การพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยใช้เทคนิคเหมืองข้อมูล
Forecasting Daily Discharge in Kievlom Dam Using Data Mining Techniques

วีรศักดิ์ ฟองเงิน[1*] วรปภา อารีราษฎร์[2] เผด็จ พรหมสาขา ณ สกลนคร[2]
Werasak Fongngen[1*] Worapapha Arreerard[2] Padej Phomasakha Na Sakolnakorn[2]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาเทคนิคเหมืองข้อมูลที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อน และ 2) เปรียบเทียบผลปริมาณน้ำจริงกับผลการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้บริหารจัดการข้อมูลน้ำที่มีอยู่ที่เป็นปัจจัยต่อการเปลี่ยนแปลงระดับน้ำ ประกอบด้วย ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำ และอัตราการระเหย โดยรวบรวมข้อมูลรายวัน ตั้งแต่ปี พ.ศ. 2535 – พ.ศ. 2559 รวม 25 ปี จำนวน 9,125 ระเบียน เพื่อนำมาพยากรณ์ด้วยเทคนิคการพยากรณ์

ผลการวิจัยพบว่า 1) เทคนิคเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ปริมาณน้ำในเขื่อนประกอบด้วย 4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย วิธีโครงข่ายประสาทเทียม วิธีแบบจำลองต้นไม้เอ็มไฟว์พีและวิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ โดยเลือกจากงานวิจัยที่เกี่ยวข้องที่มีการใช้งานมากที่สุดและ 2) ผลการเปรียบเทียบการพยากรณ์ปริมาณน้ำรายเดือนในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูลทั้ง 4 เทคนิค พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.58 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบเพื่อการพยากรณ์น้ำในเขื่อน ทั้งนี้เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคเรียงตามค่าความคลาดเคลื่อนจากน้อยไปมาก พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พีมีค่าเท่ากับ 10.58 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ มีค่าเท่ากับ 19.90 วิธีวิเคราะห์การถดถอย มีค่าเท่ากับ 20.00 และวิธีโครงข่ายประสาทเทียม มีค่าเท่ากับ 21.65 ตามลำดับ

คำสำคัญ : การพยากรณ์, เหมืองข้อมูล, เขื่อนกิ่วลม

Abstract

This research aims to 1) study the appropriate data mining techniques to forecast the amount of water in the dam, and 2) compare the actual water volume and the prediction with the amount of water in KiewLom Dam, Lampang Province, using data mining techniques. This research manages existing water data that is a factor in water level changes consist of the amount of water flowing into the dam, the amount of water in the dam, emissions and evaporated water. The data were collected daily from the year 2535 – 2559 totally 25 years with 9,125 records to use for forecasting with Techniques.

The results showed that: 1) appropriate data mining techniques used to forecast the water level in the dam consist of 4 techniques were (Regression Analysis: RA) , (Artificial Neural Network : ANN), (M5P Model Tree : M5P), and (Support Vector Machine : SVM) by choosing from the most relevant research works, and 2) comparison results of the prediction of monthly water Level in KiewLom dam, Lampang, using data mining four techniques found that M5P Model Tree

---

[1*] อาจารย์ประจำคณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏลำปาง
Lecturer of Science Faculty, Lampang Rajabhat University
[2] อาจารย์ประจำคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
Lecturer of Information Technology Faculty, Rajabhat Mahasarakham University

had Mean absolute Error as low as 10.58 and was the most appropriate way to develop a system for forecasting the water in the dam. Considering the absolute value of a Mean absolute Error from each technique in ascending order was: how the model tree M5p is equal 10.58, support vector machine has an estimated value of 19.90, regression analysis method is equal 20.00, and artificial neural networks is equivalent to 21.65.

Keywords : Forecasting, Data Mining, Kiew Lom Dam

บทนำ

สำนักงานชลประทานที่ 2 จังหวัดลำปางซึ่งเป็นหน่วยงานที่รับผิดชอบดูแลเขื่อนกิ่วลมได้มีนโยบายให้มีการนำข้อมูลน้ำในอดีตที่ผ่านมาทำการจัดเก็บให้เป็นระบบ และนำมาเป็นข้อมูลเพื่อช่วยตัดสินใจในการแจ้งเตือนประชากรกลุ่มเสี่ยงในพื้นที่ท้ายเขื่อน เพื่อลดผลกระทบที่จะเกิดขึ้น โดยงานวิจัยนี้ได้มีการบริหารจัดการกับข้อมูลน้ำที่จัดเก็บไว้ในอดีต มาประยุกต์ใช้เป็นข้อมูลช่วยในการตัดสินใจสำหรับการพยากรณ์กระบวนการคาดคะเนแนวโน้มหรือรูปแบบของการอันเกิดเหตุการณ์ต่างๆ ในอนาคตซึ่งได้จะทำให้ทราบแนวทางการวางแผนป้องกันและการรับมือเพื่อลดผลกระทบจะเกิดขึ้น จากที่ผ่านมาพบว่า จังหวัดลำปางประสบปัญหาสภาวะน้ำท่วม ในช่วงหลายปีที่ผ่านมาพื้นที่ชุมชนได้รับผลกระทบเป็นจำนวนมากไม่ว่าจากผลกระทบน้ำท่วมใหญ่ปี 2548 ปี 2552 และ ปี 2554 เนื่องจากระยะทางจากเขื่อน กิ่วลมมายังพื้นที่ชุมชนหนาแน่นในอำเภอเมืองมีระยะทางแค่ 30 กิโลเมตร ที่ผ่านมามีการแจ้งเตือนภัยล่วงหน้า แต่มีการแจ้งเตือนในเวลาที่กระชั้นชิด ไม่เพียงพอสำหรับเตรียมการรับมือภัยน้ำท่วมที่จะเกิดขึ้น

ในงานวิจัยนี้ ผู้วิจัยได้ศึกษาเทคนิคพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยการศึกษาจากงานวิจัยที่เกี่ยวข้องและเลือกเทคนิควิธีที่เหมาะสมมาประยุกต์ และ จัดการกับข้อมูลในอดีตที่เตรียมไว้เพื่อนำมาวิเคราะห์ และทดสอบกับข้อมูลย้อนหลัง 25 ปี จำนวน 9,125 ระเบียน มาสรุปผลหลายๆ เทคนิควิธี และเลือกใช้วิธีพยากรณ์ที่มีค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน และ ค่าเฉลี่ยของรากที่สองของกำลังสองน้อยที่สุด ซึ่งหมายถึงเทคนิควิธีดังกล่าวเป็นการพยากรณ์ที่มีความแม่นยำมากที่สุด นำมาจัดการนำไปสู่การพัฒนาระบบพยากรณ์น้ำในเขื่อน

ผู้วิจัยหวังว่าระบบพยากรณ์นำในเขื่อนกิ่วลม จะช่วยแจ้งเตือนภัยที่เกิดจากน้ำท่วมล่วงหน้า เพื่อช่วยลดผลกระทบที่เกิดกับทรัพย์สินของประชาชน และ ช่วยลดงบประมาณของภาครัฐในการจัดซื้อเครื่องเตือนภัยน้ำท่วม

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเทคนิคเหมืองข้อมูลที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อน
2. เพื่อเปรียบเทียบผลปริมาณน้ำจริงกับผลการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

เอกสารและงานวิจัยที่เกี่ยวข้อง

มนต์ชัย เทียนทอง (2548) อธิบายว่า การพยากรณ์ หมายถึง การคาดการณ์เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่สนใจที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศประกอบการตัดสินใจ ซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดที่จะต้องทำก่อนการวางแผน หรือการเตรียมการที่จะเริ่มทำอะไรเพื่อความถูกต้องและแม่นยำในการตัดสินใจ การพยากรณ์สามารถแบ่งตามระยะเวลาได้ 3 ประเภทคือ 1) การพยากรณ์ระยะสั้น 2) การพยากรณ์ระยะกลางและ 3) การพยากรณ์ระยะยาว ในการพยากรณ์ มีเทคนิคของการพยากรณ์ประกอบด้วย 1) การพยากรณ์เชิงปริมาณ เป็นการพยากรณ์โดยอาศัยข้อมูลหรือตัวเลขจากอดีตและพยากรณ์ไปในอนาคต และ 2). การพยากรณ์เชิงคุณภาพ เป็นกลุ่มของวิธีการพยากรณ์ที่อาศัยข้อมูลและวิธีการเชิงคุณภาพ ใช้กับลักษณะของปัญหาที่ไม่มีข้อมูลย้อนหลังหรือมีข้อมูลไม่มากพอ มาใช้สร้างตัวแบบ

ณัฐริน เจริญเกียรติบวร (2549) อธิบายว่า เหมืองข้อมูล หมายถึง กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ ขั้นตอนการทำเหมืองข้อมูลประกอบด้วยการคัดลอกข้อมูล การเตรียมข้อมูล การแปลงรูปแบบข้อมูล การทำเหมืองข้อมูล และการวิเคราะห์ข้อมูล ในงานเหมืองข้อมูลประกอบด้วยเทคนิคต่าง ๆ ได้แก่การจัดหมวดหมู่ การประเมินค่า การทำนาย การจัดกลุ่ม และ การรวมตัว

พยุง มีสัจ (2555) อธิบายว่า โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นการศึกษาระบบการทำงานของเซลล์ประสาทภายในสมองที่ประกอบด้วยเซลล์ประสาท (Neuron) และเส้นประสาท โดยที่เซลล์ประสาทจะเชื่อมต่อกันในรูปแบบโครงข่าย ซึ่งการวิเคราะห์ และประมวลผลข้อมูลของระบบประสาทนั้นจะส่งข้อมูลผ่านระบบโครงข่ายของเซลล์ประสาท และทำงานในลักษณะขนาน คือ ทำกิจกรรม หรืองานหลายอย่างได้ในเวลาเดียวกันเพื่อให้ได้มาซึ่งผลลัพธ์ที่ต้องการ โดยการทำงานของสมองในรูปแบบที่กล่าวมาในข้างต้นนั้นมีความสามารถหลายประการเช่น การสังเกต เรียนรู้ จดจำ ทำซ้ำ และแยกแยะสิ่งต่าง ๆ ซึ่งโครงข่ายประสาทเทียมได้จำลองรูปแบบการทำงาน และโครงสร้างการเชื่อมต่อดังกล่าวมา เพื่อทำให้ระบบคอมพิวเตอร์สามารถทำงานที่สมองทำได้ ซึ่งสามารถเพิ่มประสิทธิภาพในการทำงานให้มากขึ้น และช่วยลดข้อผิดพลาดที่เกิดขึ้นจากการทำงาน

พิชญ์สินี ชมพูคำ (2554) อธิบายว่า เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) คือ การหาความสัมพันธ์ของตัวแปรหนึ่งกับตัวแปรอื่น ๆ มักจะใช้เครื่องมือทางสถิติที่เรียกว่า ซึ่งเป็นกระบวนการทั้งหมดที่สร้างความสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป โดยที่ตัวแปรหนึ่งเป็นผลมาจากอีกตัวแปรหนึ่งหรือหลายตัวแปรก็ได้ โดยอาศัยข้อมูลจากการสุ่มตัวอย่างและการอนุมาน

Preis & Ostfeld (2007) อธิบายว่า แบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) เป็นต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่อง และไม่ต่อเนื่อง นอกจากนี้ที่โหนดใบ (Leaf Node) ของต้นไม้จะมีแบบจำลองเชิงเส้นที่ใช้ทำนายกลุ่มของข้อมูลที่เป็นค่าตัวเลข ในการทำนายค่าของกลุ่มข้อมูลจะวิเคราะห์ข้อมูลจากโหนดราก (Root Node) ลงมาจนกระทั่งถึงโหนดใบที่จะได้ค่าคุณลักษณะในการตัดสินใจเลือกเส้นทางจากโหนดแต่ละชั้น

Duda & HartPattern (1993) อธิบายว่า เทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (SVM) เป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุดในโลกแห่งความเป็นจริงนั้น ข้อมูลสองกลุ่มไม่ได้วางตัวในพื้นที่คุณลักษณะ เป็นเพียงสองกลุ่มและแบ่งได้โดยเส้นตรง จึงเป็นปัญหาทำให้ไม่สามารถที่จะใช้สมการซัพพอร์ตเวคเตอร์แมชชีนส์แบบเชิงเส้นได้

วิธีดำเนินการวิจัย

1. ขั้นตอนการดำเนินการวิจัยประกอบด้วย 5 ขั้นตอน ดังนี้

1.1 ศึกษาเทคนิคการพยากรณ์จากงานวิจัยที่เกี่ยวของและเลือกนำเทคนิคการพยากรณ์ที่มีการนำมาใช้ในงานวิจัยมากที่สุดจำนวน 4 เทคนิคมาเปรียบเทียบหาความแม่นยำกับข้อมูลที่เตรียมไว้

1.2 จัดเตรียมข้อมูล โดยจัดเรียงข้อมูลตามช่วงเวลารายวันตามปัจจัยที่เกี่ยวข้องเพื่อนำมาวิเคราะห์ด้วยเทคนิคที่เลือกไว้ในขั้นตอนที่ 1.1

1.3 วิเคราะห์ข้อมูลที่จัดเตรียมด้วยเทคนิคการพยากรณ์ทั้ง 4 เทคนิค

1.4 เปรียบเทียบค่าสัมบูรณ์ของความคลาดเคลื่อนทั้ง 4 เทคนิค และสรุปผลการวิเคราะห์ข้อมูลเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

1.5 สรุปผลการศึกษาและเขียนรายงานการวิจัย

เครื่องมือการวิจัย

1. โปรแกรมที่ใช้ในการวิเคราะห์ใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์ค่าสถิติการพยากรณ์ ด้วยโปรแกรม WEKA

2. เทคนิคการพยากรณ์ จำนวน 4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย (Regression Analysis :RA) เทคนิควิธีซัพพอร์ตเวคเตอร์แมชชีนส์ (Support Vector Machine : SVM) เทคนิควิธีแบบจำลองต้นไม้เอ็มไฟว์พี (M5P Model Tree: M5P) วิธีเทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

ข้อมูลที่ใช้ในการวิจัย

ข้อมูลปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง พ.ศ. 2535-2560 จำนวน 25 ปี รวมทั้งหมด 9,125 ระเบียน

สถิติที่ใช้ในการวิจัย ได้แก่

สถิติการพยากรณ์ โดยพิจารณาค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)

ผลการศึกษาเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง นำมาสรุปเทคนิคการพยากรณ์ปริมาณน้ำที่เหมาะสมได้จากงานวิจัยที่เกี่ยวข้องที่มีการนำมาทดสอบมากที่สุดจำนวน 4 เทคนิควิธี ได้แก่

1. วิธีการวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีที่ใช้การศึกษาความสัมพันธ์ระหว่างสองตัวแปร ตัวแปรหนึ่งเรียกว่าตัวแปรต้นหรือตัวแปรอิสระ (Independent variable) แทนด้วย X อีกตัวแปรหนึ่งเรียกว่าตัวแปรตาม(Dependent variable) แทนด้วย Y เป็นการดูความสัมพันธ์ว่าถ้า ตัวแปรอิสระเปลี่ยนไปแล้วตัวแปรตามเปลี่ยนไปด้วยหรือไม่

2. วิธีโครงข่ายประสาทเทียม (Artificial Neural Network) เป็นวิธีที่จำลองโมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (Neurons) และ จุดประสานประสาท (Synapses) ตามโมเดลนี้ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน

3. วิธีแบบจำลองต้นไม้เอ็มไฟว์พี (M5P Model Tree) เป็นวิธีที่จำลองต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่องและไม่ต่อเนื่อง

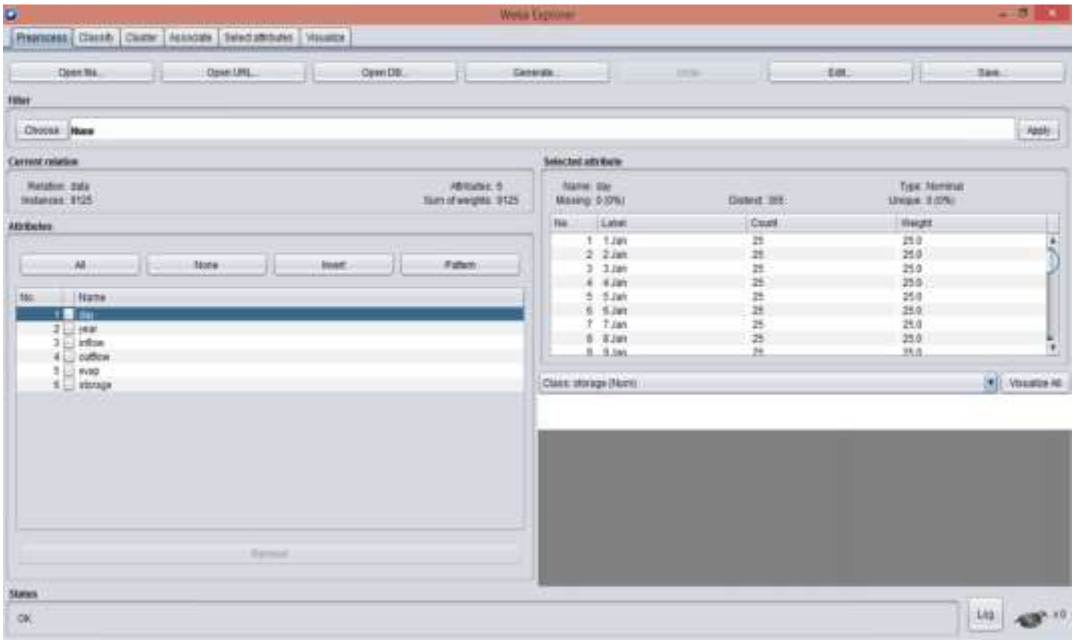
4. วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ (Support Vector Machine) เป็นวิธีที่สร้างสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด

จากผลการศึกษาที่ได้ ผู้วิจัยจะนำเทคนิคที่ได้จากการศึกษาไปพยากรณ์ข้อมูลที่ได้จัดเตรียมไว้ เพื่อเปรียบเทียบและสรุปผล และนำไปใช้ในการดำเนินการในขั้นตอนต่อไป

ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล

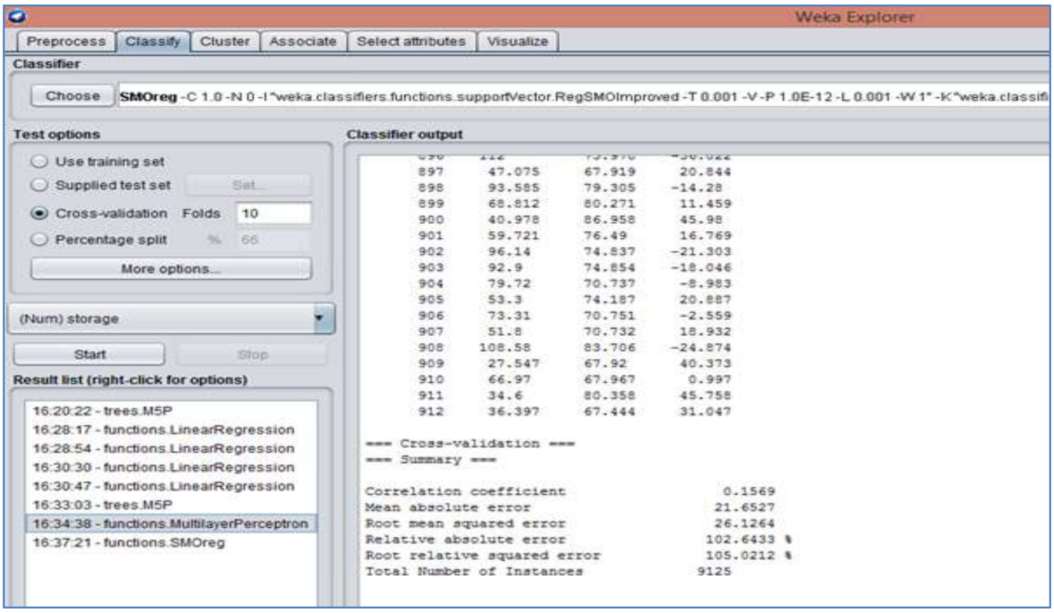
ผู้วิจัยได้นำเทคนิคการพยากรณ์ที่ได้จากการศึกษาค้นคว้า มาวิเคราะห์ด้วยข้อมูลจำนวน 9,125 ระเบียน ตามปัจจัยที่เกี่ยวข้องประกอบด้วย ข้อมูลปริมาณน้ำเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำ และ อัตราการละเหย โดยมีภาพรวมของข้อมูลที่เกี่ยวข้องดังภาพที่ 1

ภาพที่ 1 ข้อมูลภาพรวมปัจจัยที่เกี่ยวข้องสำหรับนำมาพยากร์ปริมาณน้ำเขื่อนกิ่วลม ระหว่างปี 2535 ถึงปี 2559

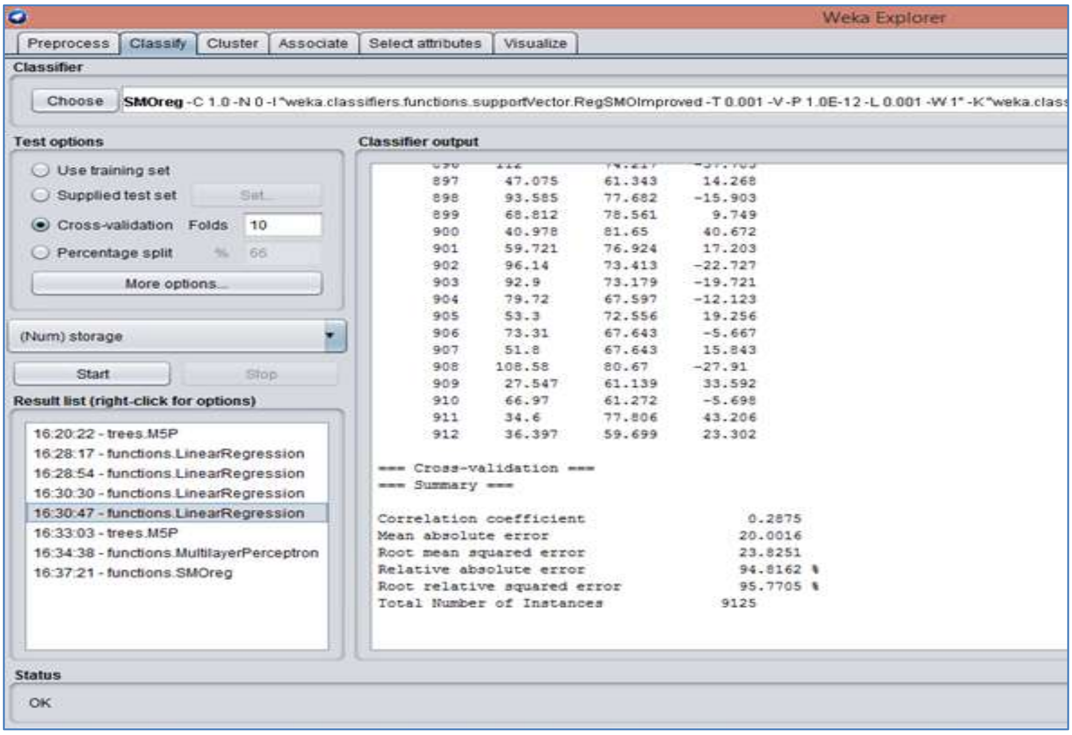


ผู้วิจัยได้นำข้อมูลจากปัจจัยที่เกี่ยวข้องประกอบด้วย วันที่ ปี ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในปล่อยออกจากเขื่อน อัตราการละเหย ปริมาณน้ำในเขื่อน ทั้งหมด 9,125 ระเบียน เป็นข้อมูลย้อนหลังรายวัน จำนวน 25 ปี ลงในโปรแกรม WEKA ทดสอบประสิทธิภาพของโมเดลเพื่อทำผลที่ได้ให้มีความน่าเชื่อถือ ผู้วิจัยได้มีการวัดประสิทธิภาพด้วยวิธี Cross-Validation 10-fold คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลัง จากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดลเพื่อทำการวิเคราะห์ทดสอบข้อมูลหาค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) มาทำการเปรียบเทียบโดยใช้โปรแกรม WEKA ผลการวิเคราะห์หาค่าความคลาดเคลื่อนทั้ง 4 วิธี ปรากฏดังภาพที่ 2-5

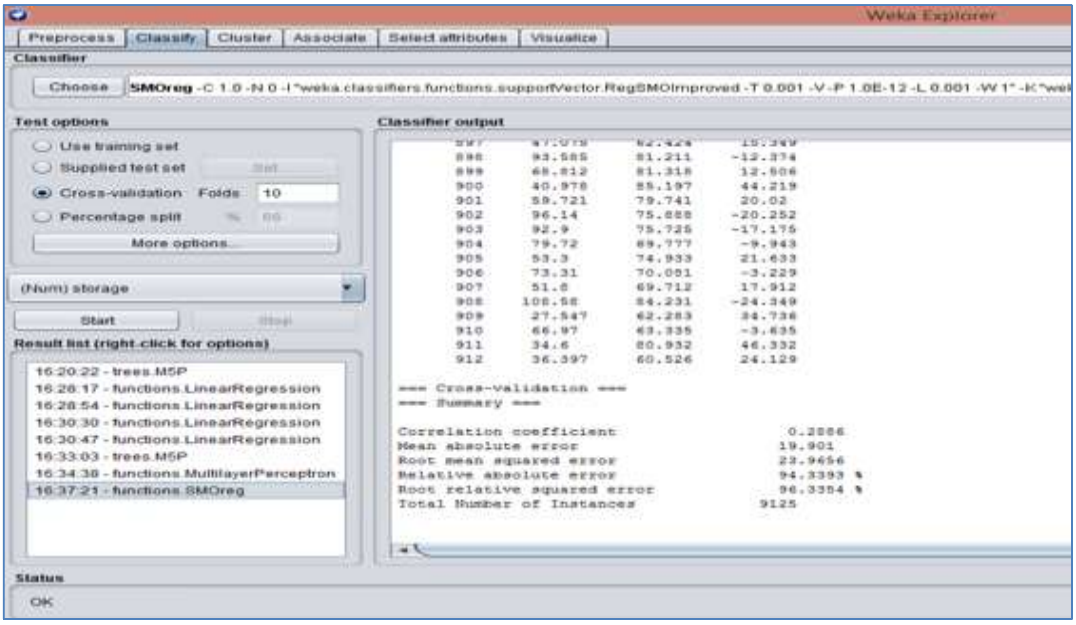
ภาพที่ 2 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีโครงข่ายประสาทเทียม



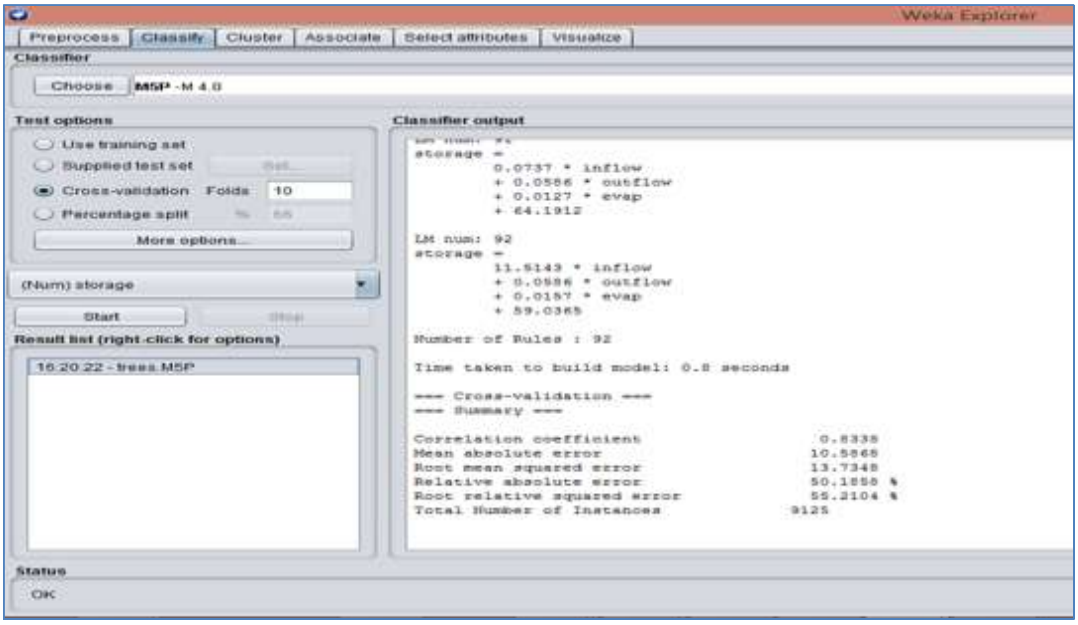
ภาพที่ 3 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีการวิเคราะห์การถดถอย



ภาพที่ 4 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีซัพพอร์ตเวคเตอร์แมชชีนส์



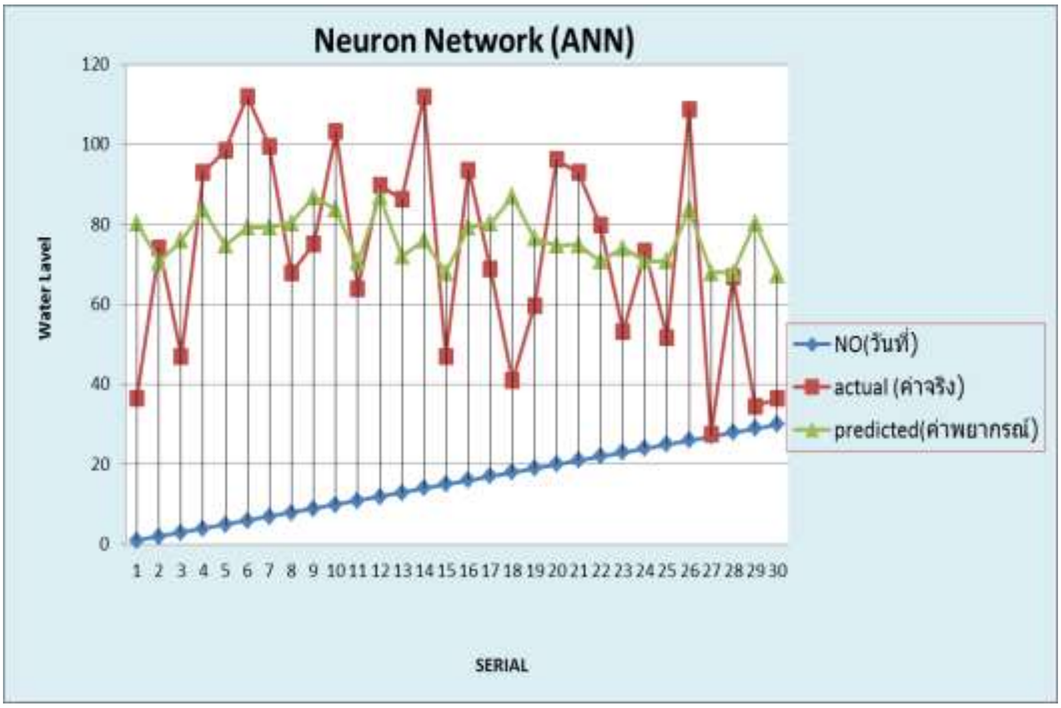
ภาพที่ 5 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี



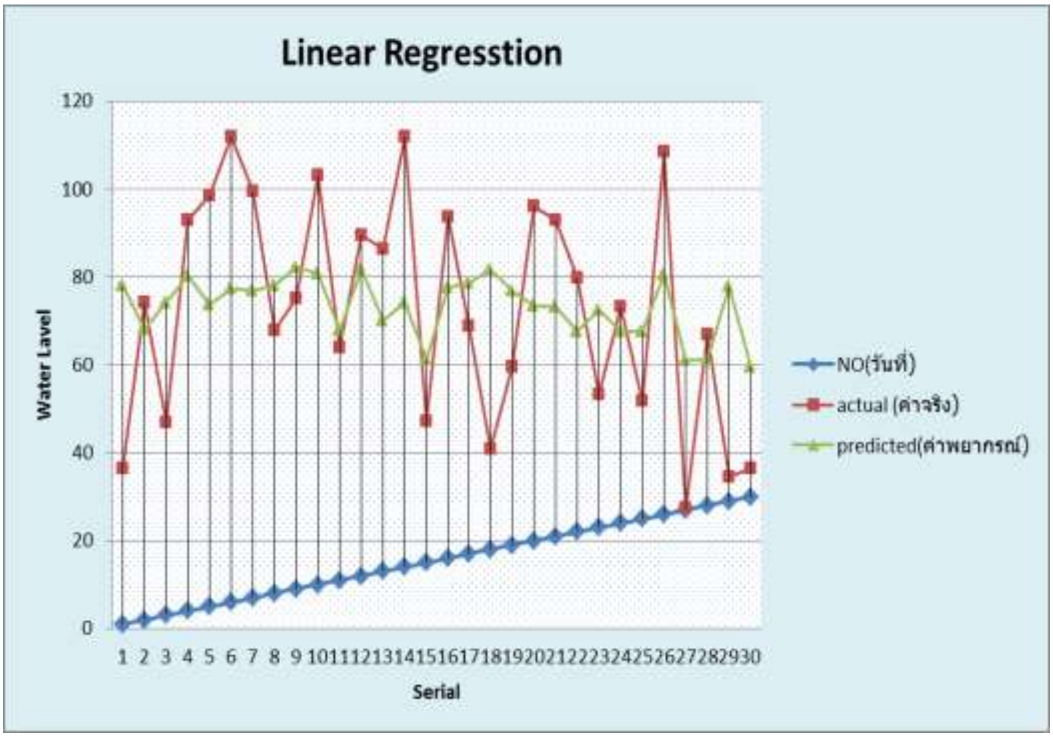
จากภาพที่ 5 แสดงผลการวิเคราะห์หาประสิทธิภาพของแต่ละโมเดล พบว่า ค่าสัมบูรณ์ของความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.58 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อน ทั้งนี้ เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าความคลาดเคลื่อนจากน้อยสุดไปหามากสุดตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนเท่ากับ 10.58 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคลาดเคลื่อนเท่ากับ 19.90 วิธีวิเคราะห์การถดถอยมีค่าความคลาดเคลื่อนเท่ากับ 20.00 และ วิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 21.65 ตามลำดับ

จากผลการวิเคราะห์หาความคลาดเคลื่อนแต่ละโมเดล ผู้วิจัยได้สร้างแบบจำลองเพื่อให้เห็นความแตกต่างความคลาดเคลื่อนระหว่าง ค่าปริมาณน้ำจริง เปรียบเทียบกับค่าที่แต่ละโมเดลมีการพยากรณ์ได้ โดยนำข้อมูลของปริมาณน้ำเดือนล่าสุดมาสร้างแบบจำลอง กับ เทคนิคการพยากรณ์ 4 วิธีได้ผลดังภาพที่ 6-9

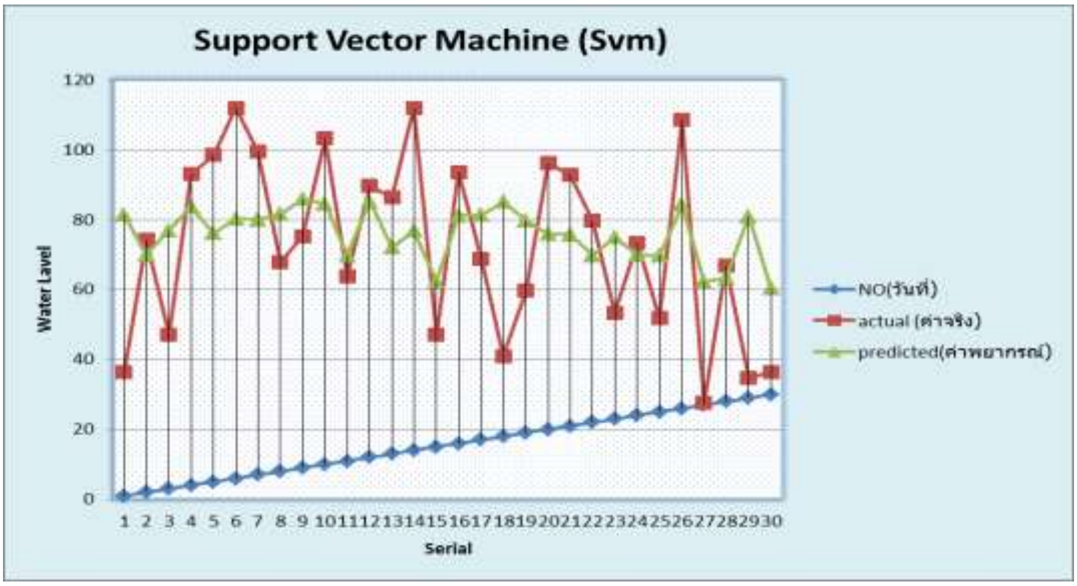
ภาพที่ 6 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีโครงข่ายประสาทเทียม



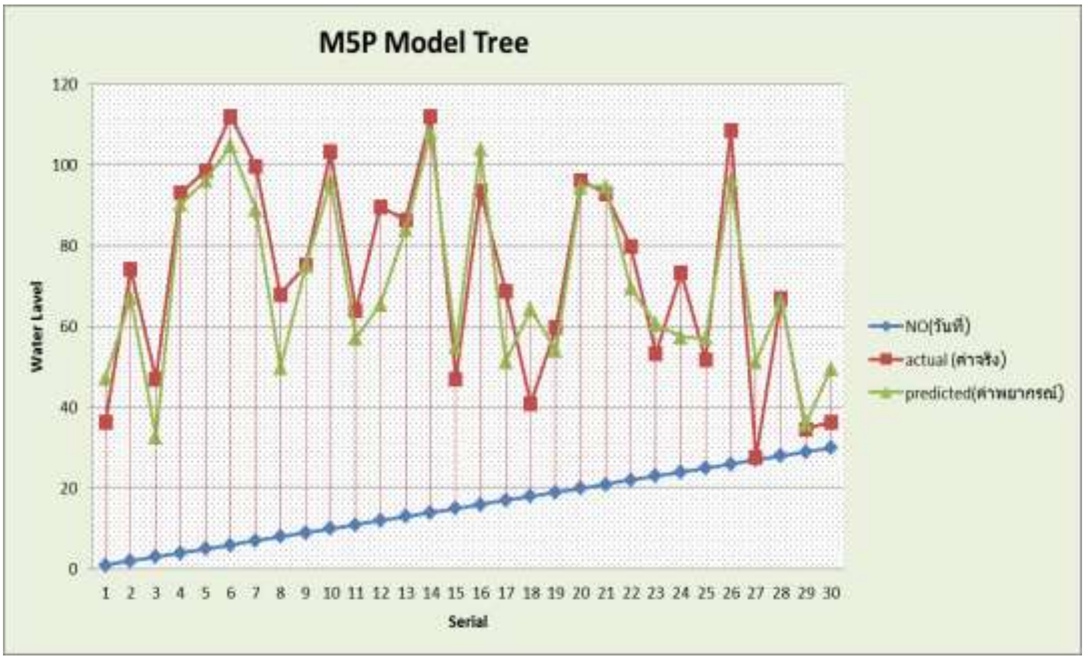
ภาพที่ 7 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีการวิเคราะห์การถดถอย



ภาพที่ 8 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีซัพพอร์ตเวคเตอร์แมชชีนส์

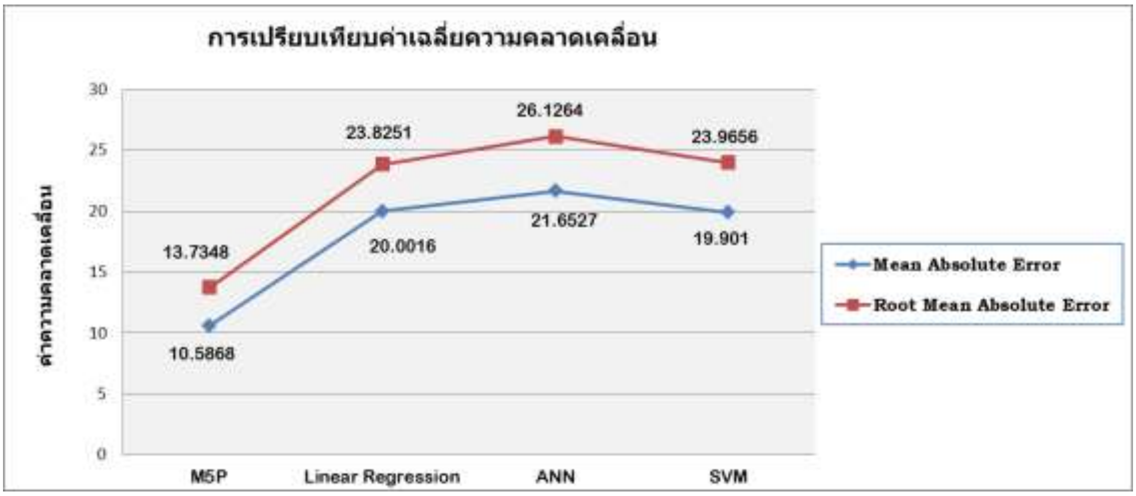


ภาพที่ 9 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี



จากภาพที่ 6-9 เป็นแบบจำลองผลการเปรียบเทียบค่าปริมาณน้ำค่าจริง กับ ค่าที่มีการพยากรณ์ แต่ละเทคนิควิธี พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนต่ำสุด จากภาพที่ 9 ค่าที่แสดงปริมาณน้ำจริง เปรียบเทียบกับค่าพยากรณ์แนวโน้มใกล้เคียงกันที่สุด เหมาะสำหรับนำไปใช้ในการพัฒนาระบบพยากรณ์น้ำ และ สอดคล้องกับการนำค่าความคลาดเคลื่อนแต่ละวิธีมาเปรียบเทียบในภาพรวมเพื่อเลือกโมเดลที่ดีที่สุดจากค่าความคลาดเคลื่อนแสดงดังภาพที่ 10

***ภาพที่* 10** ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนจาก 4 วิธี



จากผลการเปรียบเทียบการหาค่าเฉลี่ยคลาดเคลื่อนทั้ง 4 วิธี พบว่า ค่าสัมบูรณ์ของความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.58 เป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อนกิ่วลมต่อไป ทั้งนี้ เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าคลาดเคลื่อนจากน้อยไปหามากตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนเท่ากับ 10.58 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคลาดเคลื่อนเท่ากับ 19.90 วิธีวิเคราะห์การถดถอยมีค่าความคลาดเคลื่อนเท่ากับ 20.00 และ วิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 21.65

ผลการวัดความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พีเป็นวิธีที่มีค่าความคลาดต่ำสุดเหมาะสมเป็นวิธีนำไปพัฒนาระบบพยากรณ์ปริมาณน้ำในเขื่อน เมื่อได้ระบบผู้วิจัยหวังว่าระบบพยากรณ์ที่พัฒนาขึ้นประกอบด้วยการพยากรณ์ปริมาณน้ำรายวัน รายสัปดาห์ และ รายเดือน บุคคลทั่วไปสามารถตรวจสอบระดับน้ำแบบเวลาจริงและค่าที่พยากรณ์ล่วงหน้าแบบออนไลน์ได้ จะทำให้มีการเตรียมพร้อมสำหรับการรับมือ กับ ผลกระทบจากภัยน้ำท่วมที่จะเกิดขึ้นได้ทันท่วงที และ สามารถลดงบประมาณจากภาครัฐจากการจัดซื้ออุปกรณ์สำหรับเตือนภัย

สรุปและอภิปรายผล

การศึกษาเทคนิคการพยากรณ์ที่เหมาะสมในการนำมาวิเคราะห์ปริมาณน้ำในเขื่อน โดยใช้ 4 เทคนิค ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อน โดยใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)) ของวิธีโครงข่ายประสาทเทียมเท่ากับ 21.65 % และ 26.12 % วิธีวิเคราะห์การถดถอยเท่ากับ 20.00 % และ 23.82 % วิธีแบบจำลองต้นไม้เอ็มไฟว์พีเท่ากับ 10.58 % และ 13.73 % วิธีซัพพอร์ตเวคเตอร์แมชชีนส์เท่ากับ 19.90% และ 23.96 % สรุปผลการวิเคราะห์พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ และความคลาดเคลื่อนเฉลี่ยกำลังสอง มีค่าความคลาดเคลื่อนต่ำสุด เหมาะสำหรับนำไปพัฒนาระบบเพื่อใช้ในการพยากรณ์ต่อไป

ระบบพยากรณ์ปริมาณปริมาณน้ำในเขื่อนที่จะพัฒนาขึ้นโดยนำวิธีแบบจำลองต้นไม้เอ็มไฟว์พีซึ่งมาค่าความคลาดเคลื่อนต่ำสุด จะสามารถพยากรณ์ปริมาณน้ำล่วงหน้า รายวัน รายสัปดาห์ และ รายเดือน ระบบพยากรณ์จะช่วยในการแจ้งเตือนให้ทันท่วงทีและช่วยลดผลกระทบกับบ้านเรือน และ ทรัพย์สินที่จะเกิดขึ้น รวมทั้งเป็นการช่วยประหยัดงบประมาณในการจัดซื้ออุปกรณ์สำหรับเตือนภัยน้ำท่วมต่อไป

ข้อเสนอแนะ

งานวิจัยนี้ได้ทำการศึกษาเทคนิคการพยากรณ์เพียง 4 เทคนิค หากมีการศึกษาเทคนิคการพยากรณ์อื่นที่เหมาะสมในการนำมาศึกษาเปรียบเทียบอาจจะทำให้ได้ค่าการพยากรณ์ที่แตกต่างกันในแต่ละเทคนิค ทั้งนี้ ข้อมูลที่นำมาใช้ในการพยากรณ์ปริมาณน้ำในเขื่อน ผู้วิจัยใช้ข้อมูลย้อนหลัง 25 ปี ที่เก็บโดยสำนักงานชลประทานที่ 2 จังหวัดลำปางซึ่งเป็นหน่วยงานที่ดูแลเขื่อนกิ่วลม โดยแยกข้อมูลพยากรณ์จากปัจจัยที่เกี่ยวข้องที่มีการจัดเก็บข้อมูลไว้แล้ว ในการทำวิจัยที่พยากรณ์ระดับน้ำในเขื่อนอื่นที่มีลักษณะงานวิจัยที่ใกล้เคียงกันสามารถนำไปปรับใช้ให้เหมาะสมกับการพยากรณ์โดยอาจมีการเพิ่มปัจจัยที่เกี่ยวข้องที่ทำให้การพยากรณ์แม่นยำขึ้น เช่น ข้อมูลปริมาณน้ำฝน เป็นต้น

รายการอ้างอิง

ณัฐริน เจริญเกียรติบวร. (2549). *การใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับอุดมศึกษา,* ภาควิชาคอมพิวเตอร์, มหาวิทยาลัยศิลปากร .

พยุง มีสัจ.(2555) *ระบบฟัซซี่และโครงข่ายประสาทเทียม,* วิทยานิพนธ์, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

พิชญ์สินี ชมพูคำ . (2554). *การวิเคราะห์การถดถอย.* [Online] Available: http://www.hosting.cmru.ac.th/phitsinee/regression/index.php ค้นเมื่อ [2559, กันยายน 30].

มนต์ชัย เทียนทอง. (2548**).** *สถิติและวิธีการวิจัยทางเทคโนโลยีสารสนเทศ.* กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

Duda, R.O. & HartPattern, P.E. (1993). *Classification and Scene Analysis.* New York : John Wiley.

Preis, A. & Ostfeld, A. (2007). *A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds.* Journal of Hydrology .