand," *Procedia Computer Science*, vol. 200, pp. 1123–1132, 2022.
- W. Kusonkhum, K. Srinavin, N. Leungbootnak, and T. Chaitongrat, "Using a machine learning approach to predict the Thailand underground train's passenger," *Journal of Advanced Transportation*, vol. 2022, Article ID 8789067, 15 pages, 2022, doi: 10.1155/2022/8789067.
- C. Phakdeephrom and A. Phuangchaktada, "Forecasting the goods demand through a learning-based machine in retail businesses," *Journal of Innovation and Management*, pp. 1–15.
- L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- E. Asamoah et al., "Random forest machine learning for maize yield prediction," 2024.
- G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- S. Pereira et al., "Development and assessment of artificial neural network models for solar radiation forecasting," *Energy Reports*, 2024.

- M. S. A. Bakar et al., "Crop yield prediction using artificial neural networks based on climate variables," *Computers and Electronics in Agriculture*, vol. 195, 2022.
- J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- P. Mahesh and R. Soundrapandiyan, "Yield prediction for crops by gradient-based algorithms," *PLOS ONE*, 2024.
- M. A. M. M. Hridoy et al., "Predictive modelling of aquaculture water quality using IoT monitoring and machine learning," *Smart Agricultural Technology*, 2025.
- R. P. Shete et al., "IoT-driven ensemble machine learning model for accurate water quality prediction in aquaculture," *Smart Agricultural Technology*, 2025.
- T. Li et al., "Predicting aquaculture water quality using machine learning approaches," *Water*, vol. 14, no. 18, 2022.
- K. Anupama et al., "A machine learning approach to monitor water quality in aquaculture using IoT," *International Journal of Precision Engineering*, 2020.
- S. M. Naim et al., "Aquaculture water quality classification using XGBoost with explainable AI," *Water*, 2025.
- Patki, N., Wedge, R., and Veeramachaneni, K., "The Synthetic Data Vault," IEEE International Conference on Data Science and Advanced Analytics, 2016.
- I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implementation (OSDI)*, 2016, pp. 265–283.

A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. Vermont, Australia: Machine Learning Mastery, 2020.



ประวัติผู้เขียน

ชื่อ	นางสาวศมลรดา รัตนภักดี
ชื่อการค้นคว้าอิสระ	การศึกษาประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง สำหรับ การวิเคราะห์คุณภาพน้ำ กรณีศึกษาการเลี้ยงปลา
สาขาวิชา	การบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ
ประวัติ	สำเร็จการศึกษาระดับปริญญาตรี คณะสาธารณสุขศาสตร์ และสหเวชศาสตร์ หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาเวช ระเบียน วิทยาลัยเทคโนโลยีทางการแพทย์และสาธารณสุข กาญจนนา ภิเษก พ.ศ. 2566 – 2567 ทำงาน วิทยาลัยเทคโนโลยีทาง การแพทย์และสาธารณสุข กาญจนนาภิเษก ตำแหน่งปฏิบัติการด้าน การเรียนการสอนเวชระเบียน พ.ศ. 2567 – ปัจจุบัน ทำงาน โรงพยาบาลจุฬารัตน์ 9 แอร์ พอร์ตตำแหน่ง นักวิชาการเวชสถิติ

