การค้นหาปัจจัยเพื่อสร้างโมเดลสำหรับพยากรณ์การควบคุมประตูระบายน้ำ

นิธินันท์ มาตา[1] แสงดาว นพพิทักษ์[2] ศิโรรัตน์ กุลวงศ์[3] และเปรม อิงคเวชชากุล[4]

[1, 2, 3, 4] สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ บุรีรัมย์

Emails: mata0711@gmail.com, ddaow@hotmail.com, s_pooh@hotmail.com, csbongga@gmail.com

บทคัดย่อ

ปัจจุบันภัยที่เกิดจากน้ำเป็นสิ่งสำคัญที่ต้องการการควบคุมที่เหมาะสม เนื่องจากเป็นสาเหตุที่ก่อให้เกิดความเสียหายต่อชีวิตและทรัพย์สิน ทำให้นักวิจัยหลายกลุ่มพยายามสร้างโมเดลการพยากรณ์ที่เกี่ยวข้องกับน้ำ เพื่อช่วยในการตัดสินใจบริหารจัดการน้ำตามลุ่มน้ำหรืออ่างเก็บน้ำ งานวิจัยนี้จึงนำเสนอการค้นหาปัจจัยในการพยากรณ์การควบคุมประตูระบายน้ำของอ่างเก็บน้ำห้วยจรเข้มาก จังหวัดบุรีรัมย์ โดยเปรียบประสิทธิภาพโมเดลพยากรณ์การระบายน้ำจากเทคนิคการจำแนกประเภทข้อมูล พบว่าโมเดลที่มีประสิทธิภาพความแม่นยำในการจำแนกประเภทข้อมูลมากที่สุดคือ Random Forest และเมื่อใช้ร่วมกับการลดคุณลักษณะของข้อมูลแบบ Symmetric uncertainty ทำให้ประสิทธิภาพในการพยากรณ์เพิ่มขึ้น และพบว่าปัจจัยที่มีความสำคัญต่อการพยากรณ์น้ำท่วมมากที่สุดคือ ปริมาณน้ำท่าสะสม และปริมาณน้ำฝนตามลำดับ

คำสำคัญ-- การจำแนกประเภทข้อมูล; การลดคุณลักษณะของข้อมูล

1. บทนำ

อ่างเก็บน้ำห้วยจรเข้มาก เป็นอ่างเก็บน้ำชลประทาน สร้างขึ้นในปี พ.ศ. 2505 อ่างเก็บน้ำทั้งหมดเป็นพื้นที่ลุ่มมีการรับน้ำจากพื้นที่ข้างเคียง ในฤดูแล้งน้ำจะแห้งมีส่วนที่ลึกที่สุดเพียง 2 เมตร ในฤดูฝนมีน้ำลึกที่สุดประมาณ 5-7 เมตร เป็นแหล่งประมงท้องถิ่น แหล่งน้ำเพื่อการเกษตร และแหล่งผลิตน้ำประปาเพื่อใช้ในการอุปโภคบริโภคของจังหวัดบุรีรัมย์ ส่งผลให้การควบคุมระดับน้ำ (การกักเก็บ – การระบาย) เป็นสิ่งที่สำคัญเพื่อควบคุมภาวะน้ำท่วมหรือน้ำแล้งให้เหมาะสมกับธรรมชาติและลดการสูญเสียต่อประชากรในเขตพื้นที่บริการ ดังนั้นจึงต้องมีการพัฒนาโมเดลควบคุมระดับน้ำโดยผ่านการควบคุมประตูระบายน้ำ เพื่อเพิ่มประสิทธิภาพการบริหารจัดการน้ำและช่วยสนับสนุนการตัดสินใจในการระบายน้ำให้กับเกษตรกร หรือประชาชนทั่วไปเมื่อถึงฤดูแล้งหรือระบายน้ำออกเมื่อถึงฤดูฝน เพื่อบรรเทาความเสียหายที่จะเกิดขึ้นได้

ภัยจากน้ำเกิดขึ้นทั่วทุกมุมโลก เป็นสาเหตุที่ก่อให้เกิดความเสียหายต่อหลายภาคส่วน จึงทำให้นักวิจัยหลายกลุ่มพยายามสร้างโมเดลการพยากรณ์ที่เกี่ยวกับน้ำ ไม่ว่าจะเป็นการพยากรณ์เกี่ยวกับน้ำท่วมที่มีวิธีการจำนวนมาก เช่น การวิเคราะห์แนวโน้ม การวิเคราะห์การถดถอย การทำนาย ขั้นตอนวิธีทางพันธุกรรม และอัลกอริทึมทางปัญญาประดิษฐ์ การพยากรณ์การไหลของน้ำ และการนำปัจจัยที่เกี่ยวข้องกับน้ำมาช่วยสร้างโมเดลการพยากรณ์ให้กับเครื่องมือทางอุทกศาสตร์วิทยา [1] เหล่านี้สำหรับช่วยในการตัดสินใจบริหารจัดการน้ำตามลุ่มน้ำหรืออ่างเก็บน้ำ เป็นต้น ในด้านของการพยากรณ์น้ำท่วมได้มีการศึกษาและนำเสนอวิธีการหลายรูปแบบ ไม่เพียงแต่จะเป็นการนำอัลกอริทึมที่เกี่ยวกับการนำวิธีการทำเหมืองข้อมูลมาสร้างโมเดลการพยากรณ์หรือปรับปรุงอัลกอริทึมที่นิยมใช้เพื่อให้มีประสิทธิภาพในการทำนายมากขึ้น หรือแม้แต่การค้นหาปัจจัยต่างๆ ที่เกี่ยวข้องกับการพยากรณ์เพื่อให้ได้ประสิทธิภาพที่ดีกว่าเดิม ส่วนใหญ่ของงานวิจัยที่ผ่านมาที่เกี่ยวข้องกับการพยากรณ์โดยใช้การทำเหมืองข้อมูลจะใช้การจำแนกประเภทข้อมูลด้วยเทคนิคที่แตกต่างกัน แต่ในงานวิจัยครั้งนี้ได้นำเสนอการเปรียบเทียบประสิทธิภาพของการพยากรณ์จากเทคนิคการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิค ซึ่งเป็นเทคนิคที่นิยมอย่างแพร่หลายในปัจจุบัน ในแต่ละกลุ่มมีการสร้างโมเดลการจำแนกประเภทข้อมูลด้วยรูปแบบที่แตกต่างกัน ได้แก่ ได้แก่ กลุ่มของ Bayes สร้างโมเดลโดยอาศัยการคำนวนความน่าจะเป็นของข้อมูล (Naïve Bayes, Bayes Network) กลุ่มของ Function สร้างโมเดลโดยอาศัยการคำนวนทางคณิตศาสตร์และโมเดลที่ได้จะเป็นรูปแบบของสมการ (LivSVM, MLP, SMO) กลุ่มของ Lazy จะแตกต่างจากกลุ่มการจำแนกประเภทข้อมูลกลุ่มอื่น โดยไม่มีการสร้างโมเดลไว้ก่อน ใช้ข้อมูลเรียนรู้เพื่อจำแนกประเภทของข้อมูลใหม่ (IBk, KStar) กลุ่มของ Rules เป็นการสร้างโมเดลในรูปแบบของกฎ (DecisionTable, JRIP, OneR, PART) และกลุ่มของ Tree สร้างโมเดลในรูปแบบของต้นไม้ตัดสินใจ (J48, Random Forest, RandomTree, REFTree) เพื่อหาโมเดลที่มีประสิทธิภาพมากที่สุดจากกลุ่มเดียวกันก่อน หลังจากนั้นนำมาเปรียบเทียบประสิทธิภาพกับกลุ่มอื่นเพื่อหาโมเดลการจำแนกประเภทที่มีประสิทธิภาพมากที่สุด และเพิ่มประสิทธิภาพความแม่นยำของการจำแนกแนกประเภทโดยการคัดเลือกคุณลักษณะของข้อมูลพร้อมกับค้นหาปัจจัยที่สำคัญต่อการพยากรณ์

ในการพยากรณ์การควบคุมประตูระบายน้ำเกี่ยวข้องกับปัจจัย (attribute) ที่หลากหลาย ที่สำคัญปัจจัยเหล่านั้นมักจะมีความสัมพันธ์กับปัจจัยอื่น ๆ ที่เกี่ยวข้อง ซึ่งบางปัจจัยที่นำมาพยากรณ์การควบคุมประตูระบายน้ำไม่ได้มีความสำคัญต่อโมเดล หรือมีความสำคัญน้อยและเมื่อนำมาสร้างโมเดลพยากรณ์ทำให้ประสิทธิภาพความแม่นยำในการพยากรณ์ลดลง ดังนั้นงานวิจัยนี้จึงมุ่งเน้นการค้นหาปัจจัยที่สำคัญที่

สำคัญต่อการพยากรณ์การระบาย โดยมีวัตถุประสงค์หลักดังนี้ 1) เพื่อตรวจสอบความสัมพันธ์ระหว่างปัจจัยที่แตกต่างกันและระบุตัวปัจจัยที่เกี่ยวข้องมากที่สุดสำหรับการพยากรณ์ 2) เพื่อลดมิติของข้อมูลในการพยากรณ์การควบคุมประตูระบายน้ำ 3) เพื่อสร้างโมเดลการพยากรณ์การควบคุมประตูระบายน้ำที่มีประสิทธิภาพแม่นยำ ผลลัพธ์ของการพยากรณ์โดยใช้โมเดลที่ได้จากงานวิจัยนี้ต้องสามารถพยากรณ์ปริมาณน้ำที่ต้องระบายออกจากประตูระบายน้ำซึ่งอยู่ในช่วง 0.1 – 3.5 ลูกบาศก์เมตรต่อวินาที และปิดการระบายน้ำเมื่อปริมาณน้ำไม่เหมาะสมที่จะระบายออกไป

2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1. การจำแนกประเภทข้อมูล

เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ เพื่อแสดงให้เห็นความแตกต่างระหว่างกลุ่มของข้อมูลได้ และเพื่อทำนายว่าข้อมูลนี้ควรจัดอยู่ในกลุ่มใด ซึ่งโมเดลที่ใช้จำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้จะขึ้นอยู่กับการวิเคราะห์กลุ่มของข้อมูลทดลอง (Training data) โดยนำ Training data มาสอนให้ระบบเรียนรู้ว่ามีข้อมูลใดอยู่ในกลุ่มเดียวกันบ้าง ผลลัพธ์ที่ได้จากการเรียนรู้ คือ โมเดลจัดประเภทข้อมูล (Classifier model) [2, 3] มีเทคนิคที่นิยมใช้งานได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) [4], โครงข่ายประสาทเทียม (Neural Network) [5], การตัดสินใจโดยใช้กฎ (Rule-Based Decision Support System) [6] ,นาอีพเบย์ (Naïve Bayes) [7] และ K-Nearest Neighbor (KNN) [8]

2.2. การคัดเลือกคุณลักษณะของข้อมูล

เป็นการกรองคุณลักษณะของข้อมูลเพื่อลดจำนวนคุณลักษณะที่ไม่เกี่ยวข้องออกให้เหลือเฉพาะคุณลักษณะที่ความสัมพันธ์กันเท่านั้น ส่วนใหญ่จะให้ค่าความถูกต้องในการพยากรณ์สูงขึ้น [9, 10]

2.3. Information Gain (IG)

เป็นการประเมินค่าเพื่อใช้ในการแบ่งข้อมูลด้วยการคำนวณค่า Gain สำหรับแต่ละมิติข้อมูล ถ้ามิติข้อมูลใดมีค่า Gain สูงสุด จะถูกเลือกให้เป็นกลุ่มย่อยที่มีอำนาจจำแนก [3]

2.4. Gain Ratio (GR)

เป็นการประเมินความน่าเชื่อถือของมิติข้อมูลโดยการวัดGain Ratio ในแต่ละคลาสการคำนวณ GR โดยใช้ค่า SplitINFO [11]

2.5. Correlation based Feature Selection (CFS)

เป็นหลักการกรองที่ง่าย โดย CFS จะจัดอันดับกลุ่มย่อยของมิติข้อมูลตามความสัมพันธ์ที่อยู่บนพื้นฐานของฟังก์ชั่นการประมาณแบบ heuristic ซึ่งกลุ่มย่อยของมิติข้อมูลจะมีความสัมพันธ์สูงกับคลาส และไม่มีความสัมพันธ์กับคลาสอื่นๆที่สามารถรวมกันแล้วให้ค่าที่สามารถให้ผลการทำนายเป็นกลุ่ม กลุ่มเดียวที่แข็งแกร่งที่สุดโดยต้องการกลุ่มของคุณลักษณะของข้อมูลที่มีขนาดเล็กที่สุด [12]

2.6. Symmetric uncertainty (SU)

วิเคราะห์ความสัมพันธ์ที่อยู่บนพื้นฐานของฟังก์ชันการประมาณใกล้เคียงกับ CFS ซึ่งกลุ่มข้อมูลที่ผ่านการกรองจะมีความสัมพันธ์กันสูงกับคลาส โดยจะใช้การจัดอันดับจากค่าสัมประสิทธิ์ (symmetrical uncertainty) เพื่อมุ่งพิจารณาระดับคุณลักษณะกับเกณฑ์ โดยค่าสัมประสิทธิ์จะต้องมากกว่าเกณฑ์ที่กำหนด เพื่อกำหนดจำนวนตัวคุณลักษณะ โดยคุณสมบัติที่ซ้ำซ้อนจะถูกกำจัดออกไป [13, 14]

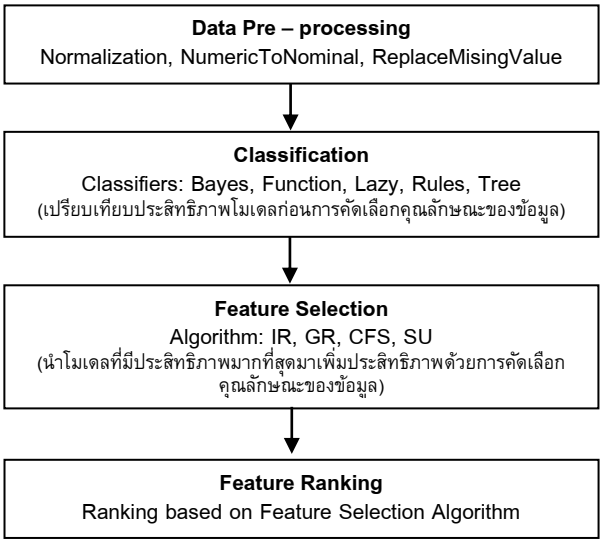
2.7. วรรณกรรมที่เกี่ยวข้อง

ก่อนหน้านี้มีนักวิจัยจำนวนมากพยายามนำวิธีการทำเหมืองข้อมูลมาสร้างโมเดลการพยากรณ์หรือปรับปรุงอัลกอริทึมที่นิยมใช้เพื่อให้มีประสิทธิภาพในการทำนายเกี่ยวกับน้ำจำนวนมาก เช่น [15] สร้างการรวมกันของ Wavelet Transforming กับโครงข่ายประสาทเทียม เพื่อสร้างความทนทานในขั้นตอนการเรียนรู้ ซึ่งก่อให้เกิดการใช้งานที่ประสบความสำเร็จมากขึ้นในการพยากรณ์น้ำท่วม และ [1] โดยอยู่บนพื้นฐานของ chaos theory (ทฤษฎีความยุ่งเหยิง) ปรับปรุงความแม่นยำซึ่งบรรลุผลที่ดีกว่ากับปัจจัยการพยากรณ์น้ำที่ซับซ้อนและมีอยู่เป็นจำนวนมากโดยใช้เวลาที่สั้น มุ่งเน้นไปที่ปัญหาการทำงานไม่เป็นเชิงเส้น และงานอื่นๆ [16, 17] นอกเหนือจากความพยายามในการนำเสนอวิธีการสร้างโมเดลแล้วยังมีบทความบางส่วนที่นำเสนอการค้นหาปัจจัยที่จำเป็นต้องใช้ในการพยากรณ์น้ำท่วม [18] เพื่อการสำรวจความสัมพันธ์ระหว่างปัจจัยที่ส่งผลต่อการเกิดน้ำท่วมและปัจจัยที่สำคัญที่สุดเพื่อเป็นตัวแทนในการพยากรณ์

นอกจากจะมีการค้นหาปัจจัยและนำอัลกอริทึมของการทำเหมืองข้อมูลมาสร้างโมเดลการพยากรณ์น้ำท่วมแล้ว ยังมีนักวิจัยบางกลุ่มพยายามนำเสนอวิธีการควบคุมการไหลของน้ำในอ่างเก็บน้ำเพื่อระบาย – กักเก็บน้ำ โดยใช้โมเดลทางอุทกศาสตร์วิทยา เช่น การควบคุมระดับน้ำของอ่างเก็บน้ำในโรงงานไฟฟ้าพลังน้ำ ด้วย non – linear predictive control (ET - NMPC) [19] การพัฒนารูปแบบการเพิ่มประสิทธิภาพโมเดลที่ใช้ในการบรรเทาอันตรายจากน้ำท่วมที่เกิดจากพายุในลุ่มน้ำ [20] นอกจากนี้งานวิจัยที่เกี่ยวกับการพยากรณ์เรื่องน้ำยังมีโมเดลทางอุทกศาสตร์วิทยาที่นิยมใช้อีก 1 โมเดล คือ MIKE11 ซึ่งเป็นโมเดลที่ใช้หาสมดุลการไหลของน้ำจากปัจจัยต่างๆ ที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่าการพยากรณ์น้ำท่วมนั้นนอกจากการสร้างโมเดลสำหรับการพยากรณ์จะเป็นสิ่งความสำคัญแล้ว ปัจจัยที่เกี่ยวข้องที่ส่งผลต่อการพยากรณ์ก็เป็นสิ่งที่สำคัญไม่แพ้กัน เนื่องจากเป็นสิ่งที่จะส่งผลต่อความแม่นยำและประสิทธิภาพของการทำนายของโมเดลที่สร้างขึ้นโดยตรง ไม่ว่าจะเป็นการพยากรณ์โดยการใช้โมเดลทางอุทกศาสตร์วิทยาหรือการพยากรณ์โดยใช้การทำเหมืองข้อมูล แต่งานวิจัยบางงานขาดการวิเคราะห์ในเรื่องของปัจจัยเหล่านี้ ดังนั้นงานวิจัยนี้ผู้วิจัยจะทำการค้นหาปัจจัยที่เกี่ยวข้องกันและมีความสำคัญในการสร้างโมเดลการพยากรณ์สำหรับสร้างโมเดลสำหรับ

พยากรณ์การควบคุมประตูระบายน้ำ โดยใช้ขั้นตอนวิธีการทำเหมืองข้อมูล เพราะการสำรวจพบว่าการระบายน้ำของอ่างเก็บน้ำมีปัจจัยที่ส่งผลต่อการระบาย – กักเก็บน้ำเป็นจำนวนมาก หากใช้โมเดลการพยากรณ์น้ำแบบ MIKE11 นั้น จะไม่เหมาะสมเนื่องจากเป็นโมเดลที่ใช้ข้อมูลหน้าตัดลำน้ำและปัจจัยในการพยากรณ์น้ำเพียงไม่กี่ปัจจัย ดังที่ยกตัวอย่างในงานวิจัยก่อนหน้า ซึ่งอาจส่งผลให้การสร้างโมเดลการพยากรณ์ผิดพลาดได้



รูปที่ 1. กรอบวิธีการดำเนินงานวิจัย

3. วิธีการดำเนินการวิจัย

วิธีการดำเนินงานวิจันเริ่มต้นโดยการเตรียมข้อมูล หลังจากนั้นหาโมเดลในการพยากรณ์ข้อมูล และทำการคัดเลือคุณลักษณะของข้อมูลเพื่อเพิ่มประสิทธิภาพให้โมเดลที่ได้ เมื่อได้คุณลักษณะของข้อมูลแล้วนำคุณลักษณะมาจัดลำดับความสำคัญ ดังแสดงในรูปที่ 1

3.1. การเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นฐานข้อมูลด้านการจัดสรรน้ำ (ระบาย – กักเก็บ) ของอ่างเก็บน้ำห้วยจรเข้มาก จังหวัดบุรีรัมย์ตั้งแต่ปี พ.ศ. 2537 – พ.ศ. 2558 ในฐานข้อมูลจะมีการเก็บข้อมูลที่เป็นปัจจัยที่ส่งผลต่อการระบายน้ำออกจากอ่างเก็บน้ำผ่านประตูระบายน้ำ ซึ่งมีทั้งหมด 7,068 ระเบียน ปัจจัยเหล่านี้จะสร้างกรณีของการระบายน้ำได้ 36 กรณี คือ การปิดการระบายน้ำ (ระบายน้ำด้วยปริมาณ 0.0 ลบ.ม./วินาที) และระบายน้ำด้วยปริมาณ 0.1 – 3.5 ลบ.ม./วินาที รายละเอียดของปัจจัยที่ส่งผลต่อการระบายน้ำเป็นดังตารางที่ 1

3.2. การสร้างโมเดลในการพยากรณ์ข้อมูล

การพยากรณ์ข้อมูลใช้การวัดค่าประสิทธิภาพจากเทคนิคในการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิคได้แก่ กลุ่มของ Bayes (Naïve Bayes, Bayes Network) กลุ่มของ Function (LivSVM, MLP, SMO) กลุ่มของ Lazy (IBk, KStar) กลุ่มของRules (DecisionTable, JRIP, OneR, PART) และกลุ่มของ Tree (J48, Random Forest, RandomTree, REFTree)

3.3. การคัดเลือกคุณลักษณะของข้อมูล

เป็นวิธีที่ใช้เพื่อลดมิติของข้อมูล หรือการกำหนดลำดับความสำคัญของคุณลักษณะของข้อมูลใหม่เพื่อเพิ่มประสิทธิภาพการพยากรณ์ และได้ปัจจัยที่มีอำนาจในการจำแนกผลลัพธ์มากที่สุด ในการทดลองนี้เปรียบเทียบการคัดเลือกคุณลักษณะของข้อมูล 4 วิธี ได้แก่ IR GR CFS และ SU

ตาราง 1. รายละเอียดของปัจจัยที่ส่งผลต่อการระบายน้ำ

ที่	ชื่อ	รายละเอียด
1	waterLevel	ระดับน้ำ (ม.ทรก.)
2	capacitance	ปริมาตรน้ำ (ลบ.ม.)
3	surfaceArea	พื้นที่ผิวน้ำ (ตร.ม.)
4	leak	การรั่วซึม (ลบ.ม/วัน)
5	inflow	ปริมาณน้ำไหลเข้าอ่าง (ลบ.ม/วินาที)
6	volatility	ระเหยออกจากอ่าง (ลบ.ม/วินาที)
7	rainAtReservoir	ปริมาณฝนตกที่อ่าง (มม.)
8	accumulateWater	ปริมาณน้ำท่าสะสม (ล้าน ลบ.ม.)
9	transplant	พื้นที่เพาะปลูกข้าว (ไร่)
10	totalRain	ปริมาณน้ำฝนสะสม (มม.)

3.5. การวัดค่าประสิทธิภาพ

ขั้นตอนการเปรียบเทียบโมเดลและวิธีเลือกคุณลักษณะข้อมูลพิจารณาจากค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ดังแสดงในสมการที่ 1 และสมการที่ 2

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

โดยที่ TP = True Positive, FP = False Positive

TN = True Negative, FN = False Negative

4. ผลการทดลอง

4.1. ผลและวิเคราะห์การเปรียบเทียบโมเดลการพยากรณ์การระบายน้ำ

จากตารางที่ 2 การเปรียบประสิทธิภาพโมเดลพยากรณ์การระบายน้ำจากเทคนิคการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิค นั้นพบว่าโมเดลที่มีประสิทธิภาพความแม่นยำในการจำแนกประเภทข้อมูลมากที่สุดคือ

โมเดลที่ได้จากกลุ่ม Tree ได้แก่ Random Forest มีค่าความถูกต้องในการพยากรณ์ 98.71% หลังจากนั้นนำโมเดลการพยากรณ์แบบ Random Forest ไปเพิ่มประสิทธิภาพความแม่นยำในการพยากรณ์ให้มากยิ่งขึ้น โดยทำการคัดเลือกคุณลักษณะของข้อมูล เพื่อให้ได้คุณลักษณะข้อมูลที่สำคัญในการพยากรณ์

ตาราง 2. ผลการวัดค่าประสิทธิภาพโมเดลพยากรณ์การระบายน้ำ

Classifiers	10- fold cross validation		
	Accuracy	Precision	RMSE
Bayes			
- Naïve Bayes	62.32 %	0.538	0.1594
- Bayes Network	69.75 %	0.773	0.1159
Function			
- LivSVM	80.31 %	0.783	0.1046
- MLP	76.45 %	0.726	0.1021
- SMO	57.08%	0.428	0.1608
Lazy			
- IBk	98.27%	0.983	0.0298
- KStar	93.63%	0.935	0.0547
Rules			
- DecisionTable	78.86%	0.785	0.132
- JRIP	86.55%	0.861	0.0814
- OneR	73.37%	0.637	0.1216
- PART	90.20%	0.9	0.0636
Tree			
- J48	89.67%	0.891	0.0648
- Random Forest	98.71%	0.987	0.0369
- RandomTree	98.16%	0.0309	0.982
- REFTree	85.97%	0.847	0.0775

4.2. ผลและวิเคราะห์การเปรียบเทียบประสิทธิภาพการลดมิติของข้อมูล

จากตารางที่ 3 หลังจากทดลองนำโมเดลการพยากรณ์แบบ Random Forest มาเพิ่มประสิทธิภาพในการพยากรณ์การระบายน้ำนั้นพบว่าโมเดลการพยากรณ์แบบ Random Forest เมื่อใช้ร่วมกับเทคนิคการลดคุณลักษณะของข้อมูลแบบ SU ทำให้ประสิทธิภาพในการพยากรณ์การระบายน้ำเพิ่มขึ้น

4.3. ผลและการวิเคราะห์ลำดับความสำคัญของปัจจัยที่เลือก โดยขึ้นอยู่กับวิธีการคัดเลือกคุณลักษณะที่แตกต่างกัน

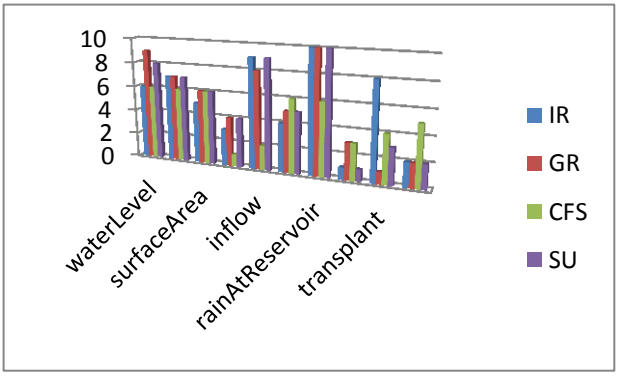
จากตารางที่ 4 แสดงให้เห็นว่าแต่ละเทคนิคของการคัดเลือกคุณลักษณะของข้อมูลมีการให้ลำดับความสำคัญของคุณลักษณะที่แตกต่างกันออกไป ในงานวิจัยนี้จึงได้นำค่าความสำคัญที่ได้จากแต่ละเทคนิคมาหาผลรวม และจัดเรียงความสำคัญของคุณลักษณะของข้อมูลใหม่ พบว่าปัจจัยที่มีความสำคัญต่อการพยากรณ์การระบายน้ำมากที่สุดคือ accumulateWater (ปริมาณน้ำท่าสะสม) รองลงมาคือ totalRain (ปริมาณน้ำฝนสะสม) และปัจจัยอื่นตามลำดับ ดังแสดงในภาพที่ 2

ตาราง 3. ผลการเปรียบเทียบประสิทธิภาพของเทคนิคการลดคุณลักษณะของข้อมูลเมื่อใช้กับโมเดลการพยากรณ์ Random Forest

Model	Accuracy	Precision	RMSE
Random Forest	98.71%	0.987	0.0369
Random Forest +IR	98.65%	0.987	0.0366
Random Forest+GR	98.65%	0.987	0.0365
Random Forest+CFS	97.99%	0.98	0.039
Random Forest+SU	98.74%	0.987	0.0366

ตาราง 4. ลำดับความสำคัญของแต่ละปัจจัย

ปัจจัย	IR	GR	CFS	SU	Sum of priority	Final Order
accumulateWater	1	3	3	1	8	1
totalRain	2	2	5	2	11	2
leak	3	4	1	4	12	3
transplant	8	1	4	3	16	4
volatility	4	5	6	5	20	5
surfaceArea	5	6	6	6	23	6
capacitance	7	7	6	7	27	7
inflow	9	8	2	9	28	8
waterLevel	6	9	6	8	29	9
rainAtReservoir	10	10	6	10	36	10



รูปที่ 2. ลำดับความสำคัญของคุณลักษณะข้อมูลจากวิธีที่แตกต่างกัน

จากตารางที่ 4 แสดงให้เห็นว่าแต่ละเทคนิคของการคัดเลือกคุณลักษณะของข้อมูลมีการให้ลำดับความสำคัญของคุณลักษณะที่แตกต่างกันออกไป ในงานวิจัยนี้จึงได้นำค่าความสำคัญที่ได้จากแต่ละเทคนิคมาหาผลรวมและจัดเรียงความสำคัญของคุณลักษณะของข้อมูลใหม่ พบว่าปัจจัยที่มีความสำคัญต่อการพยากรณ์การระบายน้ำมากที่สุดคือaccumulateWater (ปริมาณน้ำท่าสะสม) รองลงมาคือ totalRain (ปริมาณน้ำฝนสะสม) และปัจจัยอื่นตามลำดับ ดังแสดงในภาพที่ 2

5. บทสรุป

งานวิจัยนี้มีวัตถุประสงค์เพื่อเพื่อตรวจสอบความสัมพันธ์ระหว่างปัจจัยที่แตกต่างกันและระบุปัจจัยที่เกี่ยวข้องมากที่สุด สำหรับลดมิติของข้อมูลในการพยากรณ์การควบคุมประตูระบายน้ำ และสร้างโมเดลการพยากรณ์การควบคุมประตูระบายน้ำที่มีประสิทธิภาพแม่นยำ ของอ่างเก็บน้ำห้วยจรเข้มาก จังหวัดบุรีรัมย์ โดยแบ่งการทดลองออกเป็น 3 ส่วน คือ 1) การสร้างโมเดลในการพยากรณ์ข้อมูล โดยเปรียบประสิทธิภาพโมเดลพยากรณ์การระบายน้ำ จากเทคนิคการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิค นั้นพบว่าโมเดลที่มีประสิทธิภาพความแม่นยำในการจำแนกประเภทข้อมูลมากที่สุดคือโมเดลที่ได้จากกลุ่ม Tree ได้แก่ Random Forest มีค่าความถูกต้องในการพยากรณ์ 98.71% 2) การคัดเลือกคุณลักษณะของข้อมูล โดยนำโมเดลการพยากรณ์แบบ Random Forest มาเพิ่มประสิทธิภาพในการพยากรณ์การระบายน้ำ และค้นหาคุณลักษณะของข้อมูลที่สำคัญต่อการพยากรณ์ พบว่าโมเดลพยากรณ์แบบ Random Forest เมื่อใช้ร่วมกับเทคนิคการลดคุณลักษณะของข้อมูลแบบ SU ทำให้ประสิทธิภาพในการพยากรณ์การระบายน้ำเพิ่มขึ้น จาก 98.71% เป็น 98.74% 3) การจัดลำดับความสำคัญของคุณลักษณะ โดยนำผลลัพธ์จากการจัดลำดับความสำคัญของคุณลักษณะข้อมูลจากเทคนิคการเลือกคุณลักษณะของข้อมูลทั้ง 4 เทคนิค มาจัดลำดับความสำคัญใหม่ พบว่าปัจจัยที่มีความสำคัญต่อการพยากรณ์น้ำท่วมมากที่สุดคือ accumulateWater (ปริมาณน้ำท่าสะสม) รองลงมาคือ totalRain (ปริมาณน้ำฝนสะสม) และ leak (การรั่วซึม) ตามลำดับ

ผลการวิจัยในครั้งนี้สามารถนำไปสร้างเป็นระบบพยากรณ์การระบายน้ำของประตูระบายน้ำแบบอัตโนมติ โดยสามารถแยกการระบายน้ำตามปริมาณการระบายน้ำได้ตั้งแต่ 0.0 ลบ.ม/วินาที (ปิดการระบายน้ำ) ถึง 3.5 ลบ.ม/วินาที ได้ในอนาคต

เอกสารอ้างอิง

[1] H. Ping, "Wavelet neural network based on BP algorithm and its application in flood forecasting," in Granular Computing, 2009, GRC'09. IEEE International Conference on, 2009, pp. 251-253.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI magazine, vol. 17, p. 37, 1996.

[3] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining vol. 1: Pearson Addison Wesley Boston, 2006.

[4] Q. Ding, Q. Ding, and W. Perrizo, "Decision tree classification of spatial data streams using Peano Count Trees," in Proceedings of the 2002 ACM symposium on Applied computing, 2002, pp. 413-417.

[5] F. Zahedi, "An introduction to neural networks and a comparison with artificial intelligence and expert systems," Interfaces, vol. 21, pp. 25-38, 1991.

[6] H. Deng and S. Wibowo, "A Rule–Based Decision Support System for Evaluating and Selecting IS Projects," in International MultiConference of Engineers & Computer Scientists, 2008, pp. 1962-1968.

[7] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, pp. 41-46.

[8] H. Shen and K.-C. Chou, "Using optimized evidence-theoretic K-nearest neighbor classifier and pseudo-amino acid composition to predict membrane protein types," Biochemical and biophysical research communications, vol. 334, pp. 288-292, 2005.

[9] V. Nuipian, P. Meesad, and P. Boonrawd, "Improve abstract data with feature selection for classification techniques," in Advanced Materials Research, 2012, pp. 3699-3703.

[10] B.-n. Jiang, X.-Q. Ding, L.-T. Ma, Y. He, T. Wang, and W.-W. Xie, "A hybrid feature selection algorithm: Combination of symmetrical uncertainty and genetic algorithms," in The Second International Symposium on Optimization and Systems Biology, 2008, pp. 152-157.

[11] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, pp. 81-106, 1986.

[12] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," Knowledge and Data Engineering, IEEE Transactions on, vol. 15, pp. 1437-1447, 2003.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," The Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.

[14] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, pp. 1226-1238, 2005.

[15] X. Jian-cang, W. Tian-ping, Z. Jian-long, and S. Yu, "A Method of Flood Forecasting of Chaotic Radial Basis Function Neural Network," in Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on, 2010, pp. 1-5.

[16] Z. Shi, Y. Zheng, and M. Li, "Neural Network Model on Basin Flood Prevention Effect Assessment," in Intelligent Computation

Technology and Automation, 2009. ICICTA'09. Second International Conference on, 2009, pp. 83-86.

[17] K. Chaowanawatee and A. Heednacram, "Implementation of cuckoo search in RBF neural network for flood forecasting," in Computational Intelligence, Communication Systems and Networks (CICSyN), 2012 Fourth International Conference on, 2012, pp. 22-26.

[18] Q. Yang, J. Shao, M. Scholz, and C. Plant, "Feature selection methods for characterizing and classifying adaptive Sustainable Flood Retention Basins," Water research, vol. 45, pp. 993-1004, 2011.

[19] W. Zhou, H. Thoresen, and B. Glemmstad, "Application of Kalman filter based nonlinear MPC for flood gate control of hydropower plant," in Power and Energy Society General Meeting, 2012 IEEE, 2012, pp. 1-4.

[20] Y. Ding and S. S. Wang, "Optimal control of flood diversion in watershed using nonlinear optimization," Advances in Water Resources, vol. 44, pp. 30-48, 2012.