ภาคผนวก จ

งานวิจัยที่ส่งตีพิมพ์

ส่งตีพิมพ์ที่การประชุมวิชาการการจัดการเทคโนโลยีและนวัตกรรมครั้งที่ 3 คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

ผลการศึกษาการพยากรณ์ปริมาณน้ำในเขื่อน โดยใช้เทคนิคเหมืองข้อมูล

The study on Forecasting Daily Discharge in Dam Using Data Mining Techniques

วีรศักดิ์ ฟองเงิน [1*] วรปภา อารีราษฎร์ [2]เผด็จ พรหมสาขา ณ สกลนคร [3]

นักศึกษาหลักสูตรปรัชญาดุษฎีบัณฑิต [1] คณะเทคโนโลยีสารสนเทศ[2]คณะเทคโนโลยีสารสนเทศ [3]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาเทคนิคการพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อนในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล2)เพื่อศึกษาผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูลงานวิจัยนี้ได้นำข้อมูลที่เป็นปัจจัยที่มีผลต่อการเปลี่ยนแปลงระดับน้ำประกอบด้วย ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำและ อัตราการละเหย โดยรวบรวมข้อมูลรายวัน ตั้งแต่ปี พ.ศ. 2535 – พ.ศ. 2559 รวม 25 ปี จำนวน 9,300 รายการ โดยมีการแยกข้อมูลรายเดือนเพื่อนำมาทดสอบเปรียบเทียบเทคนิคเหมืองข้อมูล

ผลการวิจัยพบว่า 1) เทคนิคการพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณนำในเขื่อนประกอบด้วย4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย (Regression Analysis) วิธีและโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) วิธีแบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) และ วิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (Support vector machineSVM) และ 2) ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูลทั้ง 4 เทคนิค พบว่าวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อน (MAE) และ ค่าเฉลี่ยของค่าสัมบูรณ์ของเปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) น้อยที่สุด เป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อน ทั้งนี้เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าคลาดเคลื่อนจากน้อยไปหามากตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พีมีค่าความคาดเคลื่อนเท่ากับที่ 10.56 และ 13.54 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคาดเคลื่อนเท่ากับ 11.40 และ 14.98วิธีวิเคราะห์การถดถอยมีค่าความคาดเคลื่อนเท่ากับ 11.66% และ 14.71 วิธีโครงข่ายประสาทเทียม มีค่าความคาดเคลื่อนเท่ากับ 12.96 และ 16.55

คำสำคัญ: การพยากรณ์ เหมืองข้อมูล การวิเคราะห์การถดถอย โครงข่ายประสาทเทียม ต้นไม้เอ็มไฟว์พี ซัพพอร์ตเวคเตอร์แมชชีนส์

ABSTRACT

The purposes of the research were to 1) Study appropriate forecasting techniques to predict the level of water in the KiewLom Dam, Lampang Province by using data mining techniques, 2) compare results of each forecasting techniques from first objective, to predict the level of water.

Data used in this research were amount of water flowing into the dam, the amount of water in the dam, the amount of water in the dam Emissions, amount of water Evaporation. Data were collected daily from the year 2535 - 2559, (totally 25 years with 9300 records), and a separate monthly data were used for the comparative test data mining techniques.

The research findings showed that 1) Appropriate forecasting techniques to forecast the water level in dam in this research were regression analysis, Artificial Neural Network: ANN, Model Tree: M5P, and Support vector machine: SVM. 2)The comparing between 4 forecasting techniques showed that Model Tree: M5P had the minimal absolute value of a deviation (Mean Square Error, MSE) and the average absolute percentage error (Mean Absolute Percentage Error, MAPE) and was the most appropriate technique to develop forecasts of water in the dam. When comparing the absolute value of a deviation from the technical tolerances in ascending order, Model Tree: M5P value was equal to 10.56 and 13.54, support vector machine's value was equal to 11.40 and 14.98, regression analysis value was equal to11.66%, and 14.71 and neural network swas equal to 12.96 and 16.55.

Keyword: Forecasting, Data Mining, Regression Analysis, Artificial Neural Network, Model Tree: M5P, Support Vector Machine

บทนำ

สำนักงานชลประทานที่ 2 จังหวัดลำปาง เป็นหน่วยงานที่ดูแลรับผิดชอบดูแลเขื่อนกิ่วลมได้มีนโยบายให้มีการนำข้อมูลน้ำอดีตที่ผ่านมามีการจัดเก็บข้อมูลให้เป็นระบบ และเป็นข้อมูลช่วยตัดสินใจในการแจ้งเตือนประชากรกลุ่มเสี่ยงในพื้นที่ท้ายเขื่อน เพื่อลดผลกระทบที่จะเกิดขึ้นโดยการนำข้อมูลน้ำที่จัดเก็บไว้ในอดีต นำมาใช้เป็นข้อมูลช่วยในการตัดสินใจสำหรับการพยากรณ์กระบวนการคาดคะเนแนวโน้มหรือรูปแบบของการเกิดเหตุการณ์ต่างๆ ในอนาคตที่ได้จะทำให้ทราบว่าแนวทางการวางแผนป้องกันและการรับมือเพื่อลดผลกระทบจะเกิดขึ้น จากที่ผ่านมาพบว่า จังหวัดลำปางประสบปัญหาสภาวะ ในช่วงหลายปีที่ผ่านมา พื้นที่ชุมชนได้รับผลกระทบจากภัยน้ำท่วมเป็นจำนวนมากไม่ว่าจากผลกระทบน้ำท่วมใหญ่ปี 2548 ปี 2552 และ ปี 2554 เนื่องจากระยะทางจากเขื่อน กิ่วลมมายังอำเภอเมืองมีระยะทางแค่ 30 กิโลเมตร ที่ผ่านมามีการแจ้งเตือนภัยล่วงหน้า แต่มีการแจ้งเตือนเวลาที่กระชั้นชิด ไม่พอสำหรับเตรียมการรับมือภัยน้ำท่วมที่จะเกิดขึ้น

ผู้วิจัยจึงมีความสนใจที่จะหาวิธีแก้ปัญหาดังกล่าวขึ้นโดยการศึกษาวิจัยในเบื้องต้นเพื่อหาเทคนิคพยากรณ์ที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยการศึกษาจากงานวิจัยที่เกี่ยวข้อง นำมาทดสอบกับข้อมูลวิเคราะห์และสรุปผลหลายๆ เทคนิควิธีและเลือกใช้วิธีพยากรณ์ที่มีค่าสัมบูรณ์ของความคลาดเคลื่อนน้อยที่สุด ซึ่งหมายถึงเทคนิควิธีดังกล่าวเป็นการพยากรณ์ที่มีความแม่นยำมากที่สุดเพื่อนำไปใช้ในการพัฒนาระบบพยากรณ์ต่อไป

1. วัตถุประสงค์การวิจัย

1. เพื่อศึกษาเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูลที่เหมาะสม
2. เพื่อเปรียบเทียบผลการวิเคราะห์ปริมาณน้ำในเขื่อนโดยใช้การพยากรณ์ด้วยเทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

มนต์ชัย เทียนทอง (2548, น. 547) การพยากรณ์ หมายถึงการคาดการณ์เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่สนใจที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศประกอบการตัดสินใจซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดที่จะต้องทำก่อนการวางแผน หรือการเตรียมการที่จะเริ่มทำอะไรเพื่อความถูกต้องและแม่นยำในการตัดสินใจ การพยากรณ์สามารถแบ่งตามระยะเวลาได้ 3 ประเภทคือ การพยากรณ์ในระยะสั้น การพยากรณ์ระยะกลาง และการพยากรณ์ระยะยาว ในการพยากรณ์ มีเทคนิคของการพยากรณ์ประกอบด้วย การพยากรณ์เชิงปริมาณ เป็นการพยากรณ์โดยอาศัยข้อมูลหรือตัวเลขจากอดีตและพยากรณ์ไปในอนาคต และ การพยากรณ์เชิงคุณภาพ เป็นกลุ่มของวิธีการพยากรณ์ที่อาศัยข้อมูลและวิธีการเชิงคุณภาพ ใช้กับลักษณะของปัญหาที่ไม่มีข้อมูลย้อนหลังหรือมีข้อมูลไม่มากพอ มาใช้สร้างตัวแบบ

ณัฐริน เจริญเกียรติบวร (2549, น. 2) การทำเหมืองข้อมูล คือกระบวนการค้นหาสารสนเทศ หรือข้อความรู้ที่มีอยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์ หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่าง ๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล หารทำเหมืองข้อมูลประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์ เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์และข้อความรู้จากฐานข้อมูลขนาดใหญ่ โดยการทำเหมืองข้อมูลมีขั้นตอนการดำเนินงานหลายขั้นตอนซึ่งต้องอาศัยเทคนิคหรือวิธีการต่าง ๆ เช่น วิธีการจัดกลุ่ม การค้นหาความสัมพันธ์ การพยากรณ์

พยุง มีสัจ (2551, น. 16-23)โครงข่ายประสาทเทียม (Artificial Neural Network: ANN)เป็นการศึกษาระบบการทำงานของเซลล์ประสาทภายในสมองที่ประกอบด้วยเซลล์ประสาท (Neuron) และเส้นประสาทโดยที่เซลล์ประสาทจะเชื่อมต่อกันในรูปแบบโครงข่าย ซึ่งการวิเคราะห์ และประมวลผลข้อมูลของระบบประสาทนั้นจะส่งข้อมูลผ่านระบบโครงข่ายของเซลล์ประสาท และทำงานในลักษณะขนานคือ ทำกิจกรรม หรืองานหลายอย่างได้ในเวลาเดียวกันให้ได้มาซึ่งผลลัพธ์ที่ต้องการ โดยการทำงานของสมองในรูปแบบที่กล่าวมาในข้างต้นนั้นมีความสามารถหลายประการเช่น การสังเกต เรียนรู้ จดจำ ทำซ้ำ และแยกแยะสิ่งต่าง ๆ ซึ่งโครงข่ายประสาทเทียมได้จำลองรูปแบบการทำงาน และโครงสร้างการเชื่อมต่อดังกล่าวมา เพื่อทำให้ระบบคอมพิวเตอร์สามารถทำงานที่สมองทำได้ซึ่งสามารถเพิ่มประสิทธิภาพในการทำงานให้มากขึ้น และช่วยลดข้อผิดพลาดที่เกิดขึ้นจากการทำงาน

พิชญ์สินี ชมพูคำ (2554, น.2-3)เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) คือการหาความสัมพันธ์ของตัวแปรหนึ่งกับตัวแปรอื่น ๆ มักจะใช้เครื่องมือทางสถิติที่เรียกว่า ซึ่งเป้นกระบวนการทั้งหมดที่สร้างความสมการสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป โดยที่ตัวแปรหนึ่งเป็นผลมาจากอีกตัวแปลหนึ่งหรือหลายตัวแปรก็ได้ โดยอาศัยข้อมูลจากการสุ่มตัวอย่างและการอนุมาน

Preis, A. and A. Ostfeld. (2007,pp. 364-375) แบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P)

แบบจำลองต้นไม้เอ็มไฟว์พีเป็นต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่อง และไม่ต่อเนื่อง นอกจากนี้ที่โหนดใบ (Leaf Node) ของต้นไม้จะมีแบบจำลองเชิงเส้นที่ใช้ทำนายกลุ่มของข้อมูลที่เป็นค่าตัวเลข ในการทำนายค่าของกลุ่มข้อมูลจะวิเคราะห์ข้อมูลจากโนหดราก (Root Node) ลงมาจนกระทั่งถึงโหนดใบที่จะได้ค่าคุณลักษณะในการตัดสินใจเลือกเส้นทางจากโหนดแต่ละชั้น

R.O. Duda and P.E Hart,Pattern (1993,pp. 237) เทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (SVM) เป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกัน โดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้นเส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุดในโลกความเป็นจริงนั้น ข้อมูลสองกลุ่มไม่ได้วางตัวในพื้นที่คุณลักษณะ เป็นเพียงสองกลุ่มและแบ่งได้โดยเส้นตรงแต่ข้อมูลอาจจะจับกลุ่มกันในตำแหน่งต่าง ๆ ดังนั้น จึงเป็นปัญหาทำให้ไม่สามารถที่จะใช้สมการซัพพอร์ตเวคเตอร์แมชชีนส์แบบเชิงเส้นได้

วิธีดำเนินการวิจัย

1. ขั้นตอนการดำเนินการวิจัยประกอบด้วย 5ขั้นตอน ดังนี้

1.1 ศึกษาเทคนิคการพยากรณ์ที่เป็นไปได้และเหมาะสมในการนำมาพยากรณ์ปริมาณน้ำในเขื่อนจากเอกสารและงานวิจัยที่เกี่ยวข้องและสรุปผลเลือกเทคนิคที่เหมาะสมจำนวน 4 เทคนิค

1.2 จัดเตรียมข้อมูล โดยจัดเรียงข้อมูลตามช่วงเวลารายวันตามปัจจัยที่เกี่ยวข้องเพื่อนำมาวิเคราะห์ด้วยเทคนิคที่เลือกไว้ในขั้นตอนที่ 1.1

1.3 วิเคราะห์ข้อมูลที่จัดเตรียมด้วยเทคนิคการพยากรณ์ทั้ง 4 เทคนิค

1.4 เปรียบเทียบค่าสัมบูรณ์ของความคลาดเคลื่อนทั้ง 4 เทคนิค และสรุปผลการวิเคราะห์ข้อมูลเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

1.5 สรุปผลการศึกษาและเขียนรายงานการวิจัย

2. เครื่องมือการวิจัย

2.1 โปรแกรมที่ใช้ในการวิเคราะห์ใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์ค่าสถิติการพยากรณ์ด้วยโปรแกรม WEKA

2.2 เทคนิคการพยากรณ์ จำนวน 4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย (Regression Analysis) วิธีและโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) วิธีแบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) และ วิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (SVM)

3. กลุ่มเป้าหมาย

ข้อมูลปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง ตั้งแต่ปี พ.ศ. 2535-2560 จำนวน 25 ปี รวมทั้งหมด 9,300 รายการ

4. สถิติที่ใช้ในการวิจัยได้แก่

สถิติการพยากรณ์ โดยพิจารณาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error, MSE) และค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error, RMSE)

ผลการวิจัย

1. ผลการศึกษาเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง นำมาสรุปเทคนิคการพยากรณ์ปริมาณน้ำที่เหมาะ ได้จำนวน 4 เทคนิควิธี ได้แก่

1.1 วิธีการวิเคราะห์การถดถอย (Regression Analysis) เป็น เป็นวิธีที่ใช้เทคนิคทางสถิติในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระที่มีการกำหนดค่าที่แน่นอนไว้ล่วงหน้าและตัวแปรตามซึ่งเป็นตัวแปรที่เปลี่ยนแปลงไปตามตัวแปรอิสระและมีความสัมพันธ์ของสัมประสิทธิ์ในสมการการถดถอยเป็นแบบเชิงเส้น (นันทชัย กานตานัน, น. 2555)

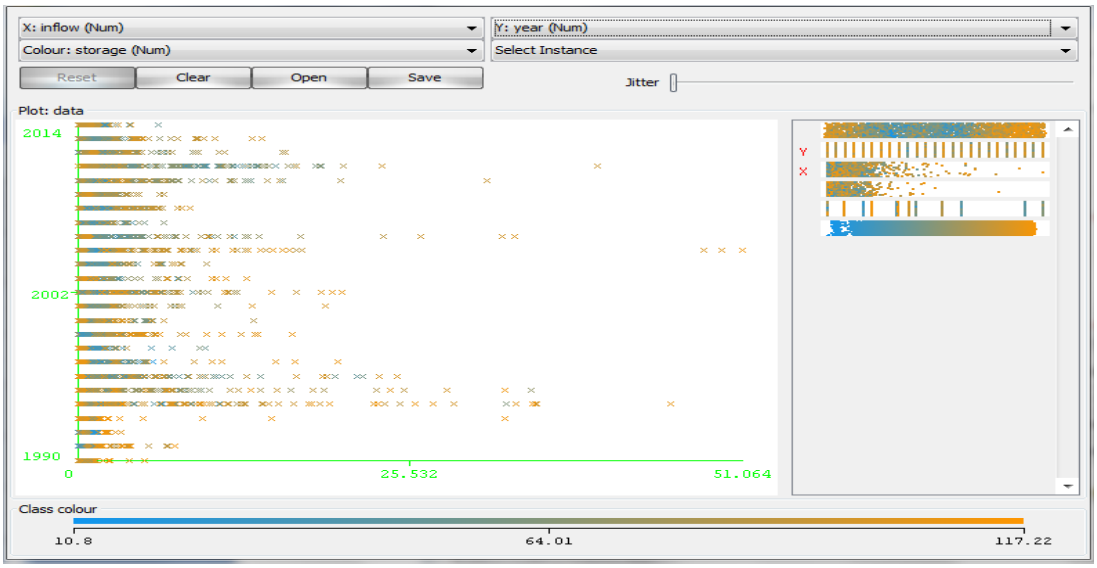
1.2 วิธีและโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นตัวแบบทางคณิตศาสตร์สำหรับจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ โดยจะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้ การจดจำรูปแบบและการสร้างความรู้ใหม่เช่นเดียวกับความสามารถที่มีในสมองมนุษย์โดยทั่วไปค่าน้ำหนักจะถูกนำไปวิเคราะห์ฟังก์ชันถ่ายโอนทำหน้าที่เปรียบเสมือนกระบวนการทำงานในตัวเซลล์ประสาท (Zou, H.F., Xia, G.P., Yang, F.T. and Wang, H.Y,2007)

1.3 วิธีแบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) เป็นวิธีที่ใช้ทำนายข้อมูลที่เป็นตัวเลขพัฒนามาจากต้นไม้ตัดสินใจจะมีความแตกต่างตรงขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลขและค่าอินพุตเป็นค่าต่อเนื่องและไม่ต่อเนื่อง ในการทำนายค่าจะวิเคราะห์ข้อมูลจากโหนดรากจนกระทั่งโหนดที่จะได้ค่าผลลัพธ์ (Preis, A. and A. Ostfeld, 2007)

1.4 วิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (SVM) เป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะออกจากกันโดยจะสร้างเส้นแบ่งที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด (R.O. Duda and P.E Hart, Pattern : 1993)จากผลการศึกษาที่ได้ ผู้วิจัยจะนำเทคนิคที่ได้จากการศึกษาไปพยากรณ์ข้อมูลที่ได้จัดเตรียมไว้ เพื่อเปรียบเทียบและสรุปผล และนำไปใช้ในการดำเนินการในขั้นตอนต่อไป

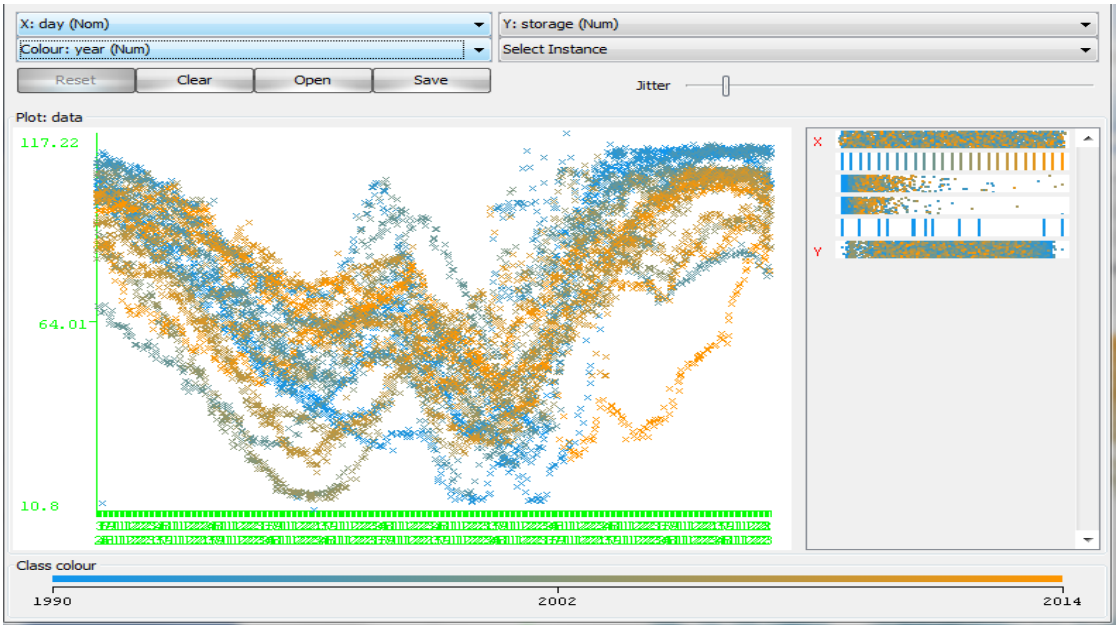
2. ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล

จากรวบรวมข้อมูลตั้งแต่ปี พ.ศ. 2535 – พ.ศ. 2559 รวม 25 ปี จำนวน 9,300 รายการ ตามปัจจัยที่เกี่ยวข้องประกอบด้วย ข้อมูลปริมาณน้ำเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำ และ อัตราการละเหย โดยมีภาพรวมของข้อมูลที่เกี่ยวข้องดังนี้



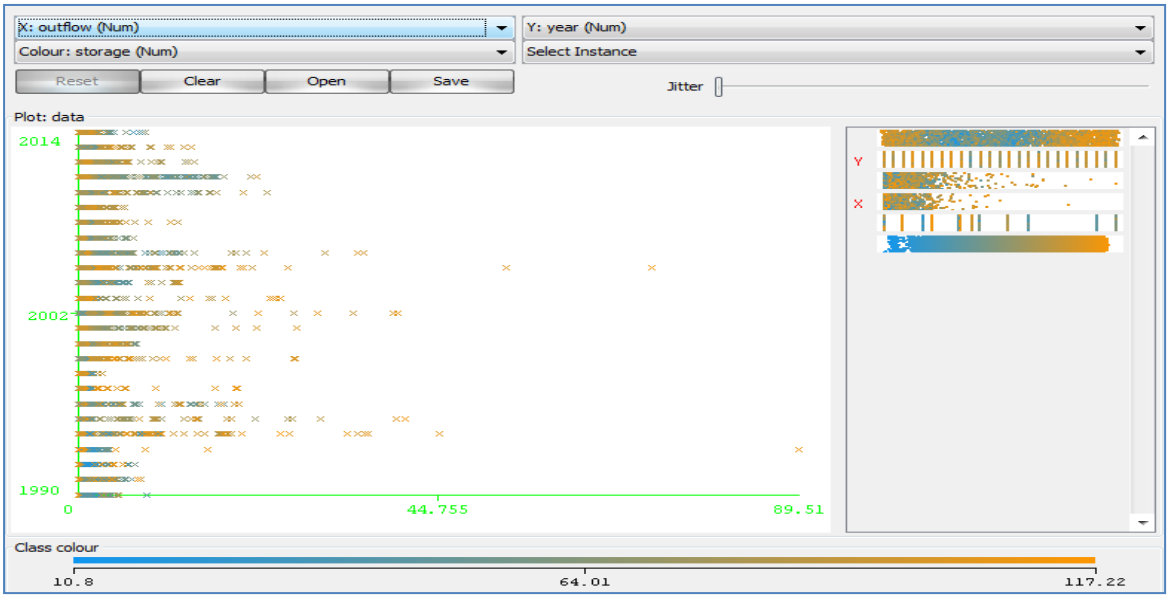
ภาพที่ 1 ข้อมูลภาพรวมปริมาณน้ำไหลเข้าเขื่อนระหว่างปี 1990 ถึงปี 2016

จากภาพที่ 1 แสดงภาพรวมปริมาณน้ำไหลเข้าเขื่อนช่วง 25 ปีที่ผ่านมาจากภาพพบว่าพบว่าปริมาณน้ำแปรผันตามช่วงฤดูกาลปริมาณน้ำในเขื่อนช่วงเดือนสิงหาคม-เดือนมกราคมจะมีปริมาณมาก



ภาพที่ 2 ข้อมูลภาพรวมปริมาณน้ำในเขื่อน ระหว่างปี 1990 ถึงปี 2016

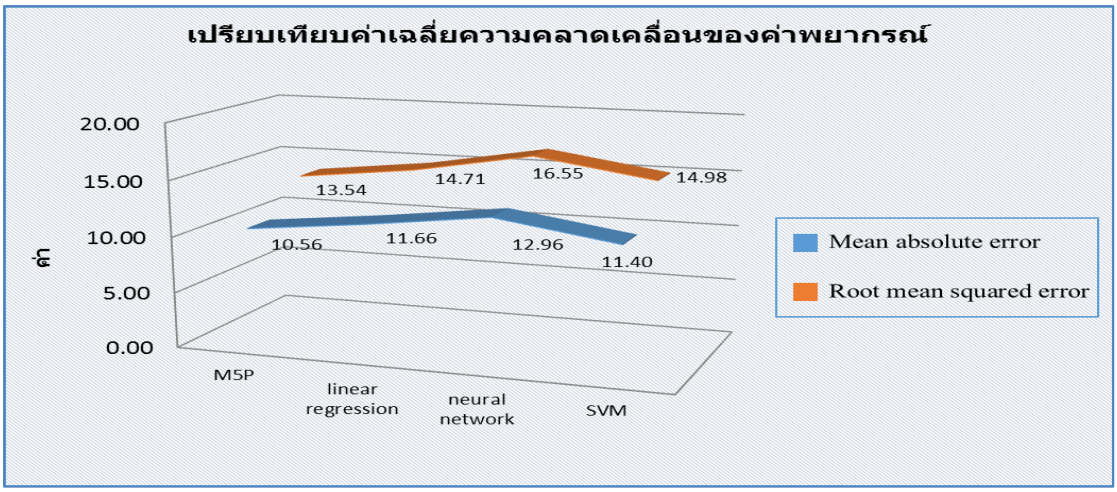
จากภาพที่ 2 แสดงภาพรวมปริมาณน้ำในเขื่อนช่วง 25 ปีที่ผ่านมาจากภาพพบว่าปริมาณน้ำแปรผันตามช่วงฤดูกาลปริมาณน้ำในเขื่อนช่วงเดือนสิงหาคม-เดือนมกราคมจะมีปริมาณมาก



ภาพที่ 3 ข้อมูลภาพรวมปริมาณการปล่อยน้ำออกจากเขื่อนระหว่างปี 1990 ถึงปี 2016

จากภาพที่ 3 แสดงภาพรวมปริมาณการปล่อยน้ำออกจากเขื่อนในช่วง 25 ปีที่ผ่านมาจากภาพพบว่าปริมาณการปล่อยน้ำจะขึ้นกับปริมาณน้ำไหลเข้าเขื่อน

ผู้วิจัยได้ทำการเปรียบเทียบผลการวิเคราะห์ข้อมูลปริมาณน้ำในเขื่อนด้วยเทคนิคเหมืองข้อมูล 4 วิธีโดยนำค่าสัมบูรณ์ของความคลาดเคลื่อนมาทำการเปรียบเทียบโดยใช้โปรแกรม WEKA แสดงดังภาพที่ 4



ภาพที่ 4 แสดงผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนจาก 4 วิธี

จากภาพที่ 4 พบว่า ค่าสัมบูรณ์ของความคลาดเคลื่อนโดยวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.56 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อน ทั้งนี้เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าคลาดเคลื่อนจากน้อยไปหามากตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พีมีค่าความคาดเคลื่อนเท่ากับ10.56 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคาดเคลื่อนเท่ากับ 10.84 วิธีวิเคราะห์การถดถอยมีค่าความคาดเคลื่อนเท่ากับ 11.12 และ วิธีโครงข่ายประสาทเทียม มีค่าความคาดเคลื่อนเท่ากับ 12.53

อภิปรายผลการวิจัย

การศึกษาเทคนิคการพยากรณ์ที่เหมาะสมในการนำมาวิเคราะห์ปริมาณน้ำในเขื่อน โดยใช้ 4 เทคนิค ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error, MSE) และ ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error, RMSE) ของวิธีโครงข่ายประสาทเทียม 12.96 %และ 16.55 % วิธีวิเคราะห์การถดถอย 11.66%และ 14.71 วิธีแบบจำลองต้นไม้เอ็มไฟว์พี 10.56 %และ 13.54 % วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ 11.40%และ 14.98 สรุปผลการวิเคราะห์พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน และ ค่าเฉลี่ยของรากที่สองของกำลังสองของความเป็นวิธีที่มีค่าความคลาดเคลื่อนต่ำสุด เหมาะสำหรับนำไปพัฒนาระบบเพื่อใช้ในการพยากรณ์ต่อไป

ข้อเสนอแนะ

งานวิจัยนี้ได้ทำการศึกษาเทคนิคการพยากรณ์เพียง 4 เทคนิค หากมีการศึกษาเทคนิคการพยากรณ์อื่นที่เหมาะสมนำมาศึกษาเปรียบเทียบอาจจะทำให้ได้ค่าการพยากรณ์ที่แตกต่างกันในแต่ละเทคนิค ทั้งนี้ข้อมูลที่นำมาใช้ในการพยากรณ์ปริมาณน้ำในเขื่อนผู้วิจัยใช้ข้อมูลย้อนหลัง 25 ปี ที่เก็บโดยสำนักงานเขื่อนกิ่วลม จังหวัดลำปาง โดยแยกข้อมูลพยากรณ์เป็นรายเดือนเนื่องจากเขื่อนกิ่วลมเป็นเขื่อนที่มีขนาดเล็กระดับน้ำมีการแปรผันอย่างรวดเร็วในการทำวิจัยที่พยากรณ์ระดับน้ำในเขื่อนอื่นที่มีลักษณะงานวิจัยที่ใกล้เคียงกันอาจนำไปปรับใช้ให้เหมาะสมกับการพยากรณ์ หรือการใช้ข้อมูลรวมกันต่อไป

เอกสารอ้างอิง

มนต์ชัย เทียนทอง. สถิติและวิธีการวิจัยทางเทคโนโลยีสารสนเทศ. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, พ.ศ. 2548.

พิชญ์สินี ชมพูคำ . (มปป.) การวิเคราะห์การถดถอย. สืบค้นจาก http://www.hosting.cmru.ac.th/phitsinee/regression/index.php

พยุง มีสัจ.ระบบฟัซซี่และโครงข่ายประสาทเทียม, วิทยานิพนธ์, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพฯ,พ.ศ. 2551.

ณัฐริน เจริญเกียรติบวร. (2549). การใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับ

นันทชัย กานตานันทะ,(2555),การพยากรณ์ด้วยวิธีการพยากรณ์เชิงสาเหตุ, วารสารวิศวกรรมศาสตร์4(1) : 35-48.

Joyce, B.R., and Weil, M. 2000. Models of Teaching.(6th ed). Massachusetts: Allyn & Bacon.Laudon, K.C.

Laudon, J. P. (2001). Essentials of management information systems: Organization and Technology in the enterprise. (4th Ed). Upper Saddle River, NJ: Prentice Hall.

Steiner, E. (1988). Methodology of theory construction. Sydney :Educology Research Associates.

Preis, A. and A. Ostfeld. 2007. A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. Journal of Hydrology (2008) 349, 364– 375Zou,

H.F., Xia, G.P., Yang, F.T. and Wang, H.Y., 2007, An investigation and comparison of artificial neural Network and time series models for Chinese food grain price forecasting, 70: 2913-2923.

ส่งตีพิมพ์ที่วารสารวิทยาการจัดการสมัยใหม่ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏลำปาง * อยู่ในระหว่างรอการตอบรับ

การพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยใช้เทคนิคเหมืองข้อมูล

Forecasting Daily Discharge in Kievlom Dam Using Data Mining Techniques

วีรศักดิ์ ฟองเงิน [1*] ผศ.ดร. วรปภา อารีราษฎร์ [2] และ ดร. เผด็จ พรหมสาขา ณ สกลนคร [3]

คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏลำปาง[1*] คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม[2,3]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาเทคนิคเหมืองข้อมูลที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อน และ 2) เปรียบเทียบผลการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้บริหารจัดการข้อมูลน้ำที่มีอยู่ที่เป็นปัจจัยต่อการเปลี่ยนแปลงระดับน้ำประกอบด้วย 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 วิธีแบบจำลองต้นไม้เอ็มไฟว์พี และวิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ โดยเลือกจากงานวิจัยที่เกี่ยวข้องที่มีการใช้งานมากที่สุด และ 2) ผลการเปรียบเทียบการพยากรณ์ปริมาณน้ำรายเดือนในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูลทั้ง 4 เทคนิค พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.58 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบเพื่อการพยากรณ์น้ำในเขื่อน ทั้งนี้เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิค พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ วิธีวิเคราะห์การถดถอย และวิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 10.58, 19.10, 20.16 และ 21.65 ตามลำดับ

คำสำคัญ: การพยากรณ์ เหมืองข้อมูล เขื่อนกิ่วลม โครงข่ายประสาทเทียม แบบจำลองต้นไม้เอ็มไฟว์พี วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ วิธีการวิเคราะห์การถดถอย

ABSTRACT

This research aims to 1) study the appropriate data mining techniques to forecasting the amount of water in the dam, and 2) compare the prediction of monthly water in Kiew Lom Dam, Lampang, using data mining techniques. This research led information on the factors affecting the level of water in town. The amount of water flowing into the dam, the amount of water in the dam, emissions and evaporated water. The Laehi Data were collected daily from the year 2535 - 2559, including 25 years by 9,125 record to use for forecasting with Techniques.

The results showed that: 1) appropriate data mining techniques to forecast the water level in the dam consists of four techniques including Regression (Regression Analysis) method, ANN (Artificial Neural Network), M5P(M5P Model Tree), and Technical SVM (Support Vector Machine) and 2) compare the prediction of water monthly in Kiew Lom dam, Lampang, using data mining techniques, the four techniques that way. M5P With absolute as low as 10.56 and was the most appropriate way to develop a system for forecasting the water in the dam. The techniques are compared with each other according to the RMSE (Root Mean Square Error) of the data. The comparison results show that M5P. The move is expected to equal the 10.58 SVM. The move is expected to equal the 19.10 regression analysis. The move is expected to equal the 20.00 and ANN. The move is expected to equal 21.65.

Keywords: Forecasting, Data Mining, KiewLom Dam, Support Vector Machine, M5P Model Tree, Artificial Neural Network, Regression Analysis

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

1. วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเทคนิคเหมืองข้อมูลที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อน
2. เพื่อเปรียบเทียบผลการพยากรณ์ปริมาณน้ำรายเดือนในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

มนต์ชัย เทียนทอง (2548, น.547) อธิบายว่า การพยากรณ์ หมายถึง การคาดการณ์เกี่ยวกับลักษณะหรือแนวโน้มของสิ่งที่สนใจที่จะเกิดขึ้นในอนาคต เพื่อใช้เป็นสารสนเทศประกอบการตัดสินใจ ซึ่งการพยากรณ์จะต้องดำเนินการเป็นส่วนแรกสุดที่จะต้องทำก่อนการวางแผน หรือการเตรียมการที่จะเริ่มทำอะไรเพื่อความถูกต้องและแม่นยำในการตัดสินใจ การพยากรณ์สามารถแบ่งตามระยะเวลาได้ 3 ประเภทคือ 1) การพยากรณ์ระยะสั้น 2) การพยากรณ์ระยะกลาง และ 3) การพยากรณ์ระยะยาว ในการพยากรณ์ มีเทคนิคของการพยากรณ์ประกอบด้วย 1) การพยากรณ์เชิงปริมาณ เป็นการพยากรณ์โดยอาศัยข้อมูลหรือตัวเลขจากอดีตและพยากรณ์ไปในอนาคต และ 2) การพยากรณ์เชิงคุณภาพ เป็นกลุ่มของวิธีการพยากรณ์ที่อาศัยข้อมูลและวิธีการเชิงคุณภาพ ใช้กับลักษณะของปัญหาที่ไม่มีข้อมูลย้อนหลังหรือมีข้อมูลไม่มากพอ มาใช้สร้างตัวแบบ

ณัฐริน เจริญเกียรติบวร (2549, น.2) อธิบายว่า เหมืองข้อมูล หมายถึง กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ ขั้นตอนการทำเหมืองข้อมูลประกอบด้วยการคัดลอกข้อมูล การเตรียมข้อมูล การแปลงรูปแบบข้อมูล การทำเหมืองข้อมูล และ

การวิเคราะห์ข้อมูล ในงานเหมืองข้อมูลประกอบด้วยเทคนิคต่าง ๆ ได้แก่การจัดหมวดหมู่ การประเมินค่า การทำนาย การจัดกลุ่ม และ การรวมตัว

พยุง มีสัจ (2551, น.16-23) 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, น.2-3) อธิบายว่า เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) คือการหาความสัมพันธ์ของตัวแปรหนึ่งกับตัวแปรอื่น ๆ มักจะใช้เครื่องมือทางสถิติที่เรียกว่า ซึ่งเป็นกระบวนการทั้งหมดที่สร้างความสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป โดยที่ตัวแปรหนึ่งเป็นผลมาจากอีกตัวแปรหนึ่งหรือหลายตัวแปรก็ได้ โดยอาศัยข้อมูลจากการสุ่มตัวอย่างและการอนุมาน

Preis, A. and A. Ostfeld. (2007,pp. 364-375) อธิบายว่า แบบจำลองต้นไม้เอ็มไฟว์พี (Model Tree: M5P) เป็นต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่อง และไม่ต่อเนื่อง นอกจากนี้ที่โหนดใบ (Leaf Node) ของต้นไม้จะมีแบบจำลองเชิงเส้นที่ใช้ทำนายกลุ่มของข้อมูลที่เป็นค่าตัวเลข ในการทำนายค่าของกลุ่มข้อมูลจะวิเคราะห์ข้อมูลจากโหนดราก (Root Node) ลงมาจนกระทั่งถึงโหนดใบที่จะได้ค่าคุณลักษณะในการตัดสินใจเลือกเส้นทางจากโหนดแต่ละชั้น

R.O. Duda and P.E Hart,Pattern (1993,p. 237) อธิบายว่า เทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (SVM) เป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้นเส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด ในโลกแห่งความเป็นจริงนั้น ข้อมูลสองกลุ่มไม่ได้วางตัวในพื้นที่คุณลักษณะ เป็นเพียงสองกลุ่มและแบ่งได้โดยเส้นตรง แต่ข้อมูลอาจจะจับกลุ่มกันในตำแหน่งต่าง ๆ ดังนั้น 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

2. เครื่องมือการวิจัย

2.1 โปรแกรมที่ใช้ในการวิเคราะห์ใช้โปรแกรมคอมพิวเตอร์ในการวิเคราะห์ค่าสถิติการพยากรณ์ ด้วยโปรแกรม WEKA

2.2 เทคนิคการพยากรณ์ จำนวน 4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย (Regression Analysis M5P) และ วิธีเทคนิคซัพพอร์ตเวคเตอร์แมชชีนส์ (Support Vector Machine : SVM)

3. ข้อมูลที่ใช้ในการวิจัย

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

4. สถิติที่ใช้ในการวิจัยได้แก่

สถิติการพยากรณ์ โดยพิจารณาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error, MSE) และ ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (Root Mean Square Error, RMSE)

5. ผลการศึกษาเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง นำมาสรุปเทคนิคการพยากรณ์ปริมาณน้ำที่เหมาะสมได้จากงานวิจัยที่เกี่ยวข้องที่มีการนำมาทดสอบมากที่สุดจำนวน 4 เทคนิควิธี ได้แก่

5.1.1 วิธีการวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีที่ใช้การศึกษาความสัมพันธ์ระหว่างสอง ตัวแปร ตัวแปรหนึ่งเรียกว่าตัวแปรต้นหรือตัวแปรอิสระ(Independent variable) แทนด้วย X อีกตัวแปรหนึ่งเรียกว่าตัวแปรตาม(Dependent variable) แทนด้วย Y เป็นการดูความสัมพันธ์ว่าถ้า ตัวแปรอิสระเปลี่ยนไปแล้วตัวแปรตามเปลี่ยนไปด้วยหรือไม่

5.1.2 วิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นวิธีที่จำลองโมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (bioelectric network) ใน

สมอง ซึ่งประกอบด้วย เซลล์ประสาท (neurons) และ จุดประสานประสาท (synapses) ตามโมเดลนี้ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน

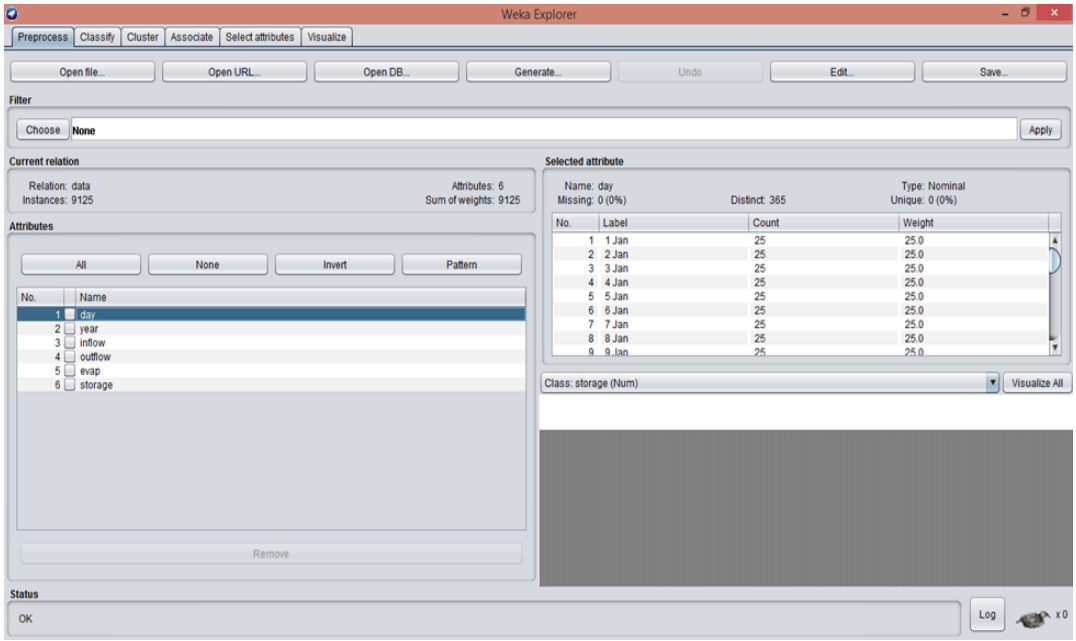
5.1.3 วิธีแบบจำลองต้นไม้เอ็มไฟว์พี (M5P Model Tree) เป็นวิธีที่จำลองต้นไม้ที่ใช้ทำนายผลข้อมูลที่เป็นตัวเลข ซึ่งพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) แต่จะมีความแตกต่างกัน ในขั้นตอนการเลือกโหนดในแต่ละชั้นของต้นไม้ และค่าคุณลักษณะเอาต์พุตเป็นค่าตัวเลข แต่ค่าคุณลักษณะอินพุตนั้นจะเป็นไปได้ทั้งค่าต่อเนื่องและไม่ต่อเนื่อง

5.1.4 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์ (Support Vector Machine : SVM) เป็นวิธีที่สร้างสมการที่ใช้ในการจำแนกค่าคุณลักษณะของสองกลุ่มที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมา และเพื่อให้ทราบว่าเส้นตรงที่แบ่งสองกลุ่มออกจากกันนั้น เส้นตรงใดเป็นเส้นตรงที่ดีที่สุดโดยเส้นตรงนั้นก็จะถูกเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น จะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด

จากผลการศึกษาที่ได้ ผู้วิจัยจะนำเทคนิคที่ได้จากการศึกษาไปพยากรณ์ข้อมูลที่ได้จัดเตรียมไว้ เพื่อเปรียบเทียบและสรุปผล และนำไปใช้ในการดำเนินการในขั้นตอนต่อไป

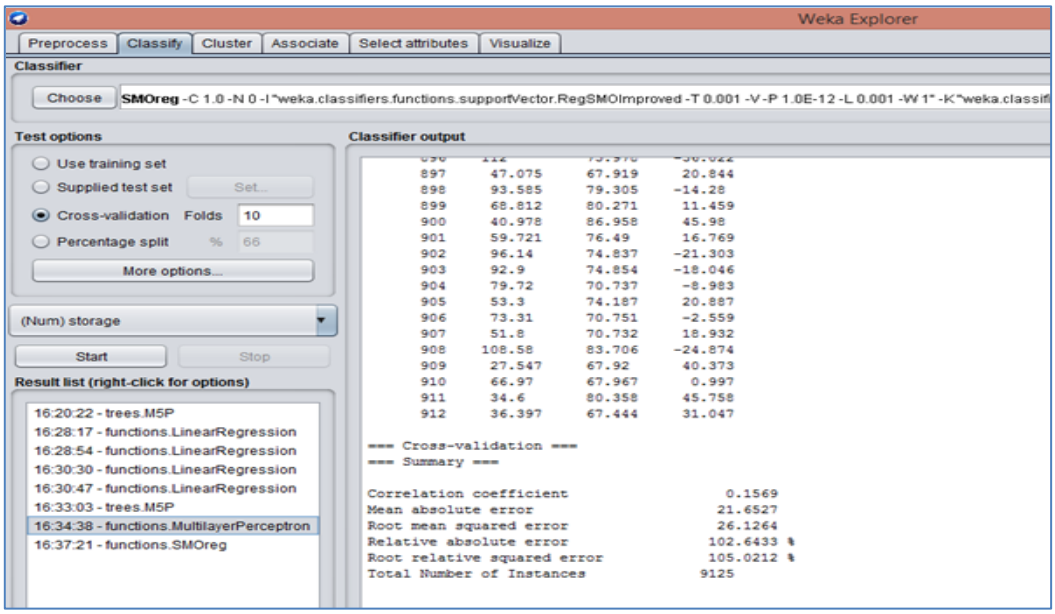
6. ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล

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

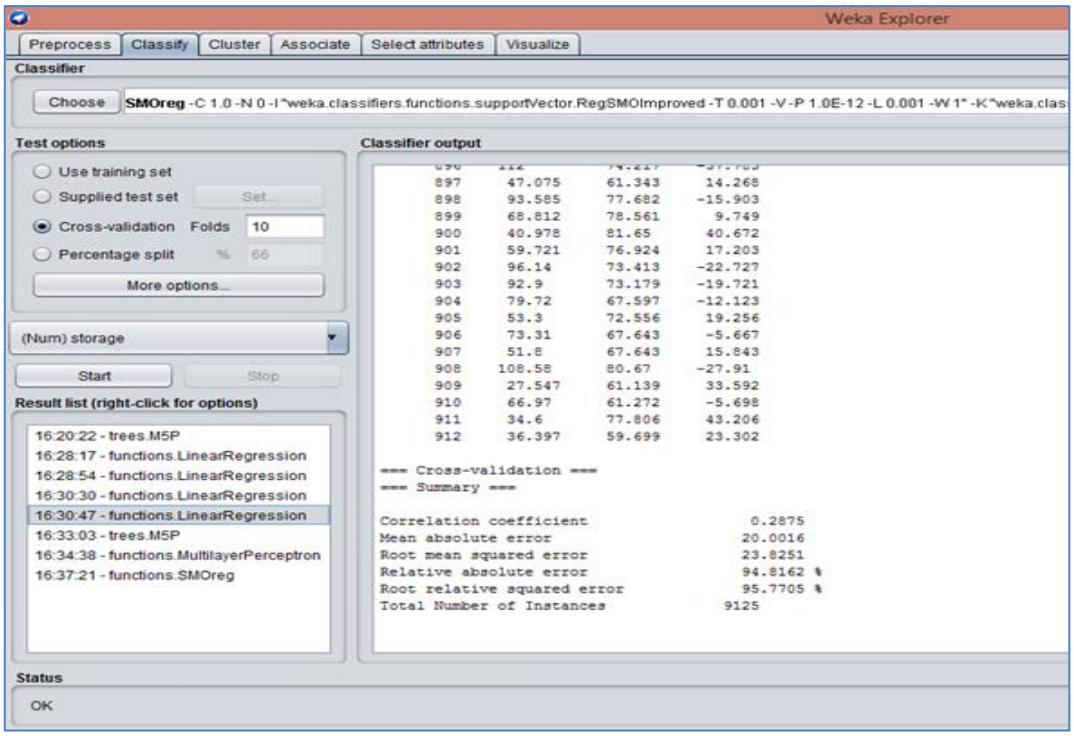


ภาพที่ 1 ข้อมูลภาพรวมปัจจัยที่เกี่ยวข้องสำหรับนำมาพยากร์ปริมาณน้ำเขื่อนกิ่วลม ระหว่างปี 2535 ถึงปี 2559

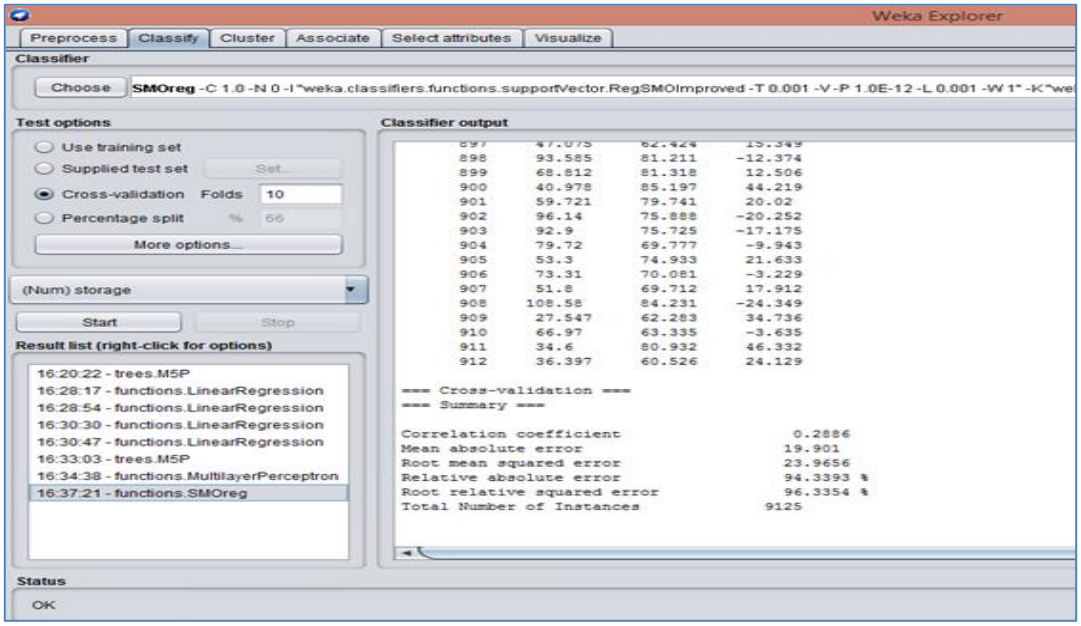
ผู้วิจัยได้นำข้อมูลจากปัจจัยที่เกี่ยวข้องประกอบด้วย วันที่ ปี ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในปล่อยออกจากเขื่อน อัตราการละเหย ปริมาณน้ำในเขื่อน ทั้งหมด 9,125 ระเบียน เป็นข้อมูลย้อนหลังรายวัน จำนวน 25 ปี ลงในโปรแกรม WEKA ทดสอบประสิทธิภาพของโมเดลเพื่อทำผลที่ได้ให้มีความน่าเชื่อถือ ผู้วิจัยได้มีการวัดประสิทธิภาพด้วยวิธี 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 ปรากฏดังนี้



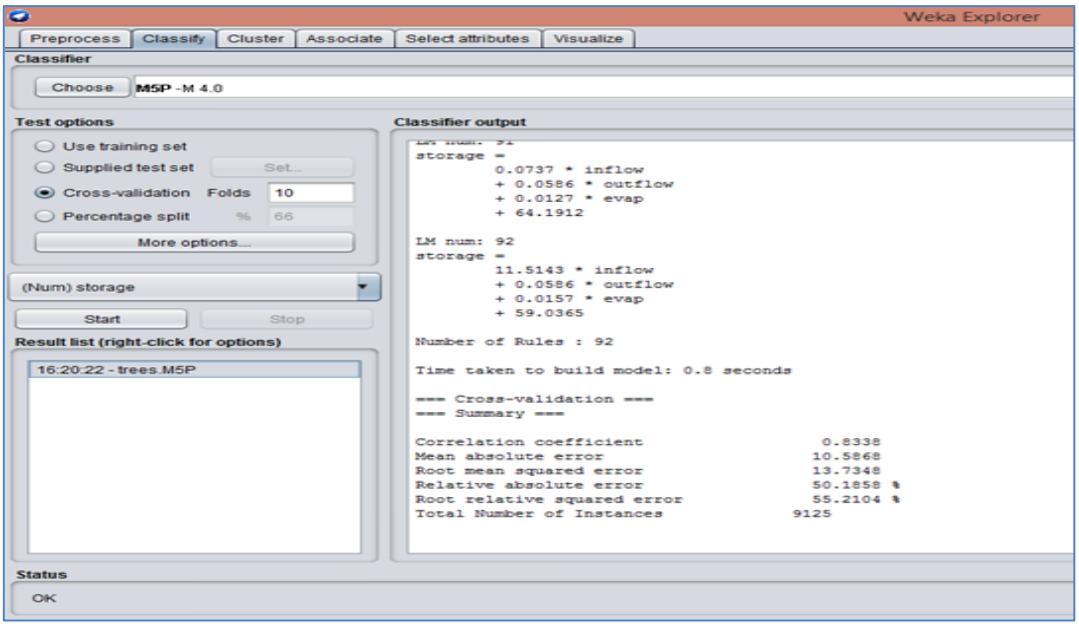
ภาพที่ 2 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีโครงข่ายประสาทเทียม



ภาพที่ 3 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีการวิเคราะห์การถดถอย



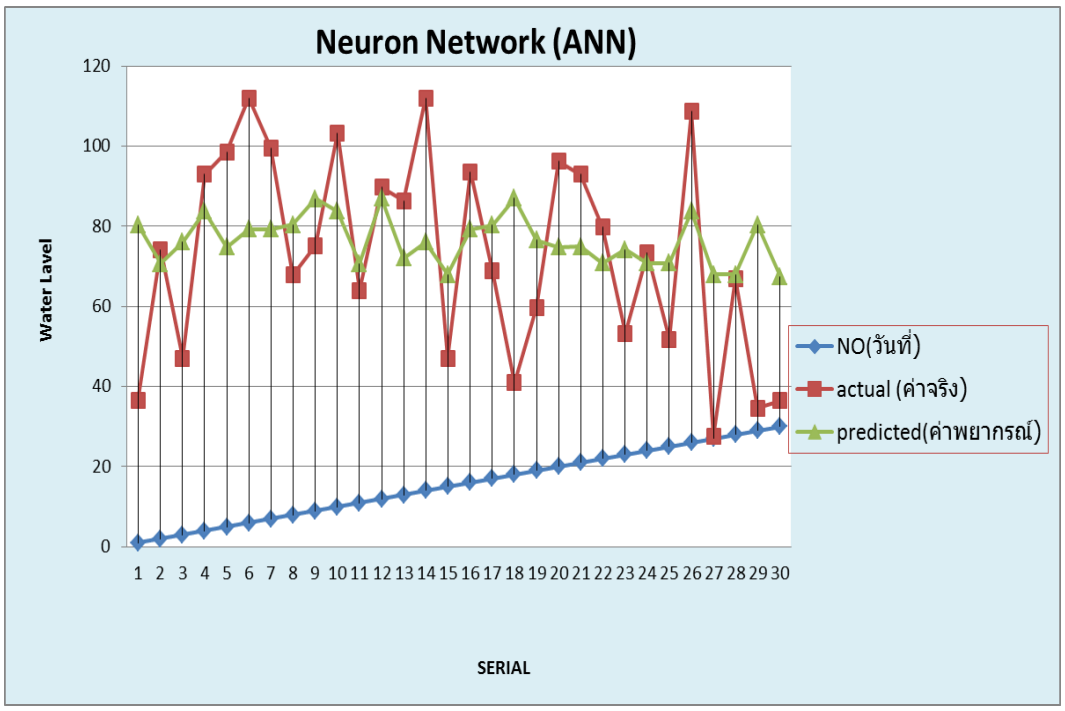
ภาพที่ 4 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีซัพพอร์ตเวคเตอร์แมชชีนส์



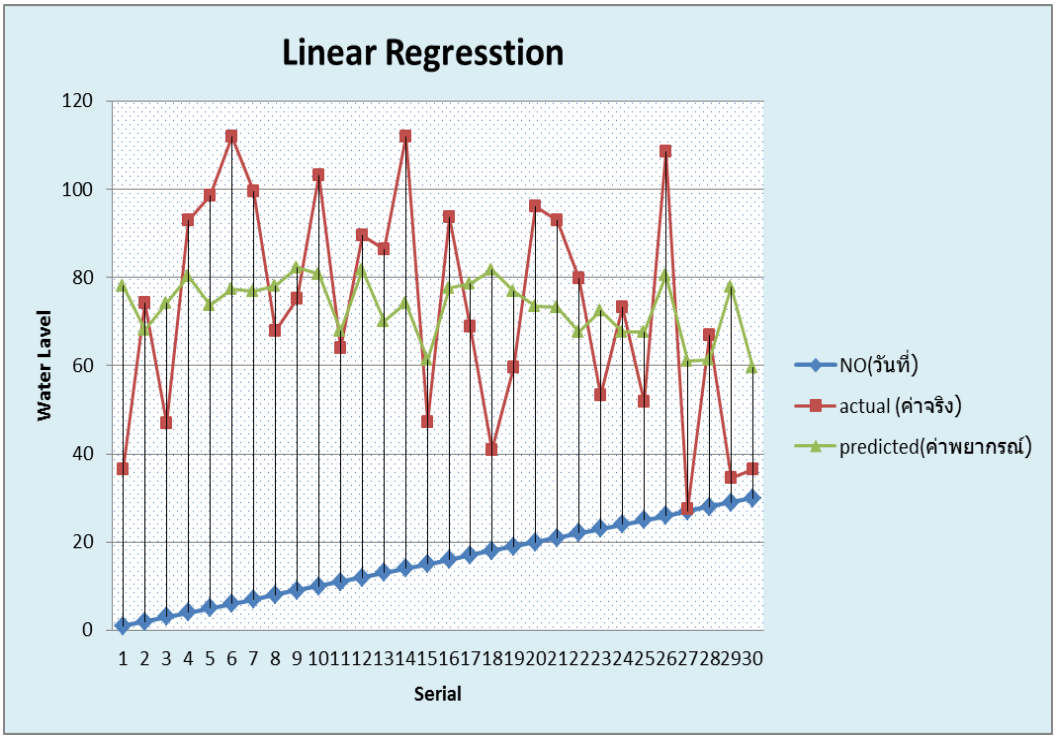
ภาพที่ 5 ผลการวิเคราะห์ความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี

จากภาพที่ 2-5 แสดงผลการวิเคราะห์หาประสิทธิภาพของแต่ละโมเดล พบว่า ค่าสัมบูรณ์ของความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.5868 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อน ทั้งนี้ เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าความคลาดเคลื่อนจากน้อยสุดไปหามากสุดตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนเท่ากับ 10.5868 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคลาดเคลื่อนเท่ากับ 19.901 วิธีวิเคราะห์การถดถอยมีค่าความคลาดเคลื่อนเท่ากับ 20.0016 และ วิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 21.6527 ตามลำดับ

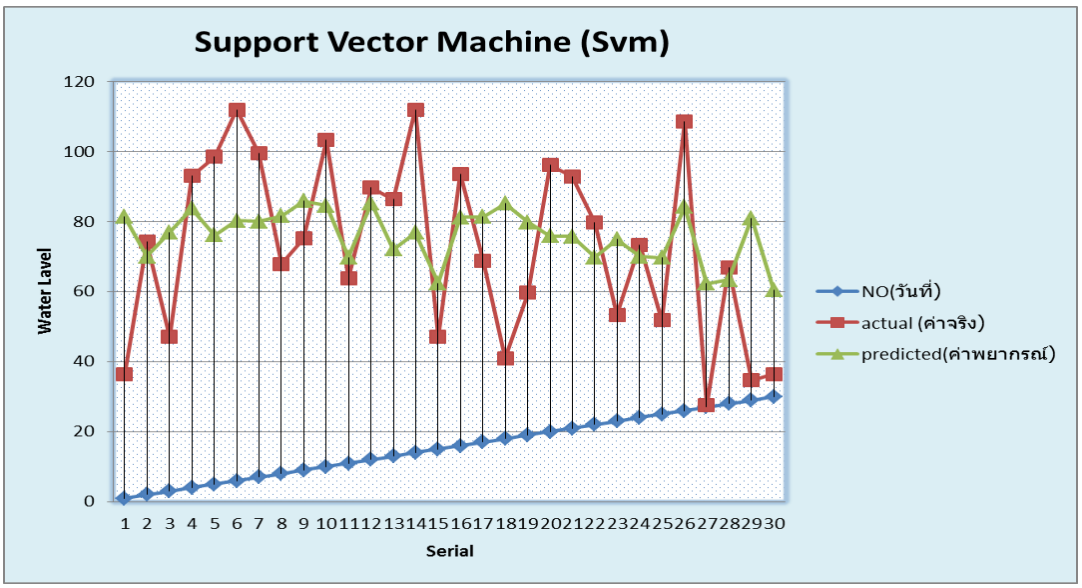
จากผลการวิเคราะห์หาความคลาดเคลื่อนแต่ละโมเดล ผู้วิจัยได้สร้างแบบจำลองเพื่อให้เห็นความแตกต่างความคลาดเคลื่อนระหว่าง ค่าปริมาณน้ำจริง เปรียบเทียบกับค่าที่แต่ละโมเดลมีการพยากรณ์ได้ โดยนำข้อมูลของปริมาณน้ำเดือนล่าสุดมาสร้างแบบจำลอง กับ เทคนิคการพยากรณ์ 4 วิธีได้ผลดังนี้



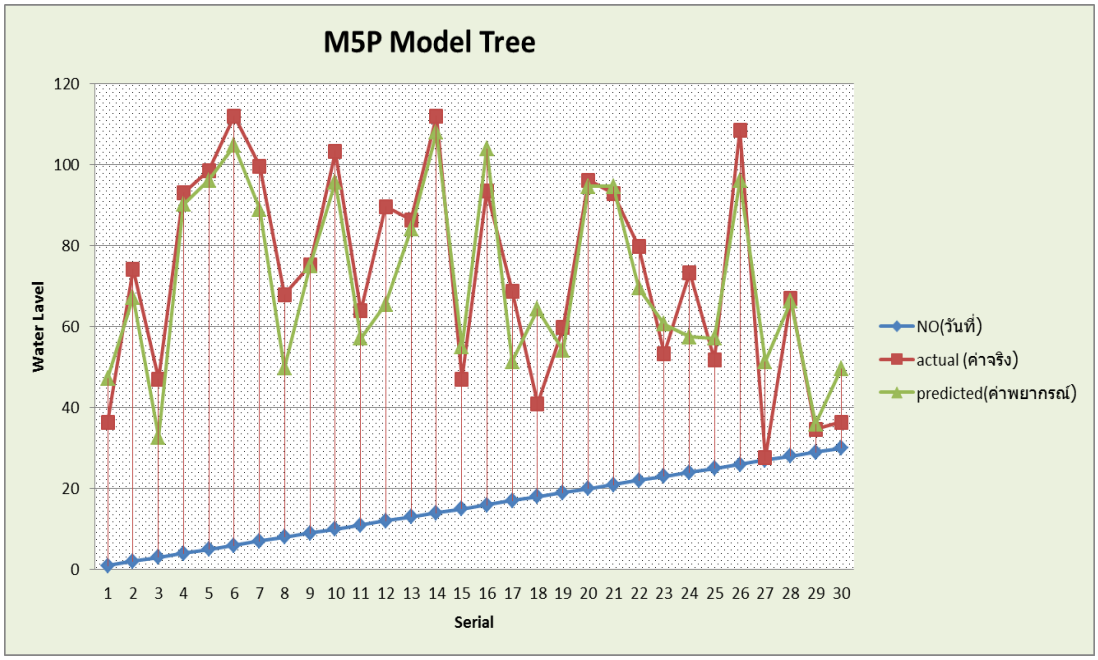
ภาพที่ 6 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีโครงข่ายประสาทเทียม



ภาพที่ 7 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีการวิเคราะห์การถดถอย

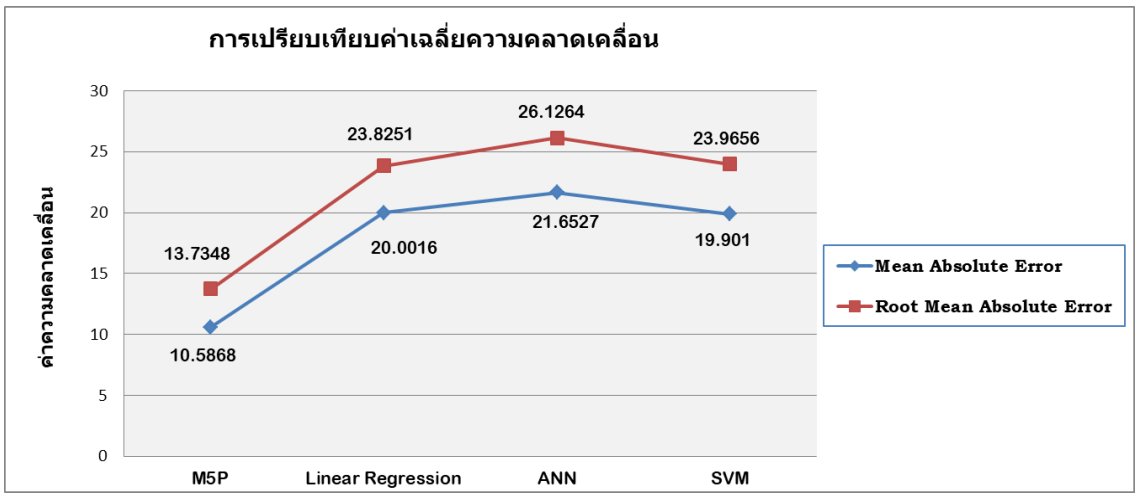


ภาพที่ 8 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีซัพพอร์ตเวคเตอร์แมชชีนส์



ภาพที่ 9 แบบจำลองผลการวิเคราะห์ค่าความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี

จากภาพที่ 6-9 เป็นแบบจำลองผลการเปรียบเทียบค่าปริมาณน้ำค่าจริง กับ ค่าที่มีการพยากรณ์ แต่ละเทคนิควิธี พบว่า วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนต่ำสุด จากภาพที่ 9 เส้นแสดงปริมาณน้ำจริงเปรียบเทียบกับค่าพยากรณ์แนวโน้มใกล้เคียงกันที่สุด เหมาะสำหรับนำไปใช้ในการพัฒนาระบบพยากรณ์น้ำและเมื่อนำแต่ละวิธีมาเปรียบเทียบในภาพรวมเพื่อเลือกโมเดลที่ดีที่สุดจากค่าความคลาดเคลื่อนแสดงดังภาพที่ 10



ภาพที่ 10 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนจาก 4 วิธี

จากผลการเปรียบเทียบการหาค่าเฉลี่ยคลาดเคลื่อนทั้ง 4 วิธี พบว่า ค่าสัมบูรณ์ของความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุด ที่ 10.5868 เป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบพยากรณ์น้ำในเขื่อนกิ่วลมต่อไป ทั้งนี้ เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคมีค่าคลาดเคลื่อนจากน้อยไปหามากตามลำดับ ดังนี้ วิธีแบบจำลองต้นไม้เอ็มไฟว์พี มีค่าความคลาดเคลื่อนเท่ากับ 10.5868 วิธีซัพพอร์ตเวคเตอร์แมชชีนส์มีค่าความคลาดเคลื่อนเท่ากับ 19.901 วิธีวิเคราะห์การถดถอยมีค่าความคลาดเคลื่อนเท่ากับ 20.0016 และ วิธีโครงข่ายประสาทเทียม มีค่าความคลาดเคลื่อนเท่ากับ 21.6527

ผลการวัดความคลาดเคลื่อนวิธีแบบจำลองต้นไม้เอ็มไฟว์พีเป็นวิธีที่มีค่าความคลาดต่ำสุดเหมาะสมเป็นวิธีนำไปพัฒนาระบบพยากรณ์ปริมาณน้ำในเขื่อน เมื่อได้ระบบผู้วิจัยหวังว่าระบบพยากรณ์ที่พัฒนาขึ้นประกอบด้วยการพยากรณ์ปริมาณน้ำรายวัน รายสัปดาห์ และ รายเดือน บุคคลทั่วไปสามารถตรวจสอบระดับน้ำแบบเวลาจริงและค่าที่พยากรณ์ล่วงหน้าแบบออนไลน์ได้ จะทำให้มีการเตรียมพร้อมสำหรับการรับมือ กับ ผลกระทบจากภัยน้ำท่วมที่จะเกิดขึ้นได้ทันท่วงที และ สามารถลดงบประมาณจากภาครัฐจากการจัดซื้ออุปกรณ์สำหรับเตือนภัย

7. สรุปและอภิปรายผล

การศึกษาเทคนิคการพยากรณ์ที่เหมาะสมในการนำมาวิเคราะห์ปริมาณน้ำในเขื่อน โดยใช้ 4 เทคนิค ผลการเปรียบเทียบเทคนิคการพยากรณ์ปริมาณน้ำในเขื่อน โดยใช้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean Square Error, MSE) และ ค่าเฉลี่ยของรากที่สองของกำลังสองของความคลาดเคลื่อน (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 ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน และ ค่าเฉลี่ยของรากที่สองของความคลาดเคลื่อนต่ำสุดซึ่ง เหมาะสำหรับนำไปพัฒนาระบบเพื่อใช้ในการพยากรณ์ต่อไป

ระบบพยากรณ์ปริมาณปริมาณน้ำในเขื่อนที่จะพัฒนาขึ้นโดยนำวิธีแบบจำลองต้นไม้เอ็มไฟว์พีซึ่งมาค่าความคลาดเคลื่อนต่ำสุด จะสามารถพยากรณ์ปริมาณน้ำล่วงหน้า รายวัน รายสัปดาห์ และ รายเดือน ระบบพยากรณ์จะ

ช่วยในการแจ้งเตือนให้ทันท่วงทีและช่วยลดผลกระทบกับบ้านเรือน และ ทรัพย์สินที่จะเกิดขึ้น รวมทั้งเป็นการช่วยประหยัดงบประมาณในการจัดซื้ออุปกรณ์สำหรับเตือนภัยน้ำท่วมต่อไป

8. ข้อเสนอแนะ

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

เอกสารอ้างอิง

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

พิชญ์สินี ชมพูคำ . (2551). การวิเคราะห์การถดถอย. [Online] Available: http://www.hosting.cmru.ac.th/phitsinee/regression/index.php ค้นเมื่อ [2559, กันยายน 30].

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

ณัฐริน เจริญเกียรติบวร. (2549). การใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับ

นันทชัย กานตานันทะ.(2555). การพยากรณ์ด้วยวิธีการพยากรณ์เชิงสาเหตุ, วารสารวิศวกรรมศาสตร์4(1): 35-48.

Joyce, B.R., and Weil, M. (2000) . Models of Teaching. 6th ed. Massachusetts: Allyn & Bacon.Laudon, K.C.

Laudon, J. P. (2001). Essentials of management information systems: Organization and Technology in the enterprise. 4th Ed. Upper Saddle River, NJ: Prentice Hall.

Steiner, E. (1988). Methodology of theory construction. Sydney :Educology Research Associates.

Preis, A. and A. Ostfeld. (2007) . A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. Journal of Hydrology (2008) 349, 364– 375Zou,

H.F., Xia, G.P., Yang, F.T. and Wang, H.Y. (2007). An investigation and comparison of artificial neural Network and time series models for Chinese food grain price forecasting, 70: 2913-2923.