ระบบพยากรณ์ความเสี่ยงน้ำล้น–น้ำแห้งของเขื่อน
ด้วยเทคนิคการเรียนรู้ของเครื่อง
DEVELOPMENT OF A MACHINE LEARNING-BASED RISK FORECASTING SYSTEM FOR DAM OVERFLOW AND WATER SHORTAGE

นายเขมภูมิ คุณเลี้ยง

นายภวดล มากสรอย

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
ปีการศึกษา 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ

ระบบพยากรณ์ความเสี่ยงน้ำล้น–น้ำแห้งของเขื่อน
ด้วยเทคนิคการเรียนรู้ของเครื่อง

นายเขมภูมิ คุณเลี้ยง
นายภวดล มากสรอย

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ
ปีการศึกษา 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ

ชื่อโครงงาน	ระบบพยากรณ์ความเสี่ยงน้ำล้น–น้ำแห้งของเขื่อนด้วยเทคนิคการเรียนรู้ของเครื่อง
โดย	นายเขมภูมิ คุณเลี้ยง นายภวดล มากสรอย
สาขาวิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.สุรีพร นวลนิ่ม

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ อนุมัติให้นับโครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

..... คณบดีคณะวิทยาศาสตร์และเทคโนโลยี
(ผู้ช่วยศาสตราจารย์ ดร.ครรชิต กำลังกล้า)

คณะกรรมการสอบโครงงาน

..... ประธานกรรมการ
(อาจารย์ ดร.ปิยะ ถิรพันธุ์เมธี)

..... อาจารย์ที่ปรึกษาโครงงาน
(อาจารย์ ดร.สุรีพร นวลนิ่ม)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.อรสา พัสดุ)

กิตติกรรมประกาศ

โครงงานนี้สำเร็จลุล่วงได้ด้วยความกรุณาจากอาจารย์สุรีพร นวลนิ่มอาจารย์ที่ปรึกษาโครงงานที่ได้ให้คำเสนอแนะ แนวคิด ตลอดจนแก้ไขบกพร่องต่าง ๆ มาโดยตลอด จนโครงงานเล่มนี้สมบูรณ์ผู้ศึกษาจึงขอกราบขอบพระคุณเป็นอย่างสูง สุดท้ายนี้ ผู้ศึกษาหวังว่าโครงงานเล่มนี้จะเป็นประโยชน์แก่ผู้ที่สนใจและสามารถนำไปต่อยอดได้ต่อไป ทั้งนี้ หากมีข้อผิดพลาดหรือบกพร่องประการใด ผู้ศึกษาขอน้อมรับไว้เพื่อเป็นบทเรียนในการ พัฒนาต่อไปในอนาคต

นายเขมภูมิ คุณเลี้ยง
นายภวดล มากสรอย

ชื่อโครงงาน ระบบพยากรณ์ความเสี่ยงน้ำล้น–น้ำแห้งของเขื่อนด้วยเทคนิคการเรียนรู้ของเครื่อง
โดย นายเขมภูมิ คุณเลี้ยง
นายภวดล มากสรอย
สาขาวิชา วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษา อาจารย์ ดร.สุรีพร นวลนิ่ม
ปีการศึกษา 2568

บทคัดย่อ

โครงงานนี้จัดทำขึ้นมีวัตถุประสงค์เพื่อพัฒนาระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ของประเทศไทย เนื่องจากการประเมินความเสี่ยงด้วยเกณฑ์ร้อยละความจุแบบคงที่ไม่สามารถสะท้อนแนวโน้มการเปลี่ยนแปลงของระดับน้ำในอนาคตได้อย่างเพียงพอ จึงนำเทคนิคการเรียนรู้ของเครื่องมาใช้เพื่อเพิ่มความสามารถในการพยากรณ์ล่วงหน้าและสนับสนุนการตัดสินใจในการบริหารจัดการน้ำ วิธีการดำเนินงานเริ่มจากการรวบรวมข้อมูลสถานการณ์น้ำรายวันจากกรมชลประทานในรูปแบบ JSON และทำการเตรียมข้อมูล รวมถึงการจัดการค่าที่หายไปและการสร้างคุณลักษณะเชิงเวลา จากนั้นกำหนดฉลากระดับความเสี่ยงเป็น 3 ระดับ ได้แก่ เสี่ยงน้ำแห้ง ปกติ และเสี่ยงน้ำล้น เพื่อใช้เป็นตัวแปรเป้าหมายในงานจำแนกประเภท แบบจำลองถูกพัฒนาด้วยเทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอน ได้แก่ Random Forest, J48 Decision Tree และ Logistic Regression โดยแบ่งข้อมูลตามช่วงเวลาเพื่อใช้ฝึกสอนและทดสอบ และประเมินประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครอบคลุม (Recall) และค่าเอฟวันสกอร์ (F1-score)

นอกจากนี้ ได้พัฒนาแดชบอร์ดต้นแบบในรูปแบบเว็บแอปพลิเคชันเพื่อแสดงผลการพยากรณ์และข้อมูลสถานการณ์น้ำในรูปแบบที่เข้าใจง่าย เพื่อสนับสนุนการตัดสินใจของผู้ใช้งานและเพิ่มประสิทธิภาพในการบริหารจัดการทรัพยากรน้ำ

คำสำคัญ การพยากรณ์ความเสี่ยง อ่างเก็บน้ำขนาดใหญ่ การเรียนรู้ของเครื่อง แดชบอร์ด

สารบัญ

	หน้า
กิตติกรรมประกาศ	ค
บทคัดย่อ	ง
สารบัญ	จ
สารบัญตาราง	ฉ
สารบัญภาพ	ช
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของโครงงาน	1
1.2 วัตถุประสงค์ของโครงงาน	2
1.3 ขอบเขตของโครงงาน	2
1.4 เครื่องมือที่ใช้ในการพัฒนาโครงงาน	3
1.5 ขั้นตอนการดำเนินโครงงาน	3
1.6 ประโยชน์ที่(คาดว่าจะ)ได้รับ	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 ข้อมูลสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่และเกณฑ์ระดับความเสี่ยง	5
2.2 วิธีการเลือกแอตทริบิวต์หรือฟีเจอร์	8
2.3 อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithms)	9
2.4 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)	10
2.5 งานวิจัยที่เกี่ยวข้อง	11
บทที่ 3 วิธีการดำเนินงาน	14
3.1 การวิเคราะห์ระบบ	14
3.2 การออกแบบระบบ	17
3.3 การสร้างโมเดล	25
3.4 การพัฒนาระบบ	27
3.5 การทดสอบและประเมินผลระบบ	28
บรรณานุกรม	29

สารบัญตาราง

ตารางที่ **หน้า**

2.1 เกณฑ์การประเมินระดับความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำ 6

ก.1 อธิบายรายละเอียดแอตทริบิวต์ข้อมูลจาก API ของกรมชลประทาน 35

สารบัญภาพ

ภาพที่		หน้า
3.1	สถาปัตยกรรมระบบ	17
3.2	แผนภาพความสัมพันธ์ของฐานข้อมูล	19
3.3	การออกแบบส่วนติดต่อผู้ใช้หน้าจอแสดงผลสำหรับประชาชนทั่วไป	22
3.4	การออกแบบส่วนติดต่อผู้ใช้หน้าจอแสดงผลสำหรับเจ้าหน้าที่	23

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของโครงงาน

ประเทศไทยพึ่งพาอ่างเก็บน้ำขนาดใหญ่เป็นแหล่งน้ำหลักของประเทศ ทั้งเพื่อการเกษตร การอุปโภคบริโภค การผลิตกระแสไฟฟ้า และการป้องกันอุทกภัย โดยกรมชลประทานเป็นหน่วยงานหลักที่รับผิดชอบอ่างเก็บน้ำขนาดใหญ่จำนวน 35 แห่ง (กรมชลประทาน, 2568) ซึ่งถือเป็นโครงสร้างพื้นฐานด้านทรัพยากรน้ำที่มีความสำคัญอย่างยิ่ง ในช่วงสิบปีที่ผ่านมา ประเทศไทยประสบปัญหาน้ำท่วมจากกรณีน้ำล้นอ่างเก็บน้ำ เช่น ปี พ.ศ. 2554 และช่วงปี พ.ศ. 2564–2565 รวมถึงภัยแล้งรุนแรงในช่วงปี พ.ศ. 2558–2559 และ พ.ศ. 2562–2563 ส่งผลให้เกิดความเสียหายทางเศรษฐกิจและสังคมในวงกว้าง (สถาบันสารสนเทศทรัพยากรน้ำ, 2558; 2559; 2562; 2563)

แม้ปัจจุบันจะมีการแจ้งเตือนความเสี่ยงโดยใช้เกณฑ์ร้อยละความจุ เช่น ระดับน้ำต่ำกว่าร้อยละ 30 มีความเสี่ยงน้ำแห้ง หรือเกินร้อยละ 80 มีความเสี่ยงน้ำล้น (กรมชลประทาน, 2568) อย่างไรก็ตาม เกณฑ์คงที่ดังกล่าวยังมีข้อจำกัด เนื่องจากไม่สามารถสะท้อนแนวโน้มการเปลี่ยนแปลงของระดับน้ำในระยะสั้นได้อย่างเพียงพอ ส่งผลให้การคาดการณ์ความเสี่ยงล่วงหน้าอาจไม่แม่นยำ ซึ่งสอดคล้องกับงานวิจัยของ Mosavi et al., (2018) และ Zhang et al., (2023) ที่ระบุว่าการอาศัยเพียงข้อมูลสถานะปัจจุบันไม่เพียงพอต่อการประเมินความเสี่ยงในอนาคตอย่างมีประสิทธิภาพ ดังนั้น เพื่อเพิ่มความสามารถจากการประเมินสถานะปัจจุบันไปสู่การพยากรณ์ความเสี่ยงล่วงหน้า จึงควรนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาประยุกต์ใช้ในการวิเคราะห์ข้อมูลย้อนหลังและแนวโน้มเชิงเวลา เพื่อสนับสนุนการแจ้งเตือนเชิงคาดการณ์และการตัดสินใจในการบริหารจัดการน้ำได้อย่างทันท่วงที

จากเหตุผลดังกล่าว โครงงานนี้จึงพัฒนาระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ โดยประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) ผ่านโปรแกรม WEKA (Waikato Environment for Knowledge Analysis) และใช้ข้อมูลปริมาณน้ำรายวันจากกรมชลประทานเป็นชุดข้อมูลในการวิเคราะห์ ระบบสามารถจำแนกระดับความเสี่ยงได้ 3 ระดับ ได้แก่ เสี่ยงน้ำแห้ง ปกติ และเสี่ยงน้ำล้น โดยอาศัยการวิเคราะห์แนวโน้มการเปลี่ยนแปลงของระดับน้ำ นอกจากนี้ ยังได้พัฒนาแดชบอร์ดต้นแบบเพื่อแสดงผลการพยากรณ์และสนับสนุนการติดตามสถานการณ์

1.2 วัตถุประสงค์ของโครงงาน

1.2.1 เพื่อพัฒนาระบบพยากรณ์ระดับความเสี่ยงสถานการณ์น้ำของอ่างเก็บน้ำขนาดใหญ่ โดยประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) พร้อมแดชบอร์ดต้นแบบสำหรับแสดงผล

1.3 ขอบเขตของโครงงาน

โครงการนี้มุ่งเน้นการพัฒนาระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ โดยใช้ข้อมูลปริมาณน้ำของกรมชลประทานและเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อจำแนกระดับความเสี่ยงเป็น 3 ระดับ ได้แก่ น้ำแห้ง (Dry Risk) ปกติ (Medium Risk) และน้ำล้น (Flood Risk) โดยมีขอบเขตการดำเนินงานดังนี้

1.3.1 การดำเนินงานจะแบ่งออกเป็น 2 ส่วนได้แก่ การพัฒนาโมเดลการพยากรณ์ความเสี่ยงด้วยอัลกอริทึมการเรียนรู้ของเครื่อง และการพัฒนาแดชบอร์ดต้นแบบเพื่อแสดงผลการพยากรณ์

1.3.2 ใช้ข้อมูลปริมาณน้ำ อัตราการระบาย และปริมาณน้ำกักเก็บรายวันจากเว็บไซต์กรมชลประทานที่ https://app.rid.go.th/reservoir/

1.3.3 กำหนดฉลากความเสี่ยง (Risk Level Labels) จากตัวแปรเชิงตัวเลข ได้แก่ เปอร์เซ็นต์ความจุของอ่างเก็บน้ำขนาดใหญ่ (%Capacity) ด้วยกฎเกณฑ์ดังนี้

1.3.3.1 < 30% หมายถึง เสี่ยงน้ำแห้ง (Dry Risk)

1.3.3.2 30% – 80% หมายถึง ระดับปกติ (Medium Risk)

1.3.3.3 ≥ 80% หมายถึง เสี่ยงน้ำล้น (Flood Risk)

โดยฉลากทั้งสามระดับถูกใช้เป็นตัวแปรเป้าหมาย (Target Variable) สำหรับการจำแนกประเภท (Classification)

1.3.4 ศึกษาและเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง 3 รูปแบบ ได้แก่ เรนดอมฟอเรสต์ (Random Forest) ต้นไม้ตัดสินใจ (J48) และการถดถอยโลจิสติก (Logistic Regression) โดยใช้ตัวชี้วัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision), ค่าความครอบคลุม (Recall) และ F1-score เพื่อคัดเลือกโมเดลที่มีความเหมาะสมที่สุดในการพยากรณ์ความเสี่ยง

1.3.5 ในส่วนของการพัฒนาแดชบอร์ดต้นแบบ (Prototype Dashboard) สำหรับแสดงผลการพยากรณ์ความเสี่ยงน้ำในอ่างเก็บน้ำขนาดใหญ่ ระบบจะดึงข้อมูลปัจจุบันหรือข้อมูลล่าสุดที่จัดเก็บไว้แล้วแสดงผลลัพธ์ความเสี่ยง (เสี่ยงน้ำแห้ง ปกติ เสี่ยงน้ำล้น) พร้อมข้อมูลประกอบ เช่น เปอร์เซ็นต์ความจุของอ่างเก็บน้ำขนาดใหญ่ โดยใช้โมเดลที่มีประสิทธิภาพสูงสุดจากการทดลองเป็น

โมเดลหลักของระบบ ทั้งนี้ ได้กำหนดขอบเขตการแสดงผลบนแดชบอร์ดโดยแบ่งมุมมองผู้ใช้งานออกเป็น 2 กลุ่มหลัก เพื่อให้ตอบสนองต่อความต้องการที่แตกต่างกัน ได้แก่

1.3.5.1 มุมมองสำหรับประชาชนทั่วไป

นำเสนอข้อมูลสถานการณ์น้ำปัจจุบันและผลการพยากรณ์ความเสี่ยงในรูปแบบกราฟิกหรือสัญลักษณ์ที่เข้าใจง่าย เพื่อให้ประชาชนสามารถติดตามสถานการณ์และเตรียมความพร้อมรับมือได้อย่างทันท่วงที

1.3.5.2 มุมมองสำหรับเจ้าหน้าที่

การแสดงข้อมูลเชิงลึกทางสถิติ กราฟแสดงแนวโน้มร้อยละความจุย้อนหลัง ปริมาณน้ำไหลเข้า (Inflow) และระบายออก (Outflow) รวมถึงผลการวิเคราะห์เปรียบเทียบจากโมเดล เพื่อสนับสนุนการตัดสินใจของเจ้าหน้าที่ในการปรับแผนการบริหารจัดการและระบายน้ำ

1.3.6 ขอบเขตที่ไม่ครอบคลุมการดำเนินงาน

1.3.6.1 ไม่รวมการพยากรณ์ค่าปริมาณน้ำในลักษณะเชิงตัวเลข (Regression)

1.3.6.2 ไม่วิเคราะห์ปัจจัยเชิงสาเหตุ เช่น ปริมาณฝนหรือการบริหารจัดการน้ำ

1.3.6.3 ไม่พัฒนาระบบแจ้งเตือนแบบเรียลไทม์ผ่านมือถือ เว้นแต่ถูกระบุเพิ่มเติม

1.3.6.4 ไม่ใช้ข้อมูลจากอ่างเก็บน้ำขนาดใหญ่ทุกแห่งทั่วประเทศในกรณีที่ข้อมูลไม่สมบูรณ์หรือมีช่องว่างข้อมูลมาก

1.4 เครื่องมือที่ใช้ในการพัฒนาโครงงาน

1.4.1 โปรแกรม WEKA สำหรับการเตรียมข้อมูล (Data Preprocessing) การสร้างและเปรียบเทียบประสิทธิภาพของโมเดลการเรียนรู้ของเครื่อง

1.4.2 ภาษา Python และไลบรารี Streamlit สำหรับการพัฒนาเว็บแอปพลิเคชันแดชบอร์ดต้นแบบ โดยใช้ร่วมกับโปรแกรม Anaconda Navigator และ Visual Studio Code

1.4.3 ระบบจัดการฐานข้อมูล MySQL สำหรับจัดเก็บข้อมูลปริมาณน้ำและผลการพยากรณ์เพื่อนำไปแสดงผลบนแดชบอร์ด

1.5 ขั้นตอนการดำเนินโครงงาน

1.5.1 การวิเคราะห์ระบบ

1.5.2 การออกแบบระบบ

1.5.3 การสร้างโมเดล

1.5.4 การพัฒนาระบบ

1.5.5 การทดสอบและประเมินผลระบบ

1.6 ประโยชน์ที่(คาดว่าจะ)ได้รับ

1.6.1 ได้ระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ที่สามารถจำแนกระดับความเสี่ยง (เสี่ยงน้ำแห้ง ปกติ เสี่ยงน้ำล้น) โดยใช้เทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอนในโปรแกรม WEKA

1.6.2 ได้ระบบแดชบอร์ดแบบเว็บแอปพลิเคชันที่สามารถแสดงผลการพยากรณ์ระดับความเสี่ยงน้ำของอ่างเก็บน้ำขนาดใหญ่ที่เลือก

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การพัฒนาระบบพยากรณ์ความเสี่ยงน้ำล้น-น้ำแห้งของเขื่อนด้วยเทคนิคการเรียนรู้ของเครื่องให้สำเร็จตามวัตถุประสงค์ที่กำหนดไว้ ผู้จัดทำได้ทำการศึกษาและรวบรวมแนวคิด ทฤษฎี และรวมถึงงานวิจัยที่เกี่ยวข้อง ดังนี้

2.1 ข้อมูลสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่และเกณฑ์ระดับความเสี่ยง

2.2 วิธีการเลือกแอตทริบิวต์หรือฟีเจอร์

2.3 อัลกอริทึมการเรียนรู้ของเครื่อง

2.4 การประเมินประสิทธิภาพของแบบจำลอง

2.5 งานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่และเกณฑ์ระดับความเสี่ยง

2.1.1 ภาพรวมของอ่างเก็บน้ำขนาดใหญ่ในประเทศไทยและระบบติดตามสถานการณ์น้ำ อ่างเก็บน้ำขนาดใหญ่ เป็นโครงสร้างพื้นฐานด้านทรัพยากรน้ำที่มีความสำคัญต่อประเทศ ทั้งในด้านการกักเก็บน้ำเพื่อการอุปโภคบริโภค การเกษตรกรรม การผลิตกระแสไฟฟ้า และการบรรเทาอุทกภัย ปัจจุบันประเทศไทยมีอ่างเก็บน้ำขนาดใหญ่ภายใต้การกำกับดูแลของหน่วยงานภาครัฐ เช่น กรมชลประทาน และการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย จำนวนทั้งสิ้น 35 แห่งทั่วประเทศ ในการติดตามสถานการณ์น้ำ ประชาชนและหน่วยงานที่เกี่ยวข้องสามารถเข้าถึงข้อมูลได้ผ่านระบบออนไลน์ เช่น เว็บไซต์รายงานสถานการณ์น้ำของกรมชลประทาน (http://app.rid.go.th/reservoir/) หรือคลังข้อมูลน้ำแห่งชาติ (https://thaiwater.net/) นอกจากนี้ กรมชลประทานได้เปิดให้บริการข้อมูลสถานการณ์น้ำในรูปแบบอินเทอร์เฟซโปรแกรมประยุกต์ (Application Programming Interface: API) ผ่าน Endpoint ที่ https://app.rid.go.th/reservoir/api/document/dam/ การเรียกใช้งาน API ดังกล่าว ระบบจะตอบกลับข้อมูลดิบ (Raw Data) ของอ่างเก็บน้ำในรูปแบบโครงสร้าง JSON ซึ่งประกอบด้วยแอตทริบิวต์ (Attributes) ที่สำคัญต่อการวิเคราะห์ เช่น ชื่ออ่างเก็บน้ำ ความจุเก็บกัก ปริมาณน้ำปัจจุบัน ร้อยละของปริมาณน้ำเทียบกับความจุอ่างเก็บน้ำ (%Capacity) ปริมาณน้ำไหลเข้าอ่าง (Inflow) ปริมาณน้ำระบายออก (Outflow) และแอตทริบิวต์อื่น ๆ ช่วยให้สามารถดึงตัวแปรสำคัญเหล่านี้มาจัดเตรียมและป้อนเข้าสู่กระบวนการเรียนรู้ของเครื่อง (Machine Learning) เพื่อวิเคราะห์และพยากรณ์ความเสี่ยงของสถานการณ์น้ำแบบอัตโนมัติได้อย่างรวดเร็วและมีประสิทธิภาพ

2.1.2 เกณฑ์การประเมินระดับน้ำและระดับความเสี่ยง

การพิจารณาและกำหนดเกณฑ์เพื่อประเมินระดับความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่สำหรับโครงงานนี้ ได้ดำเนินการอ้างอิงจากเกณฑ์วิกฤติน้ำและแนวทางการบริหารจัดการอ่างเก็บน้ำของกรมทรัพยากรน้ำ (กรมทรัพยากรน้ำ, 2567) โดยใช้ร้อยละปริมาณน้ำใช้การ หรือ ร้อยละความจุของอ่างเก็บน้ำ (%Capacity) เป็นดัชนีชี้วัดหลักในการเฝ้าระวัง โดยแบ่งเกณฑ์การเตือนภัยและการบริหารจัดการออกเป็น 3 ระดับหลัก ดังตารางที่ 2.1

ตารางที่ 2.1 เกณฑ์การประเมินระดับความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำ

ระดับความเสี่ยง	ร้อยละ ความจุอ่าง	ลักษณะและแนวทางจัดการ
เสี่ยงน้ำแห้ง	< 30	ต้องเฝ้าระวังความเสี่ยงของการขาดแคลนน้ำ งดหรือลดการระบายน้ำเพื่อการเกษตรนอกแผน เพื่อสงวนน้ำไว้สำหรับอุปโภคบริโภค
ปกติ	30-80	อยู่ในเกณฑ์สมดุลและปลอดภัย สามารถบริหารจัดการและระบายน้ำตามแผนการจัดการน้ำปกติ
เสี่ยงน้ำล้น	> 80	ต้องเฝ้าระวังความมั่นคงและเสถียรภาพของเขื่อน มีความเสี่ยงสูงที่น้ำจะเอ่อล้น ต้องเร่งระบายน้ำหรือพร่องน้ำออกเพื่อเพิ่มพื้นที่ว่าง

2.1.2.1 ระดับเสี่ยงน้ำแห้ง

เกิดขึ้นเมื่อปริมาณน้ำกักเก็บในอ่างเก็บน้ำลดลงต่ำกว่าร้อยละ 30 ของความจุ สภาวะดังกล่าวถือเป็นเกณฑ์ระดับวิกฤตที่ต้องเฝ้าระวังความเสี่ยงของการขาดแคลนน้ำอย่างใกล้ชิด และอาจต้องงดหรือลดการระบายน้ำเพื่อการเกษตรนอกแผน เพื่อสงวนน้ำไว้สำหรับอุปโภคบริโภค ในระยะนี้มีความจำเป็นต้องยกระดับการบริหารจัดการน้ำอย่างรัดกุม (กรมทรัพยากรน้ำ, 2567)

2.1.2.2 ระดับปกติ

ระดับสถานการณ์น้ำปกติคือช่วงที่ปริมาณน้ำกักเก็บในอ่างเก็บน้ำมีระดับอยู่ระหว่างร้อยละ 30 ถึง 80 ของความจุ ในสถานการณ์นี้ ถือว่าความมั่นคงของทรัพยากรน้ำอยู่ในเกณฑ์สมดุลและปลอดภัย หน่วยงานที่รับผิดชอบสามารถดำเนินการบริหารจัดการและระบายน้ำเพื่อ

ตอบสนองต่อกิจกรรมต่าง ๆ ได้ตามแผนการจัดการน้ำปกติ โดยไม่มีความเสี่ยงต่อปัญหาอุทกภัยและภัยแล้ง (กรมทรัพยากรน้ำ, 2567)

2.1.2.3 ระดับเสี่ยงน้ำล้น

ระดับเสี่ยงน้ำล้นจะถูกประกาศขึ้นเมื่อปริมาณน้ำกักเก็บในอ่างเก็บน้ำเพิ่มสูงกว่าร้อยละ 80 ของความจุ สถานการณ์นี้เป็นจุดที่ต้องเฝ้าระวังความมั่นคงและเสถียรภาพของตัวเขื่อนเป็นพิเศษ เนื่องจากมีความเสี่ยงสูงที่น้ำจะเอ่อล้นข้ามอาคารระบายน้ำล้น (Spillway) ดังนั้น จึงมีความจำเป็นต้องปรับแผนโดยการเร่งระบายน้ำหรือพร่องน้ำออกจากอ่างเก็บน้ำ เพื่อเพิ่มพื้นที่ว่างสำหรับน้ำที่คาดว่าจะไหลเข้าสมทบในอนาคต ป้องกันผลกระทบจากอุทกภัยในพื้นที่ท้ายน้ำ (กรมทรัพยากรน้ำ, 2567)

2.1.3 อิทธิพลของฤดูกาลของประเทศไทยต่อปริมาณน้ำ

การเปลี่ยนแปลงของสภาพอากาศและฤดูกาลเป็นปัจจัยภายนอกที่ส่งผลกระทบโดยตรงต่อปริมาณน้ำไหลเข้า (Inflow) และความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำ จากเกณฑ์การประกาศเข้าสู่ฤดูกาลของกรมอุตุนิยมวิทยา ประเทศไทยสามารถแบ่งฤดูกาลออกเป็น 3 ฤดูหลัก ซึ่งมีช่วงเวลาที่ชัดเจนตามประกาศในแต่ละปี (กรมอุตุนิยมวิทยา, 2567; 2568) ซึ่งมีอิทธิพลต่อวัฏจักรของน้ำดังนี้

2.1.3.1 ฤดูร้อน (Summer)

มักเริ่มต้นในช่วงปลายเดือนกุมภาพันธ์และสิ้นสุดในช่วงกลางเดือนพฤษภาคม โดยอ้างอิงจากประกาศกรมอุตุนิยมวิทยาพบว่า ฤดูร้อนในปี พ.ศ. 2567 เริ่มต้นเมื่อวันที่ 21 กุมภาพันธ์ และปี พ.ศ. 2568 เริ่มต้นเมื่อวันที่ 28 กุมภาพันธ์ ในช่วงเวลานี้ปริมาณฝนจะทิ้งช่วง ทำให้อัตราการไหลเข้าของน้ำในอ่างเก็บน้ำลดลงอย่างมีนัยสำคัญ ส่งผลให้มีความเสี่ยงในการเกิดภาวะน้ำแห้งสูงขึ้น

2.1.3.2 ฤดูฝน (Rainy)

มักเริ่มต้นในช่วงกลางเดือนพฤษภาคมและสิ้นสุดในช่วงกลางเดือนตุลาคม ตัวอย่างเช่น ในปี พ.ศ. 2567 กรมอุตุนิยมวิทยาได้ประกาศเริ่มต้นฤดูฝนอย่างเป็นทางการเมื่อวันที่ 20 พฤษภาคม และปี พ.ศ. 2568 เมื่อวันที่ 15 พฤษภาคม เป็นช่วงที่ได้รับอิทธิพลจากมรสุมตะวันตกเฉียงใต้ ส่งผลให้มีฝนตกชุกและมีปริมาณน้ำไหลเข้าอ่างเก็บน้ำในปริมาณมาก ซึ่งเป็นช่วงที่ต้องเฝ้าระวังความเสี่ยงของสภาวะน้ำล้นและต้องบริหารจัดการพร่องน้ำอย่างรัดกุม

2.1.3.3 ฤดูหนาว (Winter)

มักเริ่มต้นในช่วงปลายเดือนตุลาคมและสิ้นสุดในปลายเดือนกุมภาพันธ์ของปีถัดไป โดยมีการประกาศเข้าสู่ฤดูหนาวในปี พ.ศ. 2567 เริ่มต้นเมื่อวันที่ 29 ตุลาคม และในปี พ.ศ.

2568 เริ่มต้นเมื่อวันที่ 23 ตุลาคม ในระยะนี้ปริมาณฝนจะเริ่มลดลงอย่างชัดเจน ยกเว้นในพื้นที่ภาคใต้ที่จะมีปริมาณฝนตกอย่างต่อเนื่อง ปริมาณน้ำไหลเข้าอ่างจะเริ่มคงที่และลดระดับลงตามลำดับ

2.2 วิธีการเลือกแอตทริบิวต์หรือฟีเจอร์

ในการพัฒนาระบบจำแนกประเภทข้อมูล (Classification) ชุดข้อมูลที่นำมาศึกษามักประกอบด้วยคุณลักษณะหรือฟีเจอร์ (Feature) จำนวนมาก ซึ่งไม่ใช่ทุกฟีเจอร์จะมีนัยสำคัญในการแบ่งแยกคลาสเป้าหมาย (Target Class) การนำฟีเจอร์ที่ไม่เกี่ยวข้องมาร่วมประมวลผลมักทำให้โมเดลเกิดความซับซ้อนโดยไม่จำเป็น และอาจนำไปสู่ปัญหาการเรียนรู้เกิน (Overfitting) ดังนั้นกระบวนการคัดเลือกฟีเจอร์จึงเป็นขั้นตอนสำคัญในการลดรูปข้อมูล เพื่อคงไว้เฉพาะตัวแปรที่มีอิทธิพลและมีประโยชน์ต่อชุดข้อมูลอย่างแท้จริง (เอกสิทธิ์ พัชรวงศ์ศักดา, 2558)

2.1.1 วิธีการแบบตัวกรอง (Filter Approach)

วิธีการแบบตัวกรองเป็นแนวทางการคัดเลือกฟีเจอร์ที่อาศัยการคำนวณคุณสมบัติทางสถิติ เพื่อหาค่าน้ำหนักหรือระดับความสัมพันธ์ระหว่างฟีเจอร์แต่ละตัวกับคลาสเป้าหมาย โดยจะทำการเรียงลำดับค่าน้ำหนักและคัดเลือกเฉพาะฟีเจอร์ที่ผ่านเกณฑ์ที่กำหนดไว้ไปใช้งานต่อ

จุดเด่นสำคัญของแนวทางนี้คือกระบวนการคัดเลือกเป็นอิสระจากตัวอัลกอริทึมการเรียนรู้ (Model-independent) ทำให้ใช้เวลาในการประมวลผลน้อยและเหมาะสมกับชุดข้อมูลขนาดใหญ่ โดยเทคนิคทางสถิติที่นิยมนำมาใช้คำนวณค่าน้ำหนัก ได้แก่ การทดสอบไคสแควร์ (Chi-Square), การวิเคราะห์ค่าความสัมพันธ์ (Correlation) และอินฟอร์เมชันเกน (Information Gain)

สำหรับโครงงานนี้ จึงเลือกใช้วิธีการแบบตัวกรองร่วมกับเทคนิคอินฟอร์เมชันเกน (Information Gain) เนื่องจากเป็นวิธีที่สามารถประมวลผลข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพสูง ซึ่งหลักการนี้จะทำงานควบคู่กับการวัดค่าความไม่แน่นอนของข้อมูลที่เรียกว่า เอนโทรปี (Shannon, 1948; เอกสิทธิ์ พัชรวงศ์ศักดา, 2558) ดังนี้

2.1.2 เอนโทรปี (Entropy)

เอนโทรปีเป็นมาตรวัดความไม่แน่นอน (Uncertainty) หรือความไม่บริสุทธิ์ (Impurity) ของชุดข้อมูล หากชุดข้อมูลมีการกระจายตัวของคลาสเป้าหมาย (Target Class) อย่างสม่ำเสมอ ค่าเอนโทรปีจะสูง ซึ่งหมายถึงมีความไม่แน่นอนสูง ในทางกลับกัน หากข้อมูลส่วนใหญ่ถูกจัดอยู่ในคลาสใดคลาสหนึ่งเพียงคลาสเดียว ค่าเอนโทรปีจะต่ำ หรือมีความไม่แน่นอนต่ำ โดยสามารถคำนวณค่าเอนโทรปีของชุดข้อมูลได้จากสมการดังนี้

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2 p_i$$

2.1.3 อินฟอร์เมชันเกน (Information Gain)

อินฟอร์เมชันเกน คือค่าที่แสดงถึงปริมาณความไม่แน่นอนที่ลดลง (Reduction in Entropy) หลังจากที่ชุดข้อมูลถูกแบ่งส่วนด้วยแอตทริบิวต์ใดแอตทริบิวต์หนึ่ง แอตทริบิวต์ที่เหมาะสมที่สุดในการนำมาใช้เป็นฟีเจอร์สำหรับโมเดล คือแอตทริบิวต์ที่ให้ค่า Information Gain สูงที่สุด ซึ่งแสดงถึงความสามารถในการแยกแยะคลาสเป้าหมายออกจากกันได้อย่างชัดเจน (เอกสิทธิ์ พัชรวงศ์ศักดา, 2558) สามารถคำนวณได้จากสมการดังนี้

$$IG(S,A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{S} Entropy(S_v)$$

2.1.4 การประยุกต์ใช้ในการดำเนินงานโครงงาน

ในการดำเนินการโครงงานนี้ การประเมินประสิทธิภาพของแอตทริบิวต์ ได้แก่ ปริมาณน้ำไหลเข้าอ่าง (Inflow), ปริมาณน้ำระบายออก (Outflow) และร้อยละความจุอ่างเก็บน้ำ (% Capacity) จะถูกประมวลผลผ่านฟังก์ชัน InfoGainAttributeEval ในโปรแกรม WEKA กระบวนการนี้ช่วยยืนยันทางสถิติว่าตัวแปรใดมีอิทธิพลสูงสุดในการจำแนกระดับความเสี่ยง (Risk Level) เพื่อคัดกรองเฉพาะฟีเจอร์ที่มีนัยสำคัญเข้าสู่กระบวนการฝึกสอนโมเดลต่อไป ซึ่งช่วยลดความซับซ้อนของข้อมูลและสามารถเพิ่มความแม่นยำในการพยากรณ์ได้อย่างมีประสิทธิภาพ (Witten et al., 2016)

2.3 อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithms)

ในการดำเนินโครงงานนี้ ผู้จัดทำได้เลือกใช้อัลกอริทึมสำหรับการจำแนกประเภท (Classification) จำนวน 3 อัลกอริทึม เพื่อนำมาสร้างแบบจำลองและเปรียบเทียบประสิทธิภาพ ดังนี้

2.3.1 ต้นไม้ตัดสินใจ (Decision Tree: J48)

เป็นโมเดลที่มีโครงสร้างคล้ายแผนผังต้นไม้ ประกอบด้วย โหนดราก (Root Node) โหนดภายใน (Internal Node) และใบ (Leaf Node) ซึ่งเป็นผลลัพธ์ของคลาสเป้าหมาย อัลกอริทึม J48 ในโปรแกรม WEKA เป็นการพัฒนาต่อยอดมาจากอัลกอริทึม C4.5 ของ Quinlan (1986) ซึ่งใช้หลักการของเอนโทรปี (Entropy) และอินฟอร์เมชันเกน (Information Gain) ในการพิจารณาเลือกฟีเจอร์ที่ดีที่สุดมาทำหน้าที่เป็นโหนดสำหรับการแตกกิ่ง (Split) ข้อมูลจากบนลงล่าง (Top-Down) จนกว่าจะสามารถจัดกลุ่มข้อมูลได้สมบูรณ์

2.3.2 แรนดอมฟอเรสต์ (Random Forest)

เป็นเทคนิคการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ที่นำเสนอโดย Breiman (2001) อัลกอริทึมนี้ทำงานโดยการสร้างต้นไม้ตัดสินใจหลาย ๆ ต้น (Multiple Decision Trees) จาก

กระบวนการสุ่มตัวอย่างข้อมูลแบบใส่คืน (Bootstrap Aggregating หรือ Bagging) ร่วมกับการสุ่มเลือกกลุ่มของฟีเจอร์ในการสร้างโหนด เมื่อต้องการทำนายผลลัพธ์ อัลกอริทึมจะนำผลลัพธ์จากต้นไม้ทุกต้นมารวบรวมและใช้วิธีโหวตเสียงข้างมาก (Majority Voting) เพื่อตัดสินผลลัพธ์สุดท้าย การทำงานในลักษณะนี้ช่วยลดปัญหาการเรียนรู้เกิน (Overfitting) ที่มักพบในต้นไม้ตัดสินใจแบบต้นเดียว และช่วยเพิ่มความแม่นยำของโมเดลได้อย่างมาก (Breiman, 2001)

2.3.3 การถดถอยโลจิสติก (Logistic Regression)

เป็นโมเดลทางสถิติที่ประยุกต์ใช้สำหรับงานจำแนกประเภท โดยเฉพาะการทำนายตัวแปรตามที่เป็นข้อมูลเชิงกลุ่ม (Categorical Variable) อัลกอริทึมนี้ทำงานโดยอาศัยฟังก์ชันโลจิสติก หรือ ซิกมอยด์ (Sigmoid Function) ในการแปลงค่าผลรวมเชิงเส้นของตัวแปรอิสระให้อยู่ในรูปของค่าความน่าจะเป็น (Probability) ซึ่งจะมีค่าอยู่ระหว่าง 0 ถึง 1 ทำให้สามารถทำนายได้ว่าชุดข้อมูลนั้นๆ ควรถูกจัดอยู่ในคลาสความเสี่ยงใด โดยพิจารณาจากค่าความน่าจะเป็นที่คำนวณได้ (Hosmer & Lemeshow, 2013)

2.4 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

การพิจารณาอัลกอริทึมที่มีความเหมาะสมที่สุด โครงงานนี้ได้ใช้มาตรวัดประสิทธิภาพต่างๆ โดยอาศัยผลลัพธ์ที่ได้จากการแจกแจงในเมทริกซ์ความสับสน (Confusion Matrix) เป็นฐานในการคำนวณ

2.4.1 เมทริกซ์ความสับสน (Confusion Matrix)

ตารางที่แสดงผลเปรียบเทียบระหว่างค่าที่โมเดลพยากรณ์ได้ (Predicted Class) กับค่าความเป็นจริงของข้อมูล (Actual Class) ซึ่งประกอบด้วยค่า 4 ส่วนหลัก ได้แก่

2.4.1.1 True Positive (TP) กรณีที่โมเดลทายว่าจริง และข้อมูลจริงเป็นจริง

2.4.1.2 True Negative (TN) กรณีที่โมเดลทายว่าเท็จ และข้อมูลจริงเป็นเท็จ

2.4.1.3 False Positive (FP) กรณีที่โมเดลทายว่าจริง แต่ข้อมูลจริงเป็นเท็จ

2.4.1.4 False Negative (FN) กรณีที่โมเดลทายว่าเท็จ แต่ข้อมูลจริงเป็นจริง

2.4.2 ตัวชี้วัดประสิทธิภาพ (Performance Metrics)

จากเมทริกซ์ความสับสน สามารถนำมาคำนวณหาตัวชี้วัดประสิทธิภาพที่สำคัญ ได้แก่

2.4.2.1 ความถูกต้อง (Accuracy) อัตราส่วนของการพยากรณ์ถูกทั้งหมดเทียบกับจำนวนข้อมูลทั้งหมด

2.4.2.2 ความแม่นยำ (Precision) สัดส่วนของการพยากรณ์คลาสเป้าหมายที่ถูกต้อง เทียบกับจำนวนข้อมูลทั้งหมดที่โมเดลพยากรณ์ว่าเป็นคลาสนั้น

2.4.2.3 ความครอบคลุม (Recall) สัดส่วนของการพยากรณ์คลาสเป้าหมายที่ถูกต้อง เทียบกับจำนวนข้อมูลที่เป็นคลาสนั้นจริงๆ

2.4.2.4 F1-Measure (F1-Score) ค่าเฉลี่ยฮาร์โมนิก (Harmonic Mean) ระหว่าง Precision และ Recall ใช้สำหรับประเมินผลในกรณีที่ชุดข้อมูลมีปริมาณคลาสไม่สมดุล (Imbalanced Data)

2.4.3 การวิเคราะห์เส้นโค้ง ROC (ROC Analysis)

การวิเคราะห์ ROC (Receiver Operating Characteristic) เป็นกราฟที่แสดงประสิทธิภาพของโมเดลในการจำแนกคลาส โดยพล็อตจุดระหว่างค่าอัตราการทายถูก (True Positive Rate) บนแกน Y และอัตราการทายผิด (False Positive Rate) บนแกน X ประสิทธิภาพของโมเดลจะพิจารณาจากพื้นที่ใต้พื้นที่กราฟ (Area Under the ROC Curve: AUC) หากค่า AUC เข้าใกล้ 1.0 แสดงว่าโมเดลมีประสิทธิภาพในการจำแนกประเภทได้สมบูรณ์แบบ แต่หากค่าเข้าใกล้ 0.5 แสดงว่าโมเดลไม่มีประสิทธิภาพ (Fawcett, 2006)

2.5 งานวิจัยที่เกี่ยวข้อง

วีรศักดิ์ และคณะ (2560) จากคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ได้ศึกษาวิจัยเรื่องการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล (Dam Water Volume Forecasting using Data Mining Techniques) เพื่อเปรียบเทียบเทคนิคที่เหมาะสมสำหรับการพยากรณ์ปริมาณน้ำรายเดือนของเขื่อนกิ่วลม จังหวัดลำปาง โดยวิเคราะห์จากข้อมูลย้อนหลัง 25 ปี (พ.ศ. 2535–2559) จำนวนรวม 9,300 รายการ ประกอบด้วยตัวแปรนำเข้าหลัก ได้แก่ ปริมาณน้ำไหลเข้า ปริมาณน้ำในอ่างเก็บน้ำ การระบายน้ำ และอัตราการระเหย ผู้วิจัยได้ดำเนินการทดลองผ่านโปรแกรม WEKA เพื่อเปรียบเทียบประสิทธิภาพของ 4 เทคนิค ได้แก่ การวิเคราะห์การถดถอย (Regression Analysis), โครงข่ายประสาทเทียม (ANN), โมเดลต้นไม้เอ็มไฟว์พี (Model Tree: M5P) และซัพพอร์ตเวกเตอร์แมชชีน (SVM) ผลการวิจัยพบว่า วิธีต้นไม้เอ็มไฟว์พี (M5P) มีประสิทธิภาพสูงสุดโดยให้ค่าความคลาดเคลื่อนต่ำที่สุด รองลงมาคือ SVM, Regression และ ANN ตามลำดับ

งานวิจัยนี้มีความสอดคล้องกับโครงงานในด้านการเลือกใช้โปรแกรม WEKA และเป็นการยืนยันประสิทธิภาพของอัลกอริทึมในตระกูลต้นไม้ตัดสินใจ (Decision Tree-based) ซึ่งรวมถึง M5P ว่ามีความเหมาะสมและให้ความแม่นยำสูงในการพยากรณ์ข้อมูลทางอุทกวิทยาที่มีความซับซ้อน จึงเป็นหลักฐานสำคัญที่สนับสนุนการเลือกใช้ Random Forest และ J48 ในการดำเนินโครงงานนี้

วีระยุทธ พิมพาภรณ์ และคณะ (2559) จากมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา ได้ศึกษาวิจัยเรื่องการพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำโดยใช้วิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล (Forecasting Reservoir Inflow using Time Series Analysis with Data Mining

Techniques) เพื่อพัฒนาแบบจำลองพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำปราณบุรี จังหวัดประจวบคีรีขันธ์ โดยทำการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูล 3 รูปแบบ ได้แก่ การถดถอยเชิงเส้น (Linear Regression), โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron: MLP) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression: SVR) ผลการทดสอบพบว่า SVR มีประสิทธิภาพสูงสุดโดยให้ค่าความคลาดเคลื่อนสัมพัทธ์เฉลี่ย (Mean Magnitude of Relative Error: MMRE) ต่ำที่สุดที่ร้อยละ 18.55 นอกจากนี้ ผู้วิจัยยังระบุว่าการกำหนดช่วงข้อมูลย้อนหลัง (Time Lag) ที่ 52–65 สัปดาห์ ให้ผลลัพธ์แม่นยำที่สุดเนื่องจากสามารถจับรูปแบบอิทธิพลของฤดูกาล (Seasonality) ได้ดี

งานวิจัยนี้สอดคล้องกับแนวทางของโครงงานในด้านการประยุกต์ใช้เทคนิคเหมืองข้อมูลร่วมกับการวิเคราะห์อนุกรมเวลา และสนับสนุนแนวคิดในการนำ ปัจจัยด้านเวลาและฤดูกาล มากำหนดเป็นตัวแปรสำคัญเพื่อให้แบบจำลองสามารถพยากรณ์สถานการณ์น้ำได้อย่างแม่นยำยิ่งขึ้น

วราวุธ วุฒิวณิชย์ และคณะ (2547) ภาควิชาวิศวกรรมชลประทาน มหาวิทยาลัยเกษตรศาสตร์ ได้ศึกษาวิจัยเรื่องการพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำโดยระบบโครงข่ายประสาทประดิษฐ์ (Artificial Neural Networks: ANNs) โดยประยุกต์ใช้ในการพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำรายวัน กรณีศึกษาอ่างเก็บน้ำลำตะคองและอ่างเก็บน้ำลำพระเพลิง ผู้วิจัยได้กำหนดตัวแปรนำเข้า (Input Variables) ที่ครอบคลุมปัจจัยทางอุทก-อุตุนิยมวิทยา ได้แก่ ปริมาณน้ำไหลเข้า ปริมาณน้ำท่า ปริมาณฝน อุณหภูมิ ความชื้นสัมพัทธ์ และความกดอากาศ ผลการศึกษาพบว่า การวิเคราะห์โดยใช้ข้อมูลอุทก-อุตุนิยมวิทยาร่วมกันหลายปัจจัย ให้ความแม่นยำในการพยากรณ์สูงกว่าการใช้เพียงข้อมูลฝนหรือน้ำท่าเพียงอย่างเดียว เนื่องจากโมเดลสามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนของปัจจัยแวดล้อมได้ดีกว่า

งานวิจัยนี้จึงเป็นพื้นฐานสำคัญในการสนับสนุนแนวทางการคัดเลือกตัวแปร (Feature Selection) ของโครงงาน ซึ่งแสดงให้เห็นว่าการนำปัจจัยนำเข้าที่หลากหลาย เช่น ปริมาณน้ำไหลเข้า การระบายน้ำ และปัจจัยด้านฤดูกาล มาวิเคราะห์ร่วมกัน จะช่วยเพิ่มประสิทธิภาพและความแม่นยำในการจำแนกสถานการณ์ความเสี่ยงด้านน้ำได้อย่างมีนัยสำคัญ

Fuad Ahmad Musleh (2024) จากคณะวิศวกรรมศาสตร์ มหาวิทยาลัยบาห์เรน (University of Bahrain) ได้ทำการศึกษาวิจัยเชิงเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องเพื่อการจำแนกคุณภาพน้ำดื่ม (A Comprehensive Comparative Study of Machine Learning Algorithms for Water Potability Classification) โดยเปรียบเทียบประสิทธิภาพของอัลกอริทึม 6 รูปแบบ ได้แก่ Random Forest, J48, Logistic Regression, Bagging Classifier, IBk และ AdaBoostM1 เพื่อจำแนกความเหมาะสมของน้ำในการบริโภค (Potability Classification) จากชุดข้อมูลจำนวน 3,276 ตัวอย่าง ผ่านโปรแกรม WEKA ผลการศึกษาพบว่า อัลกอริทึม Random Forest และ J48 มี

ประสิทธิภาพสูงสุดโดยให้ค่าความถูกต้อง (Accuracy) สูงถึง 99.3% เท่ากัน ในขณะที่ Logistic Regression ให้ค่าความถูกต้องอยู่ที่ 95.8% ซึ่งน้อยกว่าโมเดลในกลุ่มต้นไม้ตัดสินใจอย่างมีนัยสำคัญ

ผลการวิจัยนี้จึงเป็นหลักฐานเชิงประจักษ์ที่สำคัญซึ่งสนับสนุนการเลือกใช้โมเดลตระกูลต้นไม้ตัดสินใจ (Decision Tree-based) ในการแก้ปัญหาการจำแนกประเภทข้อมูลที่มีความซับซ้อน ซึ่งให้ผลลัพธ์ที่แม่นยำกว่าโมเดลพื้นฐานทั่วไป

Senem Güneş Şen (2025) จากมหาวิทยาลัย Kastamonu (Kastamonu University) ประเทศตุรกี ได้ศึกษาวิจัยเรื่องการพยากรณ์ระดับน้ำในอ่างเก็บน้ำด้วยการเรียนรู้ของเครื่อง (Machine Learning-Based Water Level Forecast in a Dam Reservoir) โดยใช้เขื่อน Karaçomak ในลุ่มน้ำ Kızılırmak เป็นกรณีศึกษา เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึม 4 รูปแบบ ได้แก่ Linear Regression, Decision Tree, Random Forest และ XGBoost ในการพยากรณ์ระดับน้ำรายวัน นอกจากนี้ผู้วิจัยใช้ชุดข้อมูลอุทก-อุตุนิยมวิทยาย้อนหลัง 17 ปี (พ.ศ. 2551–2567) รวมทั้งสิ้น 5,964 รายการ ผลการศึกษาพบว่าโมเดลในกลุ่มการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ได้แก่ Random Forest และ XGBoost มีประสิทธิภาพสูงสุด โดยให้ค่าสัมประสิทธิ์การตัดสินใจ (R^2) สูงถึง 0.983 และมีค่าความคลาดเคลื่อนต่ำที่สุด ในขณะที่ Linear Regression มีประสิทธิภาพด้อยกว่าเนื่องจากข้อจำกัดในการวิเคราะห์ความสัมพันธ์ที่ไม่มีลักษณะเป็นเชิงเส้น (Non-linear)

ผลการศึกษานี้จึงเป็นเครื่องยืนยันความเหมาะสมในการเลือกใช้ Random Forest สำหรับการจัดการข้อมูลระดับน้ำที่มีความซับซ้อนและความผันผวนสูงได้เป็นอย่างดี

Peter Lamovec และคณะ (2013) จากศูนย์วิจัยวิทยาศาสตร์แห่งสโลวีเนีย (Scientific Research Center of the Slovenian Academy of Sciences and Arts) ได้ศึกษาการตรวจวัดพื้นที่น้ำท่วมด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ในกรณีศึกษาเหตุการณ์น้ำท่วมบริเวณ Ljubljana Moor ปี 2010 โดยประยุกต์ใช้ข้อมูลภาพถ่ายดาวเทียม RapidEye ร่วมกับแบบจำลองความสูงเชิงเลข (Digital Terrain Model: DTM) ผู้วิจัยได้ใช้โปรแกรม WEKA ในการทดสอบประสิทธิภาพของอัลกอริทึมจำแนกประเภท ได้แก่ J48, Random Forest และ Support Vector Machine (SMO) ผลการศึกษาพบว่า อัลกอริทึม J48 และ Random Forest มีประสิทธิภาพสูงที่สุดในการจำแนกพื้นที่น้ำท่วมออกจากพื้นที่ปกติ โดยเฉพาะอย่างยิ่งเมื่อมีการนำข้อมูลความสูง (DTM) มาวิเคราะห์ร่วมด้วย ซึ่งช่วยเพิ่มความแม่นยำในการพยากรณ์อย่างมีนัยสำคัญ

งานวิจัยนี้จึงเป็นแนวทางสำคัญในการเลือกใช้เครื่องมือ WEKA และอัลกอริทึมกลุ่มโครงสร้างต้นไม้ (Tree-based) เพื่อประยุกต์ใช้ในการบริหารจัดการภัยพิบัติด้านน้ำที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ

บทที่ 3

วิธีการดำเนินงาน

โครงงานระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ด้วยเทคนิคการเรียนรู้ของเครื่อง ผู้จัดทำได้นำเสนอขั้นตอนและกระบวนการดำเนินงานของโครงงาน ประกอบด้วยหัวข้อดังนี้

3.1 การวิเคราะห์ระบบ

3.2 การออกแบบระบบ

3.3 การสร้างโมเดล

3.4 การพัฒนาระบบ

3.5 การทดสอบและประเมินผลระบบ

3.1 การวิเคราะห์ระบบ

การพัฒนาโครงงานการพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ด้วยเทคนิคการเรียนรู้ของเครื่อง ผู้จัดทำได้ทำการวิเคราะห์ระบบเพื่อทำความเข้าใจปัญหาและความจำเป็นของการพัฒนาระบบพยากรณ์ โดยมีรายละเอียดดังนี้

3.1.1 วิเคราะห์ปัญหา

การพยากรณ์ความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำถือเป็นส่วนสำคัญในการบริหารจัดการทรัพยากรน้ำของประเทศ เนื่องจากสถานการณ์น้ำเป็นปัจจัยหลักที่ส่งผลกระทบต่อภาคการเกษตร การอุปโภคบริโภค และความปลอดภัยของประชาชนในพื้นที่ลุ่มน้ำ จึงจำเป็นต้องพัฒนาโมเดลที่สามารถคาดการณ์แนวโน้มความเสี่ยงได้อย่างแม่นยำ ซึ่งประโยชน์สามารถแบ่งได้ดังนี้

3.1.1.1 เพื่อช่วยสนับสนุนการตัดสินใจในการระบายน้ำ

การรับรู้ระดับความเสี่ยงล่วงหน้า ไม่ว่าจะเป็นความเสี่ยงน้ำล้นหรือน้ำแห้ง ช่วยให้เจ้าหน้าที่หรือผู้เกี่ยวข้องสามารถปรับแผนการระบายน้ำได้อย่างเหมาะสม เช่น หากโมเดลพยากรณ์ว่ามีความเสี่ยงน้ำล้น ก็สามารถดำเนินการพร่องน้ำล่วงหน้าเพื่อรองรับปริมาณน้ำไหลเข้าในอนาคต ซึ่งช่วยลดผลกระทบต่อพื้นที่ท้ายเขื่อน

3.1.1.2 เพื่อการบริหารจัดการความเสี่ยงภัยแล้ง

ในกรณีที่โมเดลระบุความเสี่ยงน้ำแห้ง หน่วยงานที่เกี่ยวข้องสามารถวางแผนกักเก็บน้ำหรือจัดสรรน้ำเพื่อการเกษตรและการอุปโภคบริโภคได้อย่างรอบคอบมากยิ่งขึ้น ส่งผลให้สามารถลดผลกระทบทางเศรษฐกิจจากการขาดแคลนน้ำในช่วงฤดูแล้ง

3.1.1.3 เพื่อลดข้อจำกัดของเกณฑ์การเตือนภัยแบบเดิม

การแจ้งเตือนแบบเดิมมักใช้เกณฑ์ร้อยละความจุคงที่ ซึ่งอาจไม่สามารถสะท้อนแนวโน้มของสถานการณ์น้ำในอนาคตได้อย่างแท้จริง การประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง ช่วยให้สามารถวิเคราะห์รูปแบบความสัมพันธ์ที่ซับซ้อนของข้อมูลย้อนหลัง ทำส่งผลให้การพยากรณ์มีความยืดหยุ่นและแม่นยำมากกว่าการพิจารณาจากสถานะปัจจุบันเพียงอย่างเดียว

3.1.1.4 เพื่อสนับสนุนข้อมูลสำหรับประชาชนและเกษตรกร

การนำเสนอผลการพยากรณ์ผ่านระบบแดชบอร์ด ช่วยให้ประชาชนและเกษตรกรในพื้นที่ได้รับข้อมูลสถานการณ์ความเสี่ยงที่เข้าใจง่าย ส่งเสริมให้สามารถเตรียมความพร้อมและรับมือกับภัยพิบัติทางธรรมชาติได้อย่างเหมาะสมและทันท่วงที

3.1.2 วิเคราะห์ข้อมูล

ข้อมูลที่ใช้ในโครงงานนี้เป็นข้อมูลสถานการณ์น้ำรายวันของอ่างเก็บน้ำขนาดใหญ่จากกรมชลประทาน โดยอยู่ในรูปแบบไฟล์ JSON ซึ่งผู้จัดทำได้นำข้อมูลมาวิเคราะห์เบื้องต้นเพื่อทำความเข้าใจโครงสร้าง ความครบถ้วน และความเหมาะสมของข้อมูล ก่อนนำไปพัฒนาแบบจำลองการจำแนกระดับความเสี่ยงของสถานการณ์น้ำในอ่างเก็บน้ำ โดยมีรายละเอียดดังนี้

3.1.2.1 โครงสร้างข้อมูล เป็นข้อมูลรายวันประกอบด้วยตัวแปรสำคัญ (Attributes) ได้แก่

1) วันที่ (Date) อยู่ในรูปแบบ YYYY-MM-DD ใช้สำหรับการจัดเรียงข้อมูลตามเวลาและตรวจสอบความต่อเนื่องของข้อมูล
2) ชื่ออ่างเก็บน้ำ (Dam Name) เช่น เขื่อนแม่กวงอุดมธารา
3) ปริมาณน้ำกักเก็บ (Volume) หน่วยล้านลูกบาศก์เมตร
4) ร้อยละความจุ (%Capacity) แสดงสัดส่วนปริมาณน้ำเทียบกับความจุเก็บกัก
5) ปริมาณน้ำไหลลงอ่าง (Inflow) หน่วยล้านลูกบาศก์เมตรต่อวัน
6) ปริมาณน้ำระบาย (Outflow) หน่วยล้านลูกบาศก์เมตรต่อวัน

3.1.2.2 การกำหนดฉลากความเสี่ยง (Risk Level Labels)

เพื่อให้สามารถนำข้อมูลเชิงตัวเลขไปใช้ในงานจำแนกประเภท (Classification) ผู้จัดทำได้กำหนดฉลากความเสี่ยงจากตัวแปร %Capacity โดยแบ่งออกเป็น 3 ระดับ ได้แก่

1) น้อยกว่า 30% หมายถึง เสี่ยงน้ำแห้ง (Dry Risk)
2) 30% – 80% หมายถึง ระดับปกติ (Medium Risk)
3) มากกว่าหรือเท่ากับ 80% หมายถึง เสี่ยงน้ำล้น (Flood Risk)

โดยฉลากดังกล่าวถูกใช้เป็นตัวแปรเป้าหมาย (Target Variable) สำหรับการสร้างแบบจำลอง

3.1.2.3 ช่วงเวลาข้อมูลและการแบ่งชุดข้อมูล

ผู้จัดทำใช้ชุดข้อมูลย้อนหลังรวมระยะเวลา 2 ปี (พ.ศ. 2567 – 2568) เพื่อให้ครอบคลุมรูปแบบสถานการณ์น้ำในช่วงเวลาที่หลากหลาย โดยในการฝึกสอนและทดสอบแบบจำลอง ผู้จัดทำได้นำข้อมูลทั้งสองปีมารวมกันและใช้วิธีการแบ่งชุดข้อมูลแบบกำหนดสัดส่วนร้อยละ (Percentage Split) ในอัตราส่วน 80:20 แบ่งเป็นชุดข้อมูลฝึกสอน (Training Set) ร้อยละ 80 และชุดข้อมูลทดสอบ (Testing Set) ร้อยละ 20 ตามลำดับ เพื่อประเมินประสิทธิภาพของแบบจำลอง

3.1.2.3 การตรวจสอบค่าที่หายไปและค่าผิดปกติ

จากการตรวจสอบข้อมูลรายวันของอ่างเก็บน้ำขนาดใหญ่ทั้ง 35 แห่ง พบว่าข้อมูลส่วนใหญ่มีความครบถ้วนและมีความต่อเนื่องในแต่ละวัน อย่างไรก็ตาม เพื่อให้มั่นใจว่าข้อมูลมีคุณภาพเพียงพอสำหรับการสร้างแบบจำลอง ผู้จัดทำได้ทำการตรวจสอบคุณภาพข้อมูลเพิ่มเติม โดยพิจารณาการเกิดค่าว่าง (Missing Values) และค่าที่ผิดปกติ (Outliers) ในตัวแปรสำคัญ เช่น Volume, %Capacity, Inflow และ Outflow ในกรณีที่พบวันที่ขาดข้อมูล ผู้จัดทำจะดำเนินการเติมเต็มข้อมูล (Imputation) ด้วยวิธีที่เหมาะสม เช่น การใช้ค่าจากวันก่อนหน้า (Forward Fill) หรือการใช้ค่าเฉลี่ยของช่วงเวลาใกล้เคียง ทั้งนี้เพื่อรักษาความต่อเนื่องของข้อมูลและลดผลกระทบต่อประสิทธิภาพของแบบจำลอง

3.1.2.4 การวิเคราะห์ลักษณะเชิงเวลา

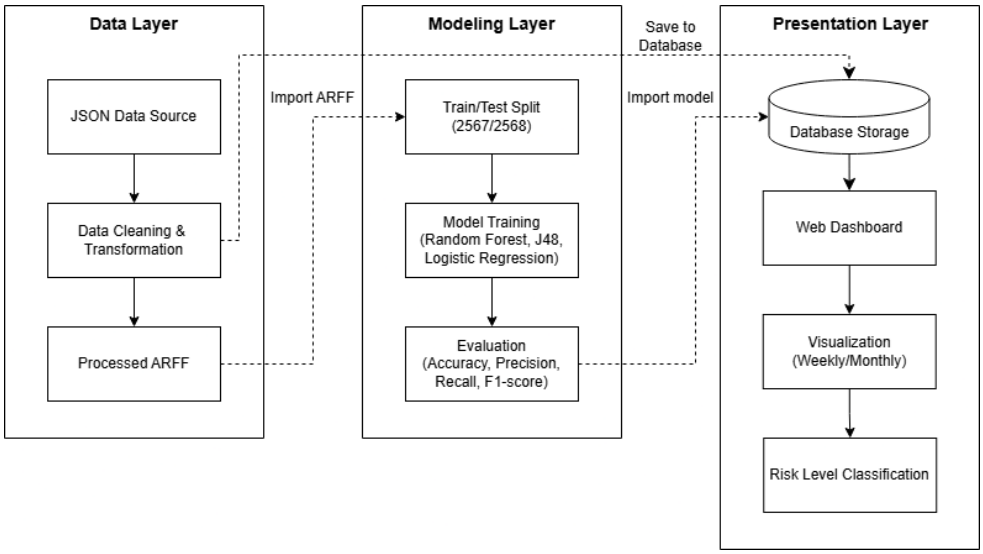
ข้อมูลปริมาณน้ำรายวันมีลักษณะเป็นข้อมูลอนุกรมเวลา (Time-series) ซึ่งมีความต่อเนื่อง โดยระดับน้ำในแต่ละวันมักได้รับอิทธิพลจากข้อมูลของวันก่อนหน้า นอกจากนี้ยังพบรูปแบบเชิงฤดูกาลที่ชัดเจน เช่น ช่วงเดือนพฤษภาคมถึงตุลาคมซึ่งเป็นฤดูฝนมีแนวโน้มระดับน้ำสูงกว่าช่วงฤดูแล้ง จากลักษณะดังกล่าว ผู้จัดทำจึงนำข้อมูลเชิงเวลาไปใช้ในการสร้างฟีเจอร์เพิ่มเติม (Feature Engineering) เพื่อช่วยให้แบบจำลองสามารถจำแนกระดับความเสี่ยงได้แม่นยำขึ้น เช่น ตัวแปรบ่งชี้ฤดูฝน (IsRainySeason) หรือคุณลักษณะจากช่วงเวลา (เช่น เดือน หรือไตรมาส)

3.2 การออกแบบระบบ

ผู้จัดทำได้ทำการออกแบบระบบการพยากรณ์ความเสี่ยงสถานการณ์น้ำในอ่างเก็บน้ำขนาดใหญ่ โดยมีรายละเอียดดังนี้

3.2.1 การออกแบบสถาปัตยกรรมระบบ

ผู้จัดทำได้ออกแบบสถาปัตยกรรมของระบบการพยากรณ์ความเสี่ยงสถานการณ์น้ำ แสดงภาพรวมการทำงานและความสัมพันธ์ขององค์ประกอบภายในระบบ โดยภาพรวมของสถาปัตยกรรมระบบแสดงดังภาพที่ 3.1



ภาพที่ 3.1 สถาปัตยกรรมระบบ

จากภาพที่ 3.1 สถาปัตยกรรมของระบบพยากรณ์ความเสี่ยงสถานการณ์น้ำถูกออกแบบแบ่งออกเป็น 3 เลเยอร์หลัก เพื่อให้ง่ายต่อการพัฒนาและบำรุงรักษา โดยมีรายละเอียดดังนี้

3.2.1.1 ส่วนการจัดการข้อมูล (Data Layer)

ทำหน้าที่ในการจัดเตรียมข้อมูลดิบให้พร้อมสำหรับการประมวลผล ประกอบด้วยขั้นตอนการดำเนินงานดังนี้

ขั้นตอนที่ 1 ดึงข้อมูลปริมาณน้ำรายวันจากอินเทอร์เฟซโปรแกรมประยุกต์ (API) ของกรมชลประทาน ซึ่งข้อมูลจะถูกรับส่งในรูปแบบโครงสร้างไฟล์เจสัน (JavaScript Object Notation: JSON) เพื่อความสะดวกรวดเร็วในการส่งผ่านข้อมูลบนเครือข่าย

ขั้นตอนที่ 2 ดำเนินการตรวจสอบคุณภาพข้อมูลเพื่อจัดการกับค่าที่สูญหาย (Missing Values) หรือค่าที่ผิดปกติ (Outliers) จากนั้นจึงแปลงรูปแบบข้อมูล (Data

Transformation) ให้เป็นไฟล์นามสกุล .arff (Attribute-Relation File Format) ซึ่งเป็นรูปแบบที่ระบุความสัมพันธ์และชนิดของข้อมูลอย่างชัดเจน เพื่อนำไปใช้ในการวิเคราะห์ด้วยโปรแกรม WEKA ในลำดับถัดไป

3.2.1.2 ส่วนประมวลผลโมเดล (Modeling Layer)

เป็นส่วนหลักในการสร้างและทดสอบประสิทธิภาพของโมเดลการเรียนรู้ของเครื่อง ประกอบด้วยขั้นตอนดังนี้

ขั้นตอนที่ 1 ดำเนินการนำชุดข้อมูลที่ผ่านการจัดเตรียมแล้ว (Processed ARFF) ซึ่งรวบรวมข้อมูลทั้งหมดเข้าด้วยกัน มาแบ่งออกเป็น 2 ส่วนด้วยเทคนิค Percentage Split ในอัตราส่วน 80:20 ได้แก่ ข้อมูลสำหรับฝึกสอน (Training Set) สัดส่วนร้อยละ 80 และข้อมูลสำหรับทดสอบ (Testing Set) สัดส่วนร้อยละ 20 เพื่อประเมินความสามารถในการพยากรณ์ข้อมูลในอนาคต

ขั้นตอนที่ 2 นำข้อมูลเข้าสู่กระบวนการฝึกสอนด้วยอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) จำนวน 3 รูปแบบ ได้แก่ เรนดอมฟอเรสต์ (Random Forest), ต้นไม้ตัดสินใจ (J48) และการถดถอยโลจิสติก (Logistic Regression) เพื่อเปรียบเทียบหาความเหมาะสมของแต่ละอัลกอริทึมต่อชุดข้อมูลสถานการณ์น้ำ

ขั้นตอนที่ 3 ตรวจสอบคุณภาพของแบบจำลองด้วยตัวชี้วัดประสิทธิภาพ (Performance Metrics) สำหรับงานจำแนกประเภท (Classification) ประกอบด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครอบคลุม (Recall) และค่า F1-score เพื่อใช้ในการคัดเลือกแบบจำลองที่มีประสิทธิภาพสูงสุดเป็นโมเดลหลักของระบบ

3.2.1.3 ส่วนการนำเสนอข้อมูล (Presentation Layer)

ทำหน้าที่ในการนำเสนอผลการพยากรณ์และข้อมูลสถิติแก่ผู้ใช้งานผ่านระบบเว็บแอปพลิเคชัน เพื่อสนับสนุนการตัดสินใจบริหารจัดการน้ำ โดยมีรายละเอียดดังนี้

การจัดเก็บข้อมูล (Database Storage) ทำหน้าที่จัดเก็บข้อมูลปริมาณน้ำจริงที่ดึงมาจากแหล่งข้อมูล และผลการพยากรณ์ที่ได้จากโมเดลลงในระบบฐานข้อมูล MySQL เพื่อเตรียมพร้อมสำหรับการดึงไปแสดงผลบนหน้าจอ

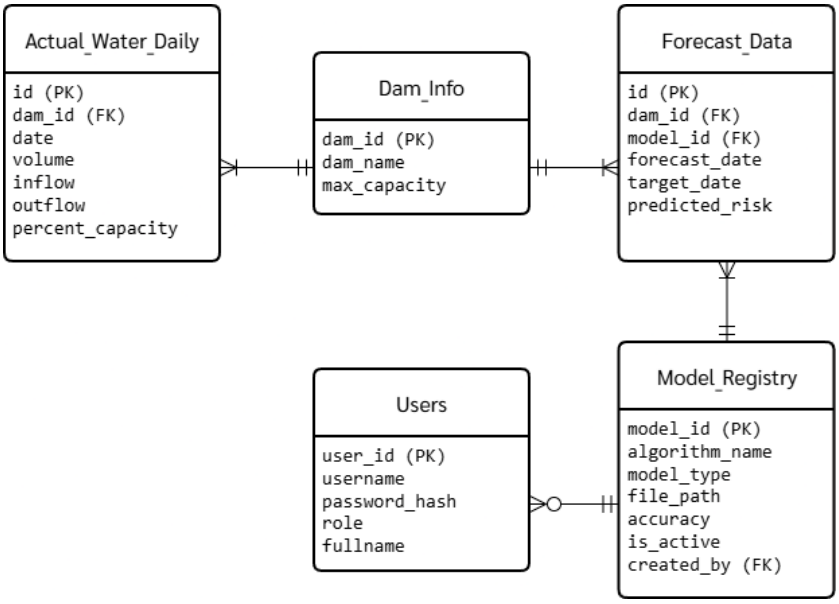
แดชบอร์ด (Web Dashboard) เป็นส่วนอินเทอร์เฟซสำหรับผู้ใช้งาน (User Interface) ที่พัฒนาด้วยภาษา Python โดยทำหน้าที่รับแบบจำลองที่ผ่านการคัดเลือก (Best Model) เข้ามาเพื่อใช้ในการพยากรณ์สถานการณ์น้ำจากข้อมูลล่าสุด

การแสดงภาพข้อมูล (Visualization) นำเสนอข้อมูลสถิติและสถานการณ์น้ำในรูปแบบกราฟ (Charts/Graphs) ที่เข้าใจง่าย ทั้งในรูปแบบรายสัปดาห์ (Weekly) และรายเดือน (Monthly) เพื่อให้เห็นแนวโน้มการเปลี่ยนแปลงของระดับน้ำ

การจำแนกระดับความเสี่ยง (Risk Level Classification) แสดงผลลัพธ์สุดท้ายจากการพยากรณ์โดยระบุระดับความเสี่ยงของอ่างเก็บน้ำเป็น 3 ระดับ ได้แก่ เสี่ยงน้ำแห้ง ปกติ และเสี่ยงน้ำล้น ตามเกณฑ์ที่กำหนดไว้ในขอบเขตโครงงาน

3.2.2 การออกแบบฐานข้อมูล

ระบบมีการจัดเก็บข้อมูลในรูปแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) โดยออกแบบโครงสร้างตารางข้อมูลและกำหนดความสัมพันธ์ระหว่างเอนทิตี (Entity Relationship) เพื่อให้รองรับการแสดงผลข้อมูลย้อนหลังและการพยากรณ์ล่วงหน้าบนแดชบอร์ด ดังภาพที่ 3.2



ภาพที่ 3.2 แผนภาพความสัมพันธ์ของฐานข้อมูล

จากภาพที่ 3.2 ระบบจัดเก็บข้อมูลในรูปแบบฐานข้อมูลเชิงสัมพันธ์ โดยแบ่งเป็น 4 ตารางหลัก ได้แก่ ตาราง Dam_Info ตาราง Actual_Water_Daily ตาราง Forecast_Data และตาราง Model_Registry เพื่อรองรับการแสดงผลข้อมูลย้อนหลัง และการจัดเก็บผลการพยากรณ์สำหรับการแสดงบนแดชบอร์ด ซึ่งความสัมพันธ์ระหว่างตารางมีดังนี้

3.2.2.1 ความสัมพันธ์ระหว่าง `Dam_Info` และ `Actual_Water_Daily` (1:N)

คีย์หลัก (PK): Dam_Info.dam_id

คีย์นอก (FK): Actual_Water_Daily.dam_id

ความสัมพันธ์: อ่างเก็บน้ำ 1 แห่ง (Dam_Info) สามารถจัดเก็บข้อมูลสถานการณ์น้ำรายวัน (Actual_Water_Daily) ได้หลายรายการตามช่วงเวลาที่ตรวจวัด

3.2.2.2 ความสัมพันธ์ระหว่าง Dam_Info และ Forecast_Data (1:N)

คีย์หลัก (PK): Dam_Info.dam_id

คีย์นอก (FK): Forecast_Data.dam_id

คำอธิบาย: อ่างเก็บน้ำ 1 แห่ง (Dam_Info) สามารถมีผลการพยากรณ์ความเสี่ยง (Forecast_Data) ได้หลายรายการตามวันที่คาดการณ์หรือตามรอบการประมวลผล

3.2.2.3 ความสัมพันธ์ระหว่าง Model_Registry และ Forecast_Data (1:N)

คีย์หลัก (PK): Model_Registry.model_id

คีย์นอก (FK): Forecast_Data.model_id

คำอธิบาย: แบบจำลองหรืออัลกอริทึม 1 ประเภท (Model_Registry) สามารถใช้สร้างผลการพยากรณ์ (Forecast_Data) ได้หลายรายการ

3.2.2.4 ความสัมพันธ์ระหว่าง Users และ Model_Registry (1:N)

คีย์หลัก (PK): Users.user_id

คีย์นอก (FK): Model_Registry.created_by

คำอธิบาย: ผู้ใช้งาน (Users) 1 บัญชี สามารถนำเข้าหรือจัดการข้อมูลแบบจำลอง (Model_Registry) ได้ตั้งแต่ 0 ถึงหลายแบบจำลอง

3.2.3 การออกแบบกระบวนการพัฒนาแบบจำลอง

ประกอบด้วยขั้นตอนการดำเนินงานเชิงเทคนิคเพื่อเปลี่ยนข้อมูลดิบให้เป็นแบบจำลองที่มีประสิทธิภาพ ดังนี้

3.2.3.1 การนำเข้าข้อมูล (Data Acquisition) ดำเนินการดึงข้อมูลปริมาณน้ำและตัวแปรที่เกี่ยวข้องจาก API ของกรมชลประทานในรูปแบบไฟล์ JSON เพื่อใช้เป็นข้อมูลชุดตั้งต้น (Raw Data) สำหรับกระบวนการเตรียมข้อมูล

3.2.3.2 การเตรียมและแปลงข้อมูล (Data Preprocessing) ดำเนินการตรวจสอบความสมบูรณ์ของข้อมูลและจัดการค่าที่สูญหาย (Missing Values) จากนั้นจึงทำการกำหนดฉลากข้อมูล (Data Labeling) เพื่อระบุระดับความเสี่ยงตามเกณฑ์เปอร์เซ็นต์ความจุที่กำหนด และทำการแปลงรูปแบบข้อมูลให้เป็นไฟล์ ARFF เพื่อให้รองรับการประมวลผลด้วยโปรแกรม WEKA

3.2.3.3 การแบ่งชุดข้อมูล (Data Splitting) นำชุดข้อมูลที่เตรียมเสร็จสิ้นซึ่งรวบรวมข้อมูลทั้ง 2 ปี มาแบ่งออกเป็น 2 ส่วนด้วยเทคนิค Percentage Split (Train/Test Split 80:20) ได้แก่ ข้อมูลสำหรับเป็นชุดฝึกสอน (Training Set) สัดส่วนร้อยละ 80 และข้อมูลสำหรับเป็นชุดทดสอบ (Testing Set) สัดส่วนร้อยละ 20 เพื่อใช้ประเมินความแม่นยำของแบบจำลองกับข้อมูลที่ไม่เคยพบมาก่อน

3.2.3.4 การสร้างและฝึกสอนแบบจำลอง (Model Training) นำชุดข้อมูลฝึกสอนเข้าสู่กระบวนการเรียนรู้ของเครื่องผ่านอัลกอริทึมการจำแนกประเภท (Classification) จำนวน 3 รูปแบบที่คัดเลือกไว้ เพื่อสร้างแบบจำลองในการพยากรณ์ระดับความเสี่ยงของสถานการณ์น้ำ

3.2.3.5 การประเมินผลและคัดเลือกแบบจำลอง (Evaluation & Selection) ทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบและวัดผลด้วยตัวชี้วัด ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครอบคลุม (Recall) และค่าเอฟวัน (F1-score) เพื่อเปรียบเทียบประสิทธิภาพและคัดเลือกแบบจำลองที่มีความเหมาะสมที่สุดไปเชื่อมต่อกับระบบฐานข้อมูลและแดชบอร์ด

3.2.4 การออกแบบแบบจำลองจำแนกระดับความเสี่ยง

ในการพัฒนาแบบจำลองเพื่อพยากรณ์ระดับความเสี่ยงสถานการณ์น้ำ ผู้จัดทำได้กำหนดรายละเอียดคุณลักษณะของข้อมูลและอัลกอริทึมที่ใช้ ดังนี้

3.2.4.1 ข้อมูลนำเข้าและตัวแปรเป้าหมาย (Input Features and Target Class) เพื่อให้แบบจำลองสามารถเรียนรู้พฤติกรรมของน้ำได้อย่างมีประสิทธิภาพ จึงได้กำหนดตัวแปรที่ใช้ในการประมวลผลดังนี้

ตัวแปรนำเข้า (Input Features) ประกอบด้วย ปริมาณน้ำไหลเข้า (Inflow) ปริมาณน้ำระบาย (Outflow) และร้อยละของความจุอ่างเก็บน้ำ (%Capacity) ซึ่งเป็นปัจจัยหลักในการตัดสินระดับน้ำ

ตัวแปรเป้าหมาย (Target Class) คือระดับความเสี่ยง 3 ระดับ ได้แก่ เสี่ยงน้ำแห้ง ปกติ เสี่ยงน้ำล้น ที่ผ่านกระบวนการทำ Data Labeling ตามเกณฑ์เปอร์เซ็นต์ความจุที่ระบุในขอบเขตโครงงาน

3.2.4.2 รายละเอียดอัลกอริทึมที่ใช้เปรียบเทียบ ผู้จัดทำเลือกใช้ 3 อัลกอริทึมที่มีจุดเด่นแตกต่างกัน เพื่อค้นหาแบบจำลองที่เหมาะสมที่สุดกับลักษณะข้อมูลชุดนี้

เรนดอมฟอเรสต์ (Random Forest) ใช้เทคนิคการสร้างต้นไม้ตัดสินใจจำนวนมาก (Ensemble Learning) แล้วนำผลลัพธ์มาโหวตร่วมกัน จุดเด่นคือมีความแม่นยำสูงและลดปัญหาการเรียนรู้เกิน (Overfitting) ได้ดีเมื่อเทียบกับต้นไม้ตัดสินใจแบบเดี่ยว (Witten et al., 2016)

อัลกอริทึมต้นไม้ตัดสินใจ (J48) เป็นการนำอัลกอริทึม C4.5 มาประยุกต์ใช้ใน WEKA จุดเด่นคือผลลัพธ์ที่ได้อยู่ในรูปแบบกฎที่มนุษย์สามารถตีความได้ง่าย (Interpretability) ช่วยให้เข้าใจเหตุผลเบื้องหลังการจำแนกความเสี่ยงน้ำได้อย่างชัดเจน (Witten et al., 2016)

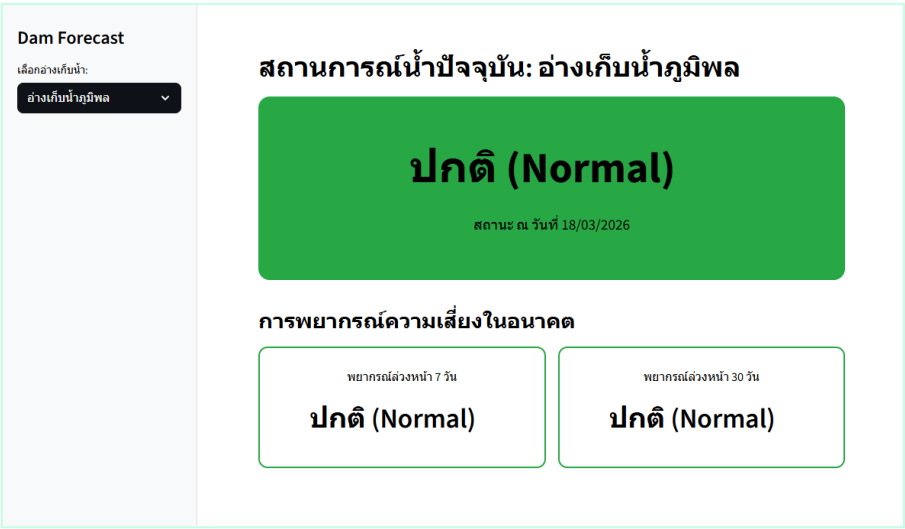
การถดถอยโลจิสติก (Logistic Regression) ใช้เป็นโมเดลพื้นฐาน (Baseline Model) เพื่อวิเคราะห์ความสัมพันธ์เชิงความน่าจะเป็นระหว่างปัจจัยนำเข้าและระดับความเสี่ยง เหมาะสำหรับงานจำแนกประเภทข้อมูลที่มีความสัมพันธ์เชิงเส้น (Witten et al., 2016)

3.2.5 การออกแบบส่วนติดต่อผู้ใช้

การออกแบบส่วนติดต่อผู้ใช้ของระบบพยากรณ์ความเสี่ยงน้ำในอ่างเก็บน้ำขนาดใหญ่ พัฒนาขึ้นในรูปแบบโปรแกรมประยุกต์บนเว็บ (Web Application) โดยมุ่งเน้นการนำเสนอข้อมูลด้วยภาพ (Data Visualization) ที่มีความชัดเจน ลดความซับซ้อนในการใช้งาน เพื่อสนับสนุนการตัดสินใจของผู้บริหารจัดการน้ำ ทั้งนี้ ระบบได้ถูกออกแบบให้มีการแบ่งส่วนติดต่อผู้ใช้ออกเป็น 2 มุมมองหลัก เพื่อให้เหมาะสมกับความต้องการของกลุ่มเป้าหมายที่แตกต่างกัน ดังนี้

3.2.5.1 มุมมองสำหรับประชาชนทั่วไป

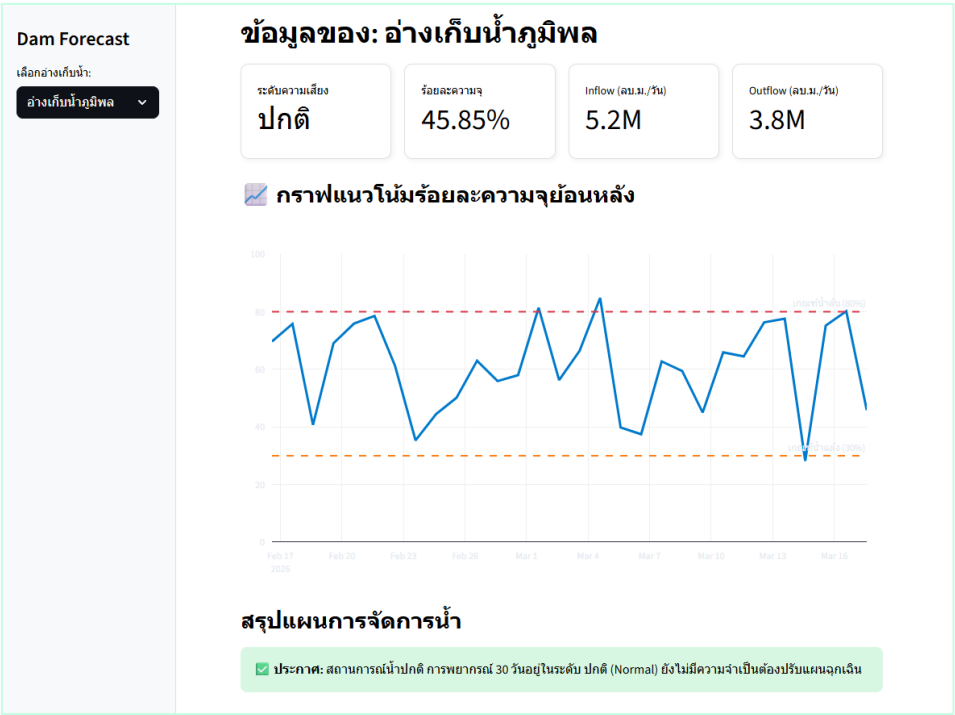
หน้าจอถูกออกแบบมาเพื่อให้ประชาชนสามารถรับรู้สถานการณ์น้ำได้อย่างรวดเร็วและเข้าใจง่าย โดยแสดงผลระดับความเสี่ยงปัจจุบันในรูปแบบกล่องสีขนาดใหญ่ที่ชัดเจน เช่น สีเขียวสำหรับสถานะปกติ พร้อมทั้งแสดงผลการพยากรณ์ความเสี่ยงในอนาคตล่วงหน้า 7 วัน และ 30 วัน โดยไม่แสดงข้อมูลตัวเลขเชิงลึกหรือกราฟทางสถิติที่ซับซ้อน เพื่อลดความสับสนและให้ประชาชนสามารถติดตามสถานการณ์ได้อย่างตรงไปตรงมา ดังภาพที่ 3.3



ภาพที่ 3.3 การออกแบบส่วนติดต่อผู้ใช้หน้าจอแสดงผลสำหรับประชาชนทั่วไป

3.2.5.2 มุมมองสำหรับเจ้าหน้าที่

หน้าจอสำหรับเจ้าหน้าที่ผู้ปฏิบัติงานถูกออกแบบมาเพื่อแสดงข้อมูลเชิงลึกทางสถิติสำหรับการวิเคราะห์และสนับสนุนการตัดสินใจ โดยจะแสดงข้อมูลสถานการณ์ปัจจุบันที่ครบถ้วน ได้แก่ ระดับความเสี่ยง, ร้อยละความจุ, ปริมาณน้ำไหลเข้า (Inflow), และปริมาณน้ำระบาย (Outflow) รวมไปถึงการแสดงกราฟแนวโน้มร้อยละความจุย้อนหลังที่มาพร้อมเส้นเกณฑ์ความเสี่ยง (เสี่ยงน้ำแห้ง 30% และ เสี่ยงน้ำล้น 80%) และส่วนสรุปแผนการจัดการน้ำเพื่อเป็นข้อเสนอแนะในการปฏิบัติงาน ดังภาพที่ 3.4



ภาพที่ 3.4 การออกแบบส่วนติดต่อผู้ใช้หน้าจอแสดงผลสำหรับเจ้าหน้าที่

3.2.6 รายละเอียดองค์ประกอบและการทำงานของหน้าจอ

หน้าจอแสดงผลหลักของระบบ (Dashboard) ได้รับการออกแบบโดยแบ่งการแสดงผลตามมุมมองของผู้ใช้งาน ได้แก่ มุมมองสำหรับประชาชนทั่วไปและมุมมองสำหรับเจ้าหน้าที่ เพื่อให้การนำเสนอข้อมูลมีความเหมาะสมกับวัตถุประสงค์การใช้งาน โดยมีรายละเอียดองค์ประกอบและการทำงานในแต่ละส่วนดังนี้

3.2.6.1 การเลือกอ่างเก็บน้ำ

ส่วนสำหรับเลือกอ่างเก็บน้ำขนาดใหญ่ที่ต้องการตรวจสอบสถานการณ์น้ำผ่านรายการแบบเลื่อนลง (Dropdown Menu) บริเวณแถบเมนูด้านข้าง (Sidebar) เมื่อมีการเปลี่ยนอ่างเก็บน้ำ ข้อมูลทั้งหมดบนแดชบอร์ดจะถูกปรับเปลี่ยนให้สอดคล้องกับอ่างเก็บน้ำที่เลือก เพื่อให้การแสดงผลมีความเฉพาะเจาะจง

3.2.6.2 การแสดงสถานการณ์น้ำปัจจุบัน

ระบบจะแสดงข้อมูลสรุปของสถานการณ์น้ำปัจจุบัน เพื่อให้ผู้ใช้งานสามารถรับรู้สถานการณ์ได้ทันที โดยมีการแสดงผลแตกต่างกันตามมุมมอง ดังนี้

1) มุมมองสำหรับประชาชนทั่วไป

แสดงระดับความเสี่ยงสถานการณ์น้ำปัจจุบัน (Risk Level) อย่างชัดเจนในรูปแบบกล่องสีขนาดใหญ่ (เช่น สีเขียวสำหรับสถานการณ์ปกติ) เพื่อให้เข้าใจง่ายโดยไม่ต้องตีความตัวเลข

2) มุมมองสำหรับเจ้าหน้าที่

แสดงข้อมูลในรูปแบบกล่องข้อมูลประกอบด้วยข้อมูลสำคัญ เช่น ระดับความเสี่ยงสถานการณ์น้ำ, ร้อยละความจุปัจจุบันของอ่างเก็บน้ำ (%Capacity), ปริมาณน้ำไหลเข้า (Inflow) และปริมาณน้ำระบายออก (Outflow) เพื่อสนับสนุนการวิเคราะห์สมดุลน้ำ

3.2.6.3 การแสดงกราฟแนวโน้มปริมาณน้ำย้อนหลังสำหรับเจ้าหน้าที่

ส่วนแสดงผลกราฟเส้นแสดงแนวโน้มร้อยละความจุของอ่างเก็บน้ำย้อนหลังตามลำดับเวลา ซึ่งเป็นข้อมูลเชิงเวลาภายในกราฟประกอบด้วย

1) แกนแนวนอนแสดงช่วงเวลาในหน่วยวัน

2) แกนแนวตั้งแสดงร้อยละความจุของอ่างเก็บน้ำ (%Capacity)

3) เส้นกราฟแสดงแนวโน้มการเปลี่ยนแปลงของระดับน้ำย้อนหลัง

นอกจากนี้ ยังมีการแสดงเส้นเกณฑ์ระดับความเสี่ยงเพื่อช่วยในการตีความข้อมูล ได้แก่ เส้นเกณฑ์ร้อยละ 30 สำหรับสถานการณ์เสี่ยงน้ำแห้ง (เส้นประล่าง) และเส้นเกณฑ์ร้อยละ 80 สำหรับสถานการณ์เสี่ยงน้ำล้น (เส้นประบน) กราฟดังกล่าวช่วยให้เจ้าหน้าที่สามารถเข้าใจแนวโน้มการเปลี่ยนแปลงของระดับน้ำในช่วงที่ผ่านมาได้อย่างชัดเจน

3.2.6.4 การแสดงผลการพยากรณ์ความเสี่ยงสถานการณ์น้ำสำหรับประชาชนทั่วไป

เพื่อให้สอดคล้องกับวัตถุประสงค์ของระบบในการพยากรณ์ความเสี่ยง แดชบอร์ดได้แสดงผลการพยากรณ์ระดับความเสี่ยงของสถานการณ์น้ำในอนาคต โดยแบ่งออกเป็น 2 ช่วงเวลา ได้แก่

1) การพยากรณ์ระยะสั้น 7 วัน

2) การพยากรณ์ระยะยาว 30 วัน

ผลการพยากรณ์จะแสดงในรูปแบบข้อความที่สรุประดับความเสี่ยง เช่น ระดับปกติ หรือระดับเสี่ยงน้ำล้น เพื่อให้ผู้ใช้งานสามารถรับรู้สถานการณ์ล่วงหน้าและใช้ประกอบการวางแผนบริหารจัดการน้ำ

3.2.6.5 ส่วนสรุปแผนการจัดการน้ำเพื่อสนับสนุนการตัดสินใจสำหรับเจ้าหน้าที่

นอกจากการแสดงข้อมูลเชิงตัวเลขและกราฟ แดชบอร์ดฝั่งเจ้าหน้าที่มีส่วนสรุปผลการวิเคราะห์ในรูปแบบข้อความ (Text Summary) บริเวณด้านล่างของหน้าจอ เพื่ออธิบายภาพรวมของสถานการณ์น้ำและแนวโน้มการพยากรณ์ ช่วยสนับสนุนการตัดสินใจของผู้ปฏิบัติงานในการปรับแผนฉุกเฉินหรือการบริหารจัดการน้ำได้อย่างเหมาะสม

3.3 การสร้างโมเดล

ในขั้นตอนนี้ ผู้จัดทำได้ดำเนินการสร้างแบบจำลองจำแนกระดับความเสี่ยงของสถานการณ์น้ำ โดยใช้โปรแกรม WEKA โดยเริ่มจากการเตรียมข้อมูลจริงจาก API ของกรมชลประทาน จัดรูปแบบข้อมูลให้พร้อมใช้งาน และทำการฝึกสอนโมเดลด้วยอัลกอริทึมที่กำหนด จากนั้นประเมินผลและเลือกโมเดลที่เหมาะสมที่สุดเพื่อใช้งานในระบบแดชบอร์ด โดยมีขั้นตอนการปฏิบัติดังนี้

3.3.1 การเตรียมและแปลงข้อมูล (Data Preprocessing & Transformation)

3.3.1.1 การดึงข้อมูลจาก API

ผู้จัดทำทำการดึงข้อมูลสถานการณ์น้ำรายวันย้อนหลังจาก API ของกรมชลประทานในรูปแบบ JSON จากนั้นจัดเก็บข้อมูลที่ได้ในรูปแบบตาราง (Tabular Format) เพื่อให้สามารถนำไปประมวลผลในขั้นตอนถัดไปได้

3.3.1.2 การคัดเลือกตัวแปรและสร้างฟีเจอร์

ผู้จัดทำเลือกใช้ตัวแปรนำเข้า (Input Features) ที่เกี่ยวข้องกับการเปลี่ยนแปลงของสถานการณ์น้ำ ได้แก่ ปริมาณน้ำไหลเข้า (Inflow) ปริมาณน้ำระบาย (Outflow) ปริมาณน้ำกักเก็บ (Volume) วันที่ (Date)

จากนั้นดำเนินการแปลงข้อมูลวันที่ (Date) ให้เป็นตัวแปรเชิงเวลา เช่น เดือน (Month) และสร้างตัวแปรฤดูกาล (IsRainySeason) เพื่อให้แบบจำลองสามารถเรียนรู้รูปแบบการเปลี่ยนแปลงตามฤดูกาลได้ชัดเจนยิ่งขึ้น

3.3.1.3 การกำหนดฉลากความเสี่ยง (Data Labeling)

ผู้จัดทำกำหนดตัวแปรเป้าหมาย (Target Class) จากค่าเปอร์เซ็นต์ความจุของอ่างเก็บน้ำ (%Capacity) ตามเกณฑ์ที่กำหนดไว้ในขอบเขตโครงงาน โดยแบ่งเป็น 3 ระดับ ได้แก่ Dry Risk, Medium Risk และ Flood Risk

3.3.1.4 การสร้างชุดข้อมูลสำหรับการคาดการณ์ล่วงหน้า (Prediction Horizon)

เพื่อรองรับการพยากรณ์ระดับความเสี่ยงล่วงหน้า ผู้จัดทำสร้างชุดข้อมูล 2 รูปแบบ ได้แก่ ชุดข้อมูลสำหรับคาดการณ์ล่วงหน้า 7 วัน โดยจับคู่ข้อมูลนำเข้าของวันปัจจุบันกับระดับความเสี่ยงที่เกิดขึ้นจริงในอีก 7 วันถัดไป ชุดข้อมูลสำหรับคาดการณ์ล่วงหน้า 30 วัน โดยจับคู่ข้อมูลนำเข้าของวันปัจจุบันกับระดับความเสี่ยงที่เกิดขึ้นจริงในอีก 30 วันถัดไป

3.3.1.5 การแปลงข้อมูลเป็นไฟล์ ARFF

เมื่อได้ข้อมูลที่ผ่านการเตรียมแล้ว ผู้จัดทำแปลงข้อมูลให้อยู่ในรูปแบบไฟล์ ARFF และกำหนดชนิดของแอตทริบิวต์ (Attribute Type) ให้ถูกต้อง เพื่อใช้เป็นข้อมูลนำเข้าของโปรแกรม WEKA

3.3.2 การฝึกสอนและคัดเลือกแบบจำลอง (Model Training & Selection)

3.3.2.1 การนำเข้าข้อมูลใน WEKA

ผู้จัดทำเปิดโปรแกรม WEKA และนำเข้าไฟล์ ARFF ผ่านเมนู Explorer ⟶ Preprocess เพื่อตรวจสอบจำนวนข้อมูล ตัวแปร และความถูกต้องของชนิดข้อมูลก่อนเริ่มฝึกสอน

3.3.2.2 การกำหนดรูปแบบการทดสอบ (Test Option)

ผู้จัดทำเลือกวิธีการประเมินผลแบบแบ่งชุดข้อมูลตามสัดส่วนร้อยละ (Percentage Split) โดยกำหนดข้อมูลร้อยละ 80 เป็นชุดฝึกสอน (Training Set) และข้อมูลร้อยละ 20 เป็นชุดทดสอบ (Testing Set) เพื่อให้การประเมินผลสอดคล้องกับสถานการณ์การใช้งานจริงและครอบคลุมความหลากหลายของข้อมูลทั้งชุด

3.3.2.3 การฝึกสอนโมเดลด้วยอัลกอริทึมที่กำหนด

ผู้จัดทำทำการฝึกสอนโมเดลด้วยอัลกอริทึมทั้ง 3 รูปแบบ ได้แก่ Logistic Regression (Baseline) Random Forest และ J48 Decision Tree โดยทำการทดลองแต่ละโมเดลภายใต้ชุดข้อมูลเดียวกัน เพื่อให้สามารถเปรียบเทียบผลได้

3.3.2.4 การประเมินผลและเลือกโมเดลที่ดีที่สุด

ผู้จัดทำประเมินผลการจำแนกด้วยตัวชี้วัด Accuracy Precision Recall และ F1-score จากผลลัพธ์ของ WEKA และเลือกโมเดลที่มีประสิทธิภาพดีที่สุดเพื่อใช้งานจริง

3.3.2.5 การบันทึกโมเดลและนำไปใช้งาน

เมื่อได้โมเดลที่เหมาะสมที่สุด ผู้จัดทำทำการบันทึกโมเดลในรูปแบบไฟล์ `.model` และจัดเก็บข้อมูลรายละเอียดของโมเดลไว้ในฐานข้อมูล เพื่อให้ระบบสามารถเรียกใช้งานโมเดลในการพยากรณ์ระดับความเสี่ยงและแสดงผลบนแดชบอร์ดได้ในขั้นตอนถัดไป

3.4 การพัฒนาระบบ

3.4.1 เครื่องมือที่ใช้ในการพัฒนา

3.4.1.1 ภาษา Python

สำหรับพัฒนาระบบส่วนหลังบ้าน (Backend) ระบบการจัดการข้อมูล และประมวลผลของระบบ

3.4.1.2 Streamlit Framework

สำหรับพัฒนาส่วนติดต่อผู้ใช้ (Frontend) และสร้างแดชบอร์ดแสดงผลข้อมูลแบบโต้ตอบ (Interactive Dashboard)

3.4.1.3 Python-WEKA-Wrapper

ไลบรารีสำหรับเชื่อมต่อระหว่าง Python Web Application กับแกนประมวลผลของ WEKA เพื่อโหลดไฟล์โมเดลและทำการพยากรณ์ข้อมูลใหม่

3.4.1.4 ฐานข้อมูล MySQL

ระบบจัดการฐานข้อมูลสำหรับจัดเก็บข้อมูลปริมาณน้ำรายวัน ผลลัพธ์การพยากรณ์ และที่อยู่โมเดล

3.4.2 การพัฒนาส่วนติดต่อผู้ใช้ (User Interface)

พัฒนาหน้าจอ Dashboard ตามการออกแบบเพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูลได้ง่าย ประกอบด้วยส่วนสำคัญดังนี้

3.4.2.1 ส่วนแสดงผลสถานะปัจจุบัน

แสดงค่าสำคัญ ได้แก่ ปริมาณน้ำไหลเข้า (Inflow), ปริมาณน้ำระบายออก (Outflow), ร้อยละความจุ (%Capacity) และระดับความเสี่ยงปัจจุบัน

3.4.2.2 ส่วนแสดงกราฟ

แสดงกราฟเส้นอนุกรมเวลาเปรียบเทียบข้อมูลจริง (Actual Data) เทียบกับเส้นเกณฑ์ความเสี่ยง เพื่อให้ผู้ใช้งานเห็นภาพสถานการณ์ล่วงหน้า

3.4.2.3 ส่วนเลือกโมเดล

พัฒนาฟังก์ชัน Dropdown Menu ที่เชื่อมต่อกับตาราง Model_Registry ในฐานข้อมูล เพื่อดึงรายชื่อโมเดลที่พร้อมใช้งานมาแสดงผลให้ผู้เลือกใช้งานได้อย่างยืดหยุ่น โดยไม่ต้องแก้ไขซอร์สโค้ดเมื่อมีการเพิ่มโมเดลใหม่

3.5 การทดสอบและประเมินผลระบบ

ผู้จัดทำได้ดำเนินการทดสอบและประเมินผลเพื่อยืนยันความถูกต้องและความสมบูรณ์ของระบบ โดยแบ่งออกเป็น 2 ส่วนหลัก คือ การประเมินประสิทธิภาพของโมเดล และ การทดสอบระบบแดชบอร์ด

3.5.1 การประเมินประสิทธิภาพโมเดล

ใช้วิธีการแบ่งชุดข้อมูลแบบกำหนดสัดส่วนร้อยละ (Percentage Split) ในอัตราส่วน 80:20 โดยใช้ข้อมูลร้อยละ 80 เป็นชุดฝึกสอน (Training Set) และข้อมูลร้อยละ 20 เป็นชุดทดสอบ (Testing Set) และการประเมินผลใช้ตัวชี้วัดประสิทธิภาพสำหรับงานจำแนกประเภท (Classification Metrics) ได้แก่ค่าความถูกต้อง (Accuracy), ความแม่นยำ (Precision), ความครอบคลุม (Recall), และ F1-score

3.5.2 การทดสอบระบบแดชบอร์ด

ทดสอบการทำงานของระบบโดยรวมเพื่อตรวจสอบความถูกต้องของฟังก์ชันการทำงาน ได้แก่

3.5.2.1 การจัดการข้อมูล

ทดสอบความถูกต้องของการดึงข้อมูลจาก API กรมชลประทานและการบันทึกข้อมูลลงฐานข้อมูล MySQL

3.5.2.2 การเชื่อมต่อโมเดล

ทดสอบความถูกต้องของการเลือกสลับโมเดลผ่านหน้าเว็บไซต์ และตรวจสอบผลลัพธ์การพยากรณ์ว่าตรงกันกับการประมวลผลในโปรแกรม WEKA

3.5.2.3 การแสดงผลและการแจ้งเตือน

ทดสอบการแสดงผลกราฟ การระบุระดับความเสี่ยง (Dry, Normal, Flood) และการตอบสนองของส่วนติดต่อผู้ใช้ (User Interface) ว่าถูกต้องและเป็นไปตามความต้องการ

บรรณานุกรม

กรมชลประทาน. (2568). ระบบฐานข้อมูลน้ำในอ่างเก็บน้ำ. [ออนไลน์]. สืบค้นจาก: https://app.rid.go.th/reservoir/rsvmiddle (วันที่สืบค้น 12 ธันวาคม 2568)

สถาบันสารสนเทศทรัพยากรน้ำ. (2554). บันทึกเหตุการณ์มหาอุทกภัย ปี 2554. [ออนไลน์]. สืบค้นจาก: https://tiwrmdev.hii.or.th/current/menu.html (วันที่สืบค้น 12 ธันวาคม 2568)

สถาบันสารสนเทศทรัพยากรน้ำ. (2564). บันทึกเหตุการณ์มหาอุทกภัย ปี 2564. [ออนไลน์].]. สืบค้นจาก: https://www.thaiwater.net/uploads/contents/current/YearlyReport2021/flood_event.html (วันที่สืบค้น 12 ธันวาคม 2568)

สถาบันสารสนเทศทรัพยากรน้ำ. (2565). บันทึกเหตุการณ์สถานการณ์น้ำไทย ปี 2565. [ออนไลน์]. สืบค้นจาก: https://www.thaiwater.net/uploads/contents/current/YearlyReport2022/flood_event.html (วันที่สืบค้น 12 ธันวาคม 2568)

สถาบันสารสนเทศทรัพยากรน้ำ. (2562). บันทึกเหตุการณ์สถานการณ์น้ำไทย ปี 2562. [ออนไลน์]. สืบค้นจาก: https://www.thaiwater.net/uploads/contents/current/2020/drought2019/summary.html (วันที่สืบค้น 12 ธันวาคม 2568)

สถาบันสารสนเทศทรัพยากรน้ำ. (2563). บันทึกเหตุการณ์สถานการณ์น้ำไทย 2563. [ออนไลน์]. สืบค้นจาก: https://www.thaiwater.net/uploads/contents/current/YearlyReport2020/drought.html (วันที่สืบค้น 12 ธันวาคม 2568)

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. Water, 10(11), 1536. [ออนไลน์]. สืบค้นจาก: https://doi.org/10.3390/w10111536 (วันที่สืบค้น 13 ธันวาคม 2568)

Zhang, Y., Pan, D., Griensven, J. V., Yang, S. X., & Gharabaghi, B. (2023). Intelligent flood forecasting and warning: A survey. Intelligence & Robotics, 3(2), 190–212. [ออนไลน์]. สืบค้นจาก: https://doi.org/10.20517/ir.2023.12 (วันที่สืบค้น 13 ธันวาคม 2568)

กรมทรัพยากรน้ำ. (ม.ป.ป., หน้า 26-27). *เกณฑ์วิกฤติน้ำ และแนวทางการบริหารจัดการ นอกเขตชลประทาน อ่างเก็บน้ำ ในพื้นที่ ภาคกลางและภาคตะวันออก*. [ออนไลน์]. สืบค้นจาก:

https://mekhala.dwr.go.th/download/6755?wpdmdl=6755&refresh=69a3ea2bf0cb21772349995 (วันที่สืบค้น 1 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2568). การเข้าสู่ฤดูหนาวของประเทศไทย พ.ศ. 2568. [ออนไลน์]. สืบค้นจาก: https://www.tmd.go.th/warning-and-events/special-events/การเข้าสู่ฤดูหนาวของประเทศไทย-พ-ศ-2568 (วันที่สืบค้น 16 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2568). การเริ่มต้นฤดูฝนของประเทศไทย พ.ศ. 2568. [ออนไลน์]. สืบค้นจาก: https://www.tmd.go.th/warning-and-events/special-events/ประกาศเริ่มต้นฤดูร้อน-ป-พ-ศ-2569 (วันที่สืบค้น 16 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2567). ประกาศกรมอุตุนิยมวิทยา เรื่อง การเข้าสู่ฤดูหนาวของประเทศไทย พ.ศ. 2567. [ออนไลน์]. สืบค้นจาก: https://www.opsmoac.go.th/news-preview-462891791955 (วันที่สืบค้น 16 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2567). ประกาศเริ่มต้นฤดูฝน ปี พ.ศ. 2567. [ออนไลน์]. สืบค้นจาก: https://www.tmd.go.th/warning-and-events/special-events/ประกาศเริ่มต้นฤดูฝน-ป-พ-ศ-2567 (วันที่สืบค้น 16 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2567). ประกาศเริ่มต้นฤดูร้อน ปี พ.ศ. 2567. [ออนไลน์]. สืบค้นจาก: https://www.tmd.go.th/warning-and-events/special-events/ประกาศเริ่มต้นฤดูร้อน-ป-พ-ศ-2567 (วันที่สืบค้น 16 มีนาคม 2569)

กรมอุตุนิยมวิทยา. (2568). ประกาศเริ่มต้นฤดูร้อน ปี พ.ศ. 2568. [ออนไลน์]. สืบค้นจาก: https://tmd.go.th/warning-weather/content/ประกาศเริ่มต้นฤดูร้อน-ป-พ-ศ-2568 (วันที่สืบค้น 16 มีนาคม 2569)

เอกสิทธิ์ พัชรวงศ์ศักดา. (ม.ป.ป.). การคัดเลือกfeature (feature selection) ด้วยวิธี Information Gain. [ออนไลน์]. สืบค้นจาก: https://th.linkedin.com/pulse/การคัดเลือก-feature-selection-ด้วยวิธี-information-gain-pacharawongsakda (วันที่สืบค้น 19 กุมภาพันธ์ 2569)

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3), 379-423. สืบค้นจาก: https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf (วันที่สืบค้น 19 กุมภาพันธ์ 2569)

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Elsevier 3rd Edition. [ออนไลน์]. สืบค้นจาก: https://tjzhifei.github.io/links/DM3.pdf (วันที่สืบค้น 25 มกราคม 2569)

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324 (วันที่สืบค้น 3 มีนาคม 2569)

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons. https://books.google.co.th/books?id=bRoxQBIZRd4C (วันที่สืบค้น 3 มีนาคม 2569)

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251 (วันที่สืบค้น 3 มีนาคม 2569)

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010 (วันที่สืบค้น 3 มีนาคม 2569)

วีรศักดิ์ ฟองเงิน, วรปภา อารีราษฎร์ และเผด็จ พรหมสาขา ณ สกลนคร. (2560). การพยากรณ์ปริมาณน้ำในเขื่อน โดยใช้เทคนิคเหมืองข้อมูล. วารสารวิชาการ การจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม, 4(1), 27–33. สืบค้น 25 มกราคม 2568. จาก https://ph02.tci-thaijo.org/index.php/itm-journal/article/view/115233 (วันที่สืบค้น 25 มกราคม 2569)

วีระยุทธ พิมพาภรณ์, ปรียาภรณ์ พูลทอง และ บุษกร แก้ววิเชียร. (2559). การพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำโดยใช้วิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล. ใน *การประชุมวิชาการระดับชาติ มหาวิทยาลัยรังสิต ประจำปี 2559 (RSU National Research Conference 2016)* (หน้า 625-633). ปทุมธานี: มหาวิทยาลัยรังสิต. [ออนไลน์]. สืบค้นจาก: https://rsucon.rsu.ac.th/files/proceedings/nation2016/G2-19.pdf (วันที่สืบค้น 25 มกราคม 2569)

วราวุธ วุฒิวณิชย์, สันติ ทองพำนัก, นิมิต เฉิดฉันท์พิพัฒน์, อารียา ฤทธิมา และ นัทพันธุ์ เกษมพันธุ์. (2547). การพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำโดยระบบโครงข่ายประสาทประดิษฐ์. เรื่องเต็มการประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 42: สาขาวิศวกรรมศาสตร์. *การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 42*. (หน้า 24-31). สำนักงานกองทุนสนับสนุนงานวิจัย. [ออนไลน์]. สืบค้นจาก: https://kukr.lib.ku.ac.th/kukr_es/index.php?/BKN/search_detail/result/9044 (วันที่สืบค้น 25 มกราคม 2569)

Ahmad Musleh, F. (2024). A Comprehensive Comparative Study of Machine Learning Algorithms for Water Potability Classification. International Journal of

Computing and Digital Systems, 15(1), 1189–1200. [ออนไลน์]. สืบค้นจาก: https://doi.org/10.12785/ijcds/150184 (วันที่สืบค้น 25 มกราคม 2569)

Lamovec, P., Matjaž, M., & Oštir, K. (2013). Detection of Flooded Areas using Machine Learning Techniques: Case Study of the Ljubljana Moor Floods in 2010. Disaster Advances, 6, 4–11. [ออนไลน์]. สืบค้นจาก: https://www.researchgate.net/publication/254558083 (วันที่สืบค้น 25 มกราคม 2569)

Şen, S. G. (2025). Machine Learning-Based Water Level Forecast in a Dam Reservoir: A Case Study of Karaçomak Dam in the Kızılırmak Basin, Türkiye. Sustainability, 17(18). [ออนไลน์]. สืบค้นจาก: https://doi.org/10.3390/su17188378 (วันที่สืบค้น 25 มกราคม 2569

ภาคผนวก ก
โครงสร้างข้อมูลสถานการณ์น้ำจาก API กรมชลประทาน

1. ภาพรวมของ API (API Overview)

โครงงานนี้ดำเนินการเชื่อมต่อและดึงข้อมูลสถานการณ์น้ำรายวันของอ่างเก็บน้ำขนาดใหญ่ ทั้ง 35 แห่งทั่วประเทศ จากศูนย์ปฏิบัติการน้ำอัจฉริยะ (SWOC) กรมชลประทาน ผ่านอินเทอร์เฟซโปรแกรมประยุกต์ (API) โดยมีรายละเอียดการเข้าถึงข้อมูลดังนี้

1.1 Endpoint (ข้อมูลปัจจุบัน): https://app.rid.go.th/reservoir/api/dam/public

1.2 Endpoint (ข้อมูลย้อนหลังตามวันที่): https://app.rid.go.th/reservoir/api/dam/public/{YYYY-MM-DD}

1.3 รูปแบบข้อมูลที่ตอบกลับ (Response Format): JSON (JavaScript Object Notation)

1.4 เมธอด (Method): GET

2. ตัวอย่างโครงสร้างข้อมูล JSON (JSON Response Example)

เมื่อระบบส่งคำร้องขอ (HTTP GET Request) ไปยัง API ระบบจะได้รับการตอบกลับมาในรูปแบบโครงสร้างข้อมูล JSON โดยมีตัวอย่างโครงสร้างดังต่อไปนี้


```
{
  "date": "2024-03-18",
  "total": 35,
  "data": [
    {
      "region": "ภาคเหนือ",
      "dam": [
        {
          "id": "100104",
          "name": "เขื่อนแม่กวงอุดมธารา",
          "owner": "กรมชลประทาน",
          "capacity": 295.00,
          "storage": 263.00,
          "active_storage": 249.00,
          "dead_storage": 14.00,
          "volume": 86.48,
          "percent_storage": 32.88,
          "inflow": 0.148,
```

"outflow": 0.045
}
]
}
]
}

3. พจนานุกรมข้อมูล (Data Dictionary)

คำอธิบายรายละเอียดของคุณลักษณะข้อมูล (Attributes) ที่ปรากฏในโครงสร้าง JSON ซึ่งถูกนำมาใช้สำหรับการวิเคราะห์และสร้างฟีเจอร์ของแบบจำลอง มีรายละเอียดดังตารางที่ ก.1

ตารางที่ ก.1 อธิบายรายละเอียดแอตทริบิวต์ข้อมูลจาก API ของกรมชลประทาน

ชื่อฟิลด์ (Field Name)	ชนิดข้อมูล (Data Type)	คำอธิบาย (Description)
date	String	วันที่ (รูปแบบ YYYY-MM-DD)
total	Integer	จำนวนอ่างเก็บน้ำขนาดใหญ่ทั้งหมด
region	String	ภาคของอ่างเก็บน้ำขนาดใหญ่
id	String	รหัสอ่างเก็บน้ำขนาดใหญ่
name	String	ชื่ออ่างเก็บน้ำขนาดใหญ่
owner	String	เจ้าของอ่างเก็บน้ำ
capacity	Float	ปริมาณน้ำสูงสุด (ล้าน ลบ.ม.)
storage	Float	ปริมาณน้ำเก็บกัก (ล้าน ลบ.ม.)
active_storage	Float	ปริมาณน้ำใช้การ (ล้าน ลบ.ม.)
dead_storage	Float	ปริมาณใช้การต่ำสุด (ล้าน ลบ.ม.)
volume	Float	ปริมาณน้ำในเขื่อน (ล้าน ลบ.ม.)
percent_storage	Float	เปอร์เซนต์ปริมาณน้ำต่อปริมาณกักเก็บ
inflow	Float	ปริมาณน้ำไหลเข้าเขื่อน
outflow	Float	ปริมาณน้ำระบาย