สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาโดยใช้การเรียนรู้ของเครื่องจักร

ESTIMATION OF WATER LEVEL IN CHAO PHRAYA RIVER BY MACHINE LEARNING TECHNIQUES

T146469

โดย

สิริเพ็ญ จึงเจริญรัตน์

SIRIPEN JUNGJAREANTRAT

อาจารย์ที่ปรึกษา

ผศ.ดร. กิตติ์สุชาต พสุภา

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เลขหมู่.....
เลขทะเบียน 146469
วัน.เดือน.ปี 23 มิ.ย. 2560

b.1298419798
1.....

รายงานนี้เป็นส่วนหนึ่งของวิชาการศึกษาอิสระ 2

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ESTIMATION OF WATER LEVEL IN CHAO PHRAYA RIVER BY MACHINE LEARNING TECHNIQUES

SIRIPEN JUNGJAREANTRAT

A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS OF THE COURSE

INDEPENDENT STUDY 2

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2016

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองการศึกษาอิสระ 2 (Independent Study 2)

เรื่อง

การพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาโดยใช้การเรียนรู้ของเครื่องจักร

Estimation of Water Level in Chao Phraya River by Machine Learning Techniques

นางสาวสิริเพ็ญ จึงเจริญรัตน์

รหัสประจำตัว 54660543

ขอรับรองว่ารายงานฉบับนี้ข้าพเจ้าไม่ได้คัดลอกมาจากที่ใด

รายงานฉบับนี้ได้รับการตรวจสอบและอนุมัติให้เป็นส่วนหนึ่งของการศึกษาวิชาการศึกษาอิสระ

หลักสูตรวิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศ)

ภาคเรียนที่ 2 ปีการศึกษา 2558

.....อาจารย์ที่ปรึกษา

(ผศ.ดร.กิตติ์สุชาต พสุภา)

.....กรรมการสอบ

(ผศ.ดร.ธีรพงศ์ ลีลานุภาพ)

.....กรรมการสอบ

(ดร.สิงหะ ฉวีสุข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาโดยใช้การเรียนรู้ของเครื่องจักร
นักศึกษา	นางสาวสิริเพ็ญ จึงเจริญรัตน์
รหัสนักศึกษา	54660543
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	เทคโนโลยีระบบสารสนเทศ
ปีการศึกษา	2558
อาจารย์ที่ปรึกษา	ผศ.ดร.กิตติ์สุชาต พสุภา

บทคัดย่อ

ในปัจจุบันการพยากรณ์ระดับน้ำในแม่น้ำเจ้าพระยา พยากรณ์โดย กรมอุทกศาสตร์ กองทัพเรือ โดยใช้วิธี Harmonic Tidal Analysis ซึ่งใช้วิธีนี้ให้ผลการทำนายที่มีความคลาดเคลื่อนเป็นอย่างมาก แต่ยังแสดงถึงแนวโน้มของระดับน้ำเบื้องต้นได้ ในงานวิจัยนี้จึงใช้อัลกอริทึมการเรียนรู้ของเครื่องจักรมาใช้ในการพยากรณ์ระดับน้ำ เพื่อปรับปรุงให้ประสิทธิภาพในการทำนายใกล้เคียงกับความเป็นจริง อัลกอริทึมที่นำมาทดสอบได้แก่ การถดถอยเชิงเส้น, Support Vector Regression, k-Nearest Neighbors และ Random Forest โดยข้อมูลนำเข้า คือ ข้อมูลระดับน้ำของสถานีวัดระดับน้ำต่าง ๆ โดยเป็นอนุกรมเวลาย้อนหลัง 24 48 หรือ 72 ชั่วโมง เพื่อพยากรณ์ล่วงหน้า 24 ชั่วโมง โดยพยากรณ์ที่สถานีน้ำในแม่น้ำเจ้าพระยา 2 สถานี คือ หน้ากองบัญชากองทัพเรือ ป้อมพระจุลจอมเกล้า จากการทดลองนี้ ใช้ข้อมูลอนุกรมเวลาย้อนหลัง 72 และ อัลกอริทึม Support Vector Regression รวมกับ Radial Basis Function Kernel สามารถพยากรณ์ระดับน้ำแล้วมีความคลาดเคลื่อนน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Estimation of Water Level in Chao Phraya River by Machine Learning Techniques
Student	Miss Siripen Jungjareantrat
Student ID	54660543
Degree	Master of Science
Program	Information Technology
Major	Information System Technology
Academic Year	2015
Advisor	Assistant Professor Dr. Kitsuchart Pasupa

ABSTRACT

Presently, the model that is employed by the Hydrographic Department of the Royal Thai Navy for predicting the level of water in Chao Phraya river uses Harmonic Tidal Analysis in its calculation. The model can predict the overall trend well but with high individual prediction errors. This research utilized several machine learning algorithm namely, Linear Regression, Support Vector Regression, k-Nearest Neighbors and Random Forest, to generate the prediction model that could achieve higher accuracy of level water prediction. Input specification was defined as previous 24, 48 or 72 hour data to predict 24 hour later at the Royal Thai Navy Headquarters station and the Chulachomklao Fort station. The experiment showed that using 72 hour lagged data with support vector machine with Radial basis function led to the best prediction

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การศึกษาอิสระฉบับนี้สำเร็จได้ด้วยความกรุณาของ ผศ. ดร. กิติ์สุชาต พสุภา ซึ่งเป็นอาจารย์ที่ปรึกษา สำหรับความช่วยเหลือ ให้คำชี้แนะ ช่วยแก้ปัญหาตลอดจนให้ความรู้และประสบการณ์ที่ดีแก่ข้าพเจ้า ขอขอบคุณ ดร. สิงหะ ฉวีสุข สำหรับคำแนะนำ และ ความช่วยเหลือด้านต่างๆ ขอขอบคุณ พี่ ๆ เพื่อน ๆ น้อง ๆ ใน Data Science and Machine Learning Laboratory สำหรับความช่วยเหลือ และกำลังใจที่ดีต่อข้าพเจ้าตลอดมา ขอขอบคุณการสนับสนุนการทำการศึกษาอิสระ จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ขอขอบคุณ กรมอุทกศาสตร์ กองทัพเรือ และ สำนักการระบายน้ำ กรุงเทพมหานคร สำหรับข้อมูลระดับน้ำในแม่น้ำเจ้าพระยาที่ใช้ในการทำวิจัย สุดท้ายนี้ขอขอบคุณครอบครัวและเพื่อน ๆ ของข้าพเจ้าสำหรับการสนับสนุน การส่งเสริมในทุก ๆ ด้านในงานวิจัย รวมถึงกำลังใจที่ได้ตลอดเวลาที่ทำงานวิจัย จนสำเร็จลุล่วงเป็นอย่างดี คุณงามความดีอันใดที่เกิดจากการศึกษาอิสระฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูบาอาจารย์ที่เคารพทุกท่านที่ได้มอบวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ข้าพเจ้า

สิริเพ็ญ จึงเจริญรัตน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อ I

ABSTRACT II

กิตติกรรมประกาศ III

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 ขอบเขตของการศึกษา 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 3

1.5 ขั้นตอนของการศึกษา 3

1.6 เครื่องมือที่ใช้ในการพัฒนา 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 การวิเคราะห์น้ำขึ้น-น้ำลง 4

2.2 การเรียนรู้ของเครื่องจักร 4

2.3 กระบวนการแบ่งเซตของข้อมูล 10

2.4 งานวิจัยที่เกี่ยวข้อง 12

บทที่ 3 วิธีดำเนินการวิจัย 14

3.1 การเตรียมข้อมูล 14

3.2 วิธีการทดลอง 17

3.3 การทดสอบความแม่นยำของผลพยากรณ์ที่ได้ 18

บทที่ 4 ผลการทดลอง 19

4.1 ผลและวิเคราะห์การเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำท่า 19

4.2 ผลและการวิเคราะห์ข้อมูลย้อนหลังที่มีผลต่อการพยากรณ์ 21

4.3 ผลและการวิเคราะห์ปัจจัยที่มีผลต่อการพยากรณ์ 24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

4.4 ผลพยากรณ์ระดับน้ำ.....27
บทที่ 5 สรุปผลและข้อเสนอแนะ.....29
บรรณานุกรม.....30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูปที่ หน้า

2.1 แสดงความคิดพื้นฐานของ SVM ที่ใช้แก้ปัญหาการแบ่งกลุ่ม 7
2.2 แสดงการพยากรณ์ค่าโดยซัพพอร์ตเวกเตอร์แบบถดถอย 7
2.3 แสดง Random Forest 10
2.4 แสดงจำนวนตัวอย่างข้อมูลชุดสอนและข้อมูลทดสอบ 10
2.5 แสดง 5-fold cross-validation 11
3.1 แสดงตำแหน่งของสถานีวัดระดับน้ำที่ใช้ในการวิเคราะห์ในแม่น้ำเจ้าพระยา 15
4.1 แสดงค่าสัมประสิทธิ์ของข้อมูลนำเข้าย้อนหลัง 72 ชั่วโมงสถานีกองบัญชาการกองทัพเรือ 22
4.2 แสดงค่าสัมประสิทธิ์ของข้อมูลนำเข้าย้อนหลัง 72 ชั่วโมงสถานีป้อมพระจุลจอมเกล้า 23
4.3 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัยของสถานีกองบัญชาการกองทัพเรือ 25
4.4 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัยของสถานีป้อมพระจุลจอมเกล้า 26
4.5 แสดงกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีกองบัญชาการกองทัพเรือ 27
4.6 แสดงกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีป้อมพระจุลจอมเกล้า 28

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 สถานีวัดระดับน้ำที่ใช้ในการวิเคราะห์ในแม่น้ำเจ้าพระยา 16
3.2 การปรับค่าพารามิเตอร์ของแต่ละอัลกอริทึม 17
4.1 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือ..... 20
4.2 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำสถานีป้อมพระจุลจอมเกล้า 21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

แม่น้ำเจ้าพระยาเป็นแม่น้ำสายสำคัญสายหนึ่งของประเทศไทย มีลำน้ำสาขามากกว่า 200 สาขา และมีพื้นที่รับน้ำทั้งหมดประมาณ 20,125 ตารางกิโลเมตร (กรมชลประทาน, 2554) แม่น้ำเจ้าพระยาเกิดจากการรวมตัวของแม่น้ำสายหลัก 2 สายจากภาคเหนือ คือ แม่น้ำปิงและแม่น้ำน่าน ไหลมาบรรจบที่จังหวัดนครสวรรค์ จากนั้นไหลผ่านที่ราบภาคกลางในเขต 11 จังหวัด ได้แก่ นครสวรรค์, ชัยนาท, สิงห์บุรี, อ่างทอง, พระนครศรีอยุธยา, ปทุมธานี, นนทบุรี, และกรุงเทพมหานคร ก่อนออกสู่อ่าวไทยที่จังหวัดสมุทรปราการ รวมเป็นระยะทางประมาณ 372 กิโลเมตร สำหรับคนไทย แม่น้ำเจ้าพระยาคือวงจรชีวิต และหัวใจของวัฒนธรรมไทย เหตุการณ์สำคัญมากมายในประวัติศาสตร์ไทย เกิดขึ้นบนสองฝั่งฟากแม่น้ำนี้ นอกจากด้านประวัติศาสตร์และวัฒนธรรมแล้ว แม่น้ำเจ้าพระยายังเป็นเส้นทางคมนาคมและขนส่งที่สำคัญของประเทศ ตลอดจนเส้นทางที่ไหลผ่าน กระแสน้ำได้พัดพาดินตะกอนอันอุดมสมบูรณ์มาด้วย จึงก่อให้เกิดที่ราบลุ่มอันกว้างใหญ่และเหมาะกับการทำเกษตร ซึ่งมีพื้นที่ประมาณ 14,786 ตารางกิโลเมตร (สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร, 2556)

น้ำถือเป็นปัจจัยสำคัญในการทำเกษตร เพราะว่าพื้นที่การเกษตรมีความต้องการใช้น้ำสูง แต่ว่าปัจจุบันมีพื้นที่ชลประทานที่สนับสนุนภาคการเกษตรเพียงร้อยละ 17.8 เท่านั้น ส่วนพื้นที่การเกษตรที่เหลืออีกร้อยละ 82.2 จะต้องพึ่งน้ำฝนตามธรรมชาติ ซึ่งในแต่ละปีมีปริมาณน้ำฝนประมาณ 760,000 ล้านลูกบาศก์เมตร โดยประมาณร้อยละ 71.7 จะระเหยกลับไปในอากาศและไหลซึมลงใต้ดิน ส่วนร้อยละ 28.3 หรือประมาณ 215,000 ล้านลูกบาศก์เมตร (วิชัย สุภาโสด, 2549) จะกลายเป็นน้ำท่า เนื่องจากความแตกต่างของสภาพภูมิประเทศ ปริมาณน้ำฝน ปริมาณน้ำท่า สภาพดิน การใช้ที่ดิน การเกษตรกรรม ทำให้เกิดปัญหาเรื่องน้ำที่แตกต่างกัน ไม่ว่าจะเป็นปัญหาเรื่องน้ำหลาก น้ำท่วม หรือ อุทกภัย เพราะมีน้ำมากในฤดูฝน และเกิดปัญหาการขาดแคลนน้ำ หรือ น้ำแล้ง เพราะมีน้ำน้อยในฤดูแล้ง ส่งผลให้รายได้ของเกษตรกรมีความไม่แน่นอนตามไปด้วย ดังนั้นจึงมีความจำเป็นอย่างยิ่งที่ต้องมีระบบการบริหารจัดการน้ำที่มีประสิทธิภาพ การบริหารจัดการน้ำได้แก่ การสร้างเขื่อน อ่างเก็บน้ำ การชลประทาน และการสร้างแบบจำลองในการวิเคราะห์เกี่ยวกับน้ำเพื่อช่วยในการตัดสินใจบริหารการจัดการน้ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเทศไทยมีหน่วยงานที่พยากรณ์ระดับน้ำ คือ กรมอุทกศาสตร์ กองทัพเรือ จากการสอบถามผู้เชี่ยวชาญทางด้านการพยากรณ์ระดับน้ำ ที่กรมอุทกศาสตร์ กองทัพเรือ ใช้วิธีการวิเคราะห์น้ำขึ้น-น้ำลง (Harmonic Tidal Analysis: HTA) (เชิดวงศ์ แสงศุภวานิช, 2555) การพยากรณ์ระดับน้ำในแม่น้ำเจ้าพระยาทำการพยากรณ์ล่วงหน้า 1 ปีเพื่อใช้ในการเดินเรือเข้าแม่น้ำเจ้าพระยา การพยากรณ์ระดับน้ำด้วยวิธีดังกล่าวมีความคลาดเคลื่อนจากระดับน้ำที่วัดจริงเป็นอย่างมาก แต่ยังมีประโยชน์ในการแสดงแนวโน้มของระดับน้ำเบื้องต้นได้

จากอดีตถึงปัจจุบันมีงานวิจัยจำนวนมากที่นำการเรียนรู้ของเครื่องจักร (Machine Learning) มาใช้ในการพยากรณ์ระดับน้ำ การเรียนรู้ของเครื่องจักรสามารถพยากรณ์ระดับน้ำใกล้เคียงกับข้อมูลจริง (ทวีศักดิ์ วังไพศาล และ จักรฤทธิ์ ตรีนาจ, 2555) (นิธินันท์ มาตา, แสงดาว นพพิทักษ์ และ ศิโรรัตน์ กุลวงศ์, 2558) (S. Phuphong and C. Surussavadee, 2013) (P. Markpeng, P. Wongnimmarn, N. Champreeda, P. Vateekul and K. Sarinnapakorn, 2014) ดังนั้นในงานวิจัยนี้ได้นำการเรียนรู้ของเครื่องจักร มาใช้ในการสร้างแบบจำลองของการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา เพื่อให้ได้ข้อมูลที่ใกล้เคียงความเป็นจริงและน่าเชื่อถือมากขึ้น ซึ่งในงานวิจัยนี้ได้เลือกใช้อัลกอริทึมการเรียนรู้ของเครื่องจักร ที่มีการใช้งานอย่างแพร่หลายในการพยากรณ์ มาสร้างแบบจำลองของการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา ได้แก่ การถดถอยเชิงเส้น (Linear Regression), ซัพพอร์ตเวกเตอร์แบบถดถอย (Support Vector Regression: SVR), k-Nearest Neighbors (k-NN) และ Random Forest (RF) เลือก 4 อัลกอริทึมนี้

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. ทดสอบอัลกอริทึมการเรียนรู้ของเครื่องจักร เช่น การถดถอยเชิงเส้น, SVR, k-NN และ RF เพื่อเปรียบเทียบอัลกอริทึมที่สามารถพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาใกล้เคียงกับข้อมูลจริง
2. หาปัจจัยที่ส่งผลต่อการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา

1.3 ขอบเขตของการศึกษา

1. ศึกษาอัลกอริทึมการเรียนรู้ของเครื่องจักร 4 อัลกอริทึม ได้แก่ การถดถอยเชิงเส้น, SVR, k-NN และ RF
2. ศึกษาปัจจัยที่ส่งผลต่อการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา โดยได้ศึกษาข้อมูลระดับน้ำท่าในแม่น้ำเจ้าพระยา ในจังหวัดกรุงเทพมหานคร ในช่วงเวลาปี 2553 ถึงปี 2554 กรมอุทกศาสตร์ กองทัพเรือ และ สำนักการระบายน้ำ กรุงเทพมหานคร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถสร้างแบบจำลองและพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาได้มีประสิทธิภาพดีขึ้นจากการพยากรณ์ระดับน้ำโดยใช้วิธี HTA
2. สามารถพบปัจจัยที่ส่งผลต่อการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา

1.5 ขั้นตอนของการศึกษา

1. รวบรวมข้อมูลที่นำมาใช้ในการพยากรณ์
2. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
3. กำหนดขอบเขตวิเคราะห์และทำการสร้างแบบจำลองที่ใช้พยากรณ์ระดับน้ำท่าโดยใช้อัลกอริทึมต่าง ๆ
4. ทดสอบและวิเคราะห์ผล
5. สรุปผลและจัดทำเอกสาร

1.6 เครื่องมือที่ใช้ในการพัฒนา

- Matlab

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎี งานวิจัยที่เกี่ยวข้อง และการเรียนรู้เครื่องจักรที่ใช้ในวิชาการศึกษาอิสระ 2 โดยมีรายละเอียดดังต่อไปนี้

2.1 การวิเคราะห์น้ำขึ้น-น้ำลง

การวิเคราะห์น้ำขึ้น-น้ำลง (Harmonic Tidal Analysis : HTA) คือ การวิเคราะห์ข้อมูลน้ำขึ้น-น้ำลงย้อนหลังและนำมาแยกองค์ประกอบของน้ำขึ้น-น้ำลง (Tidal Constituents) ซึ่งน้ำขึ้น-น้ำลง ณ ที่ใดที่หนึ่งจะประกอบไปด้วยองค์ประกอบหลายองค์ประกอบ จากนั้นจะได้แบบจำลองของระดับน้ำซึ่งนำมาใช้พยากรณ์ในอนาคตได้ (เชิดวงศ์ แสงศุภวานิช, 2555)

2.2 การเรียนรู้ของเครื่องจักร

การเรียนรู้ของเครื่องจักร (Machine Learning) คือ การทำให้เครื่องจักรเรียนรู้ได้จากข้อมูลตัวอย่าง สามารถแบ่งประเภทของการเรียนรู้ของเครื่องจักรออกเป็น 3 ประเภทหลัก ๆ ได้แก่

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ นำข้อมูลที่มีป้ายกำกับ (Label) มาใช้สอนให้กับการเรียนรู้ของเครื่องจักร การเรียนรู้ประเภทนี้ เช่น
   - การจำแนกข้อมูล (Classification)
   - การพยากรณ์ค่า (Regression)
2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คือ การหาโครงสร้างความสัมพันธ์ของโดยไม่มีป้ายกำกับของข้อมูล การเรียนรู้ประเภทนี้ เช่น
   - การหากฎความสัมพันธ์ (Association Rule)
   - การแบ่งกลุ่มข้อมูล (Clustering)
3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) คือ เมื่อคอมพิวเตอร์ตอบถูกจะมีการให้รางวัล (Reward) ถ้าคอมพิวเตอร์ตอบผิดจะมีการลงโทษ (Penalty)

การเรียนรู้ของเครื่องจักรสามารถแก้ปัญหาได้มากมาย เช่น การคำนวณเชิงอารมณ์ (Affective Computing), ชีวสารสนเทศศาสตร์ (Bioinformatics), เคมีสารสนเทศศาสตร์ (Chemoinformatics), คอมพิวเตอร์วิทัศน์ (Computer Vision), การค้นคืนสารสนเทศ (Information Retrieval), การวินิจฉัยทางการแพทย์ (Medical Diagnosis), การประมวลผลภาษาธรรมชาติ (Natural Language Processing), การหาค่าเหมาะสมที่สุด (Optimization), การรู้จำเสียงพูด (Speech Recognition) เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต้น ตัวอย่างของอัลกอริทึมทางการเรียนรู้ของเครื่องจักร เช่น การถดถอยเชิงเส้น, โครงข่ายประสาทเทียม (Neural Network), ต้นไม้ตัดสินใจ (Decision Tree), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM), k-NN, RF และ อื่น ๆ

ในงานวิจัยนี้ใช้การเรียนรู้แบบมีผู้สอนทำการพยากรณ์ค่า โดยใช้ 4 อัลกอริทึมดังต่อไปนี้

2.2.1 การถดถอยเชิงเส้น

การถดถอยเชิงเส้น (Linear regression) คือ ความสัมพันธ์ระหว่างตัวแปรสองตัวแปร คือ ตัวแปรอิสระ (x) และ ตัวแปรตาม (y) ที่มีความสัมพันธ์ในลักษณะเชิงเส้น (Linear)

สมมติว่ามีข้อมูลชุดสอน (Training Data) $\{(\bar{x}_1, y_1),(\bar{x}_2, y_2),...,(\bar{x}_n, y_n)\}$ โดยที่ $\bar{x}$ คือ ตัวแปรอิสระ (Independent Variable) มีขนาดเวกเตอร์เท่ากับ m , $\bar{x}_i = \left[x_1^{(i)}, x_2^{(i)}, \ldots, x_m^{(i)}\right]$, โดยมี $\bar{x}$ จำนวน n ตัวอย่าง และ y คือ ตัวแปรตาม (Dependent Variable)

สมการการถดถอยเชิงเส้น สามารถแสดงดังสมการที่ (2.1)

$$y_i = x_i w_1 + x_i w_2 + \ldots + x_i w_m + b \quad ; \quad i = 1,2,...,n \tag{2.1}$$

โดยที่ w คือ ค่าสัมประสิทธิ์, b คือ ไบอัส (Bias) ซึ่ง $\bar{w}$ สามารถหาได้โดยใช้ความลาดชันของเกรเดียน (Gradient Descent) อย่างไรก็ตาม $\bar{w}$ ยังสามารถหาได้โดยใช้ Least Square มีสมการดังนี้

$$\bar{w} = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \bar{y} \tag{2.2}$$

โดยที่ $\bar{y} = \left[y_1, y_2, \ldots, y_n\right]^T$ และ $\tilde{X} = \left[\bar{1}\ X\right]$ ซึ่ง $X \in \Re^{n \times m}$

มีการนำ Regularization มาแก้ปัญหาการเรียนรู้ในการฝึกสอน เมื่อแบบจำลองนั้นพอดีกับข้อมูลเกินไป (Over-fitting) มีสมการดังนี้

$$\bar{w} = \left(\tilde{X}^T \tilde{X} + \lambda \tilde{I}\right)^{-1} \tilde{X}^T Y \tag{2.3}$$

โดยที่ λ = ตัวแปรของ Regularization และ

$$\tilde{I} = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$$

เราเรียกสมการนี้ว่า Regularized Least Square (RLS)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 Kernel Function

Kernel Function คือ การเปลี่ยนแปลงข้อมูลที่มีมิติ (Dimension) ต่ำกว่าให้มีมิติสูงขึ้น ซึ่งสามารถเลือกใช้ได้ตามความเหมาะสมกับประเภทของข้อมูลที่ใช้งาน ถ้าเลือกฟังก์ชันแบบเชิงเส้นซึ่งจะเหมาะกับข้อมูลที่มีความสัมพันธ์เชิงเส้น ได้แก่ Linear Kernel และฟังก์ชันแบบไม่เป็นเชิงเส้นซึ่งจะเหมาะกับข้อมูลที่มีความสัมพันธ์ไม่เป็นเชิงเส้น เช่น Radial Basis Function Kernel, Polynomial Kernel และ Sigmoid Kernel

Kernel Function มีหลายรูปแบบ เช่น

1. Linear Kernel

$$k\left(x_i, x_j\right) = x_i^T x_j \tag{2.4}$$

2. Radial Basis Function Kernel (RBF)

$$k(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{2.5}$$

3. Polynomial Kernel

$$k(x_i, x_j) = (1 + x_i \cdot x_j)^d \tag{2.6}$$

4. Sigmoid Kernel

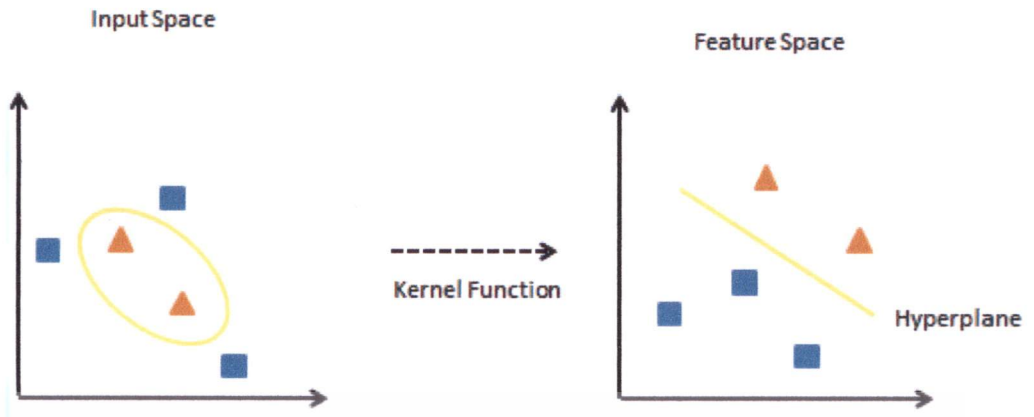
$$k(x_i, x_j) = \tanh\left(s\left(x_i \cdot x_j\right) + c\right) \tag{2.7}$$

ในงานวิจัยนี้ได้เลือกใช้ Linear Kernel ที่เป็นฟังก์ชันแบบเชิงเส้น และ RBF ที่เป็นฟังก์ชันแบบไม่เป็นเชิงเส้น

2.2.3 ซัพพอร์ตเวกเตอร์แบบถดถอย

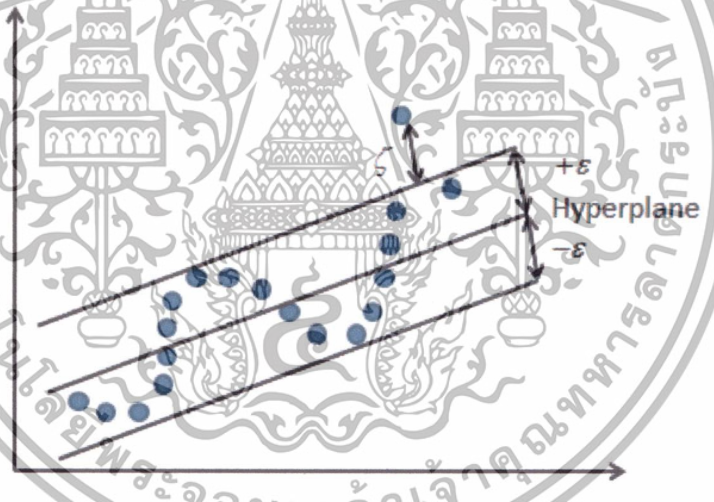
ความคิดพื้นฐานของ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) คือ การแปลงข้อมูลชุดสอน ให้มีมิติที่สูงขึ้นโดยใช้ Kernel Function ดังรูปที่ 2.1 และหลังจากนั้นทำการสร้างระนาบ (Hyperplane) ที่เหมาะสมที่สุด โดยแบ่งกลุ่มโดยพยายามที่จะใช้ระยะห่างระหว่างกลุ่มมากที่สุด (Maximum Margin) เพื่อให้เกิดข้อผิดพลาดน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 แสดงความคิดพื้นฐานของ SVM ที่ใช้แก้ปัญหาการแบ่งกลุ่ม

ซัพพอร์ตเวกเตอร์แบบถดถอย (Support Vector Regression: SVR) ได้นำความคิดพื้นฐานของ SVM มาประยุกต์ใช้คือ จะทำการสร้างระนาบที่เหมาะสมที่สุด และทำการพยากรณ์ค่าให้อยู่ในระนาบที่เหมาะสม โดยที่ระนาบมีค่าตั้งแต่ $+\varepsilon$ ถึง $-\varepsilon$ และมีการอนุญาตให้พยากรณ์ค่าเกินกว่าระนาบที่กำหนดไว้ เรียกว่า Soft Margin โดยที่ Soft Margin มีค่า ζ ดังรูปที่ 2.2



รูปที่ 2.2 แสดงการพยากรณ์ค่าโดยซัพพอร์ตเวกเตอร์แบบถดถอย

2.2.4 k-Nearest Neighbors (k-NN)

k-NN เป็นอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูล คือ การวัดระยะห่างระหว่างระเบียน (Record) ของข้อมูลชุดสอนกับระเบียนของข้อมูลชุดทดสอบ ถ้าระเบียนของข้อมูลทดสอบอยู่ใกล้กับข้อมูลชุดสอนนั้นมากที่สุด จะจัดอยู่ในกลุ่มเดียวกัน สามารถนำมาประยุกต์ใช้การพยากรณ์ค่าได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการทำงานของ k-NN มีดังนี้

อัลกอริทึม 1 k-NN

---

1. กำหนดค่าเริ่มต้น $(X,Y,\bar{x}_{test},k)$: $\bar{x}_{test}$ คือ ข้อมูลทดสอบ (Test Data) มีขนาดเวกเตอร์เท่ากับ m, $\left[x_{test(1)},x_{test(2)},\ldots,x_{test(m)}\right]$, k = จำนวนระเบียนของข้อมูลชุดสอนที่อยู่ใกล้กับข้อมูลทดสอบมากที่สุด
2. ฟังก์ชัน k-NN $(X,Y,\bar{x}_{test},k)$
3. for $i=1$ to n do
4. คำนวณระยะห่างระหว่าง $\bar{x}$ กับ $\bar{x}_{test}$
5. end for
6. จัดเรียงลำดับของระยะห่าง ตามจำนวน k ที่กำหนดไว้
7. การคำนวณหาผลลัพธ์
8. จบฟังก์ชัน

หมายเหตุ การคำนวณหาผลลัพธ์มีหลายวิธี ยกตัวอย่าง เช่น นำผลลัพธ์ของข้อมูลชุดสอน ตามจำนวน k ที่กำหนดไว้ มาหาค่าเฉลี่ย หรือ นำผลลัพธ์ของข้อมูลชุดสอน ตามจำนวน k ที่กำหนดไว้ มาหาค่าเฉลี่ยถ่วงน้ำหนัก (Inverse Distance Weighted) ซึ่งเพื่อนบ้านที่ใกล้ที่สุดจะมีค่าเฉลี่ยถ่วงน้ำหนักมาก

การหาระยะห่างระหว่างข้อมูล (Distance Function) มีโดยการหลายวิธีเช่น

- Euclidean Distance

$$d(\bar{x},\bar{x}_{test})=\sqrt{\sum_{i=1}^{m}\left(\bar{x}_i-\bar{x}_{test(i)}\right)^2} \tag{2.8}$$

- Minkowski Distance

$$d(\bar{x},\bar{x}_{test})=\left(\sum_{i=1}^{m}\left(\bar{x}_i-\bar{x}_{test(i)}\right)^p\right)^{\frac{1}{p}} \tag{2.9}$$

ค่า p คือ $p-norm$

- City Block Distance หรือ Manhattan

$$d(\bar{x},\bar{x}_{test})=\sum_{i=1}^{m}\left|\bar{x}_i-\bar{x}_{test(i)}\right| \tag{2.10}$$

โดยที่ $\bar{x}_i$ คือ แอตทริบิวต์ที่ i ของข้อมูลชุดสอน และ $\bar{x}_{test(i)}$ คือ แอตทริบิวต์ที่ i ของข้อมูลชุดทดสอบ โดยข้อมูล $\bar{x}$ และ $\bar{x}_{test}$ มีจำนวนแอตทริบิวต์เท่ากับ m การหาระยะห่างระหว่างข้อมูลที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นิยมใช้มากที่สุดคือ Euclidean Distance และงานวิจัยนี้ใช้การหาระยะทางแบบ Euclidean Distance ด้วย การกำหนดค่า k ควรกำหนดค่า k ให้เหมาะสมเพื่อได้ค่าเป้าหมายที่ใกล้เคียงมากที่สุด

2.2.5 Random Forest

Random Forest (RF) นำมาใช้การจำแนกข้อมูลและการพยากรณ์ค่าข้อมูล โดยสร้างต้นไม้ตัดสินใจจำนวนมากในเวลาที่ทำการฝึกสอนข้อมูล เมื่อทำการฝึกสอนข้อมูลแล้วก็หลายจะได้แบบจำลองที่เป็นเฉพาะของต้นไม้ตัดสินใจต้นนั้น ๆ

ต้นไม้ตัดสินใจต้นเดียวมักมีความแปรปรวนสูง (High Variance) หรือ ค่าเอนเอียงสูง (High Bias) แต่ว่า RF จะแก้ปัญหาความแปรปรวนสูงหรือค่าเอนเอียงสูงได้ เนื่องจาก RF เป็นการรวมกันของต้นไม้ตัดสินใจหลาย ๆ ต้น ทำให้ผลลัพธ์ที่ได้มีประสิทธิภาพมากกว่าการใช้ต้นไม้ตัดสินใจเพียงต้นเดียวและมีโอกาสน้อยที่เกิดปัญหา Over-fitting

ขั้นตอนการทำงานของ RF มีดังนี้

อัลกอริทึม 2 RF

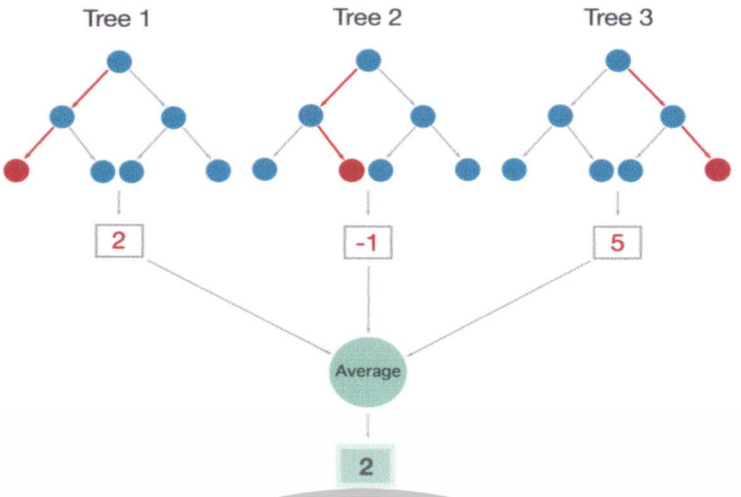
---

1. กำหนดค่าเริ่มต้น $(X, Y, \bar{x}_{test}, nTree)$: $nTree$ = จำนวนต้นไม้ตัดสินใจ
2. ฟังก์ชัน k-NN $(X, Y, \bar{x}_{test}, nTree)$
3. for $i = 1$ to $nTree$ do
4. สร้างแบบจำลองของต้นไม้ตัดสินใจ 1 ต้น โดยใช้ข้อมูล X
5. end for
6. นำ $\bar{x}_{test}$ หาผลลัพธ์จากแบบจำลองของต้นไม้ตัดสินใจแต่ละต้น
7. นำผลลัพธ์ที่ได้ของต้นไม้แต่ละต้น มาคำนวณหาผลลัพธ์
8. จบฟังก์ชัน

หมายเหตุ การคำนวณหาผลลัพธ์ของข้อมูลทดสอบมีหลายวิธี ยกตัวอย่าง เช่น นำผลลัพธ์ที่ได้จากแบบจำลองมาหาค่าเฉลี่ย หรือ นำผลลัพธ์ที่ได้จากแบบจำลองมาทำการลงคะแนน และดูจากเสียงข้างมากเป็นหลัก

แม้ว่าใช้ต้นไม้ตัดสินใจเหมือนกันแต่ข้อมูลที่ใช้การสร้างแบบจำลองต่างกันก็ทำให้แบบจำลองที่สร้างขึ้นมาได้มีลักษณะต่างกัน ยกตัวอย่าง เช่น กำหนดให้มีต้นไม้ตัดสินใจ 3 ต้น แต่ละต้นทำการสร้างแบบจำลอง ซึ่งแบบจำลองไม่เหมือนกันทั้ง 3 ต้น เมื่อได้แบบจำลองแล้วนำแบบจำลองของต้นไม้ตัดสินใจแต่ละต้นมาหาผลลัพธ์ ต้นไม้ตัดสินใจต้นที่ 1 ผลลัพธ์เท่ากับ 2 ต้นไม้ตัดสินใจต้นที่ 2 ผลลัพธ์เท่ากับ -1 ต้นไม้ตัดสินใจต้นที่ 3 ผลลัพธ์เท่ากับ 5 แล้วผลลัพธ์ของต้นไม้ตัดสินใจ 3 ต้น มาหาค่าเฉลี่ย ผลลัพธ์เท่ากับ 2 ดังรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



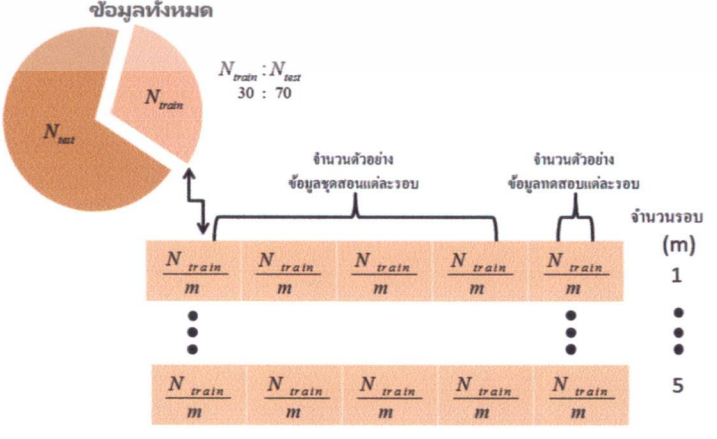
รูปที่ 2.3 แสดง Random Forest

2.3 กระบวนการแบ่งเซตของข้อมูล

กระบวนการแบ่งเซตของข้อมูล คือ การแบ่งข้อมูลออกเป็นส่วน ๆ เพื่อจุดประสงค์ต่าง ๆ เช่น การแบ่งข้อมูลเป็นข้อมูลชุดสอน และ ข้อมูลชุดทดสอบ

2.3.1 Split Test รวมกับ Cross-validation

การใช้วิธีแบ่งเซตข้อมูลแบบ Split Test รวมกับ Cross-validation โดยจะทำการแบ่งเซตข้อมูลแบบ Split Test ก่อน แล้วในการทำ Split Test แต่ละครั้ง จะทำการสุ่มเลือกข้อมูลออกเป็น ข้อมูลชุดสอน (Training Data) และ ข้อมูลทดสอบ (Test Data) โดยหลังจากนั้นจะนำข้อมูลชุดสอนมาทำ Cross-validation เพื่อทำการหาค่าพารามิเตอร์ที่ดีที่สุด ดังรูปที่ 2.5 ต่อมานำพารามิเตอร์ที่ดีที่สุดมาสร้างแบบจำลองโดยใช้ข้อมูลชุดสอน หลังจากนั้นทดสอบแบบจำลองด้วยข้อมูลทดสอบ โดยข้อดีของการทำ Split Test รวมกับ Cross-validation คือ แก้ปัญหา Over-fitting และสามารถลดเวลาในการหาแบบจำลองที่เหมาะสม



รูปที่ 2.4 แสดงจำนวนตัวอย่างข้อมูลชุดสอนและข้อมูลทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 Split Test

Split Test คือ การแบ่งข้อมูลทั้งหมดออกเป็น 2 ส่วน โดยการสุ่มเลือก ซึ่งได้แก่ ข้อมูลชุดสอน กับข้อมูลทดสอบ โดยไม่มีอัตราส่วนที่แน่นอนสำหรับการแบ่งข้อมูล แต่วิธีได้รับความนิยมแพร่หลาย คือ การแบ่งข้อมูลชุดสอนมากกว่าข้อมูลทดสอบ เพราะอาจจะทำให้การพยากรณ์หรือ การจัดกลุ่ม มีความใกล้เคียงกับข้อมูลจริง ซึ่งอัตราส่วนที่นิยมมีหลายแบบ เช่น 50:50, 60:40, 70:30, 80:20, 90:10 โดยข้อดีของการทำด้วยวิธี Split Test หากมีข้อมูลปริมาณน้อย จะสามารถทำการประมวลผลได้รวดเร็ว แต่อย่างไรก็ตามการทำ Split Test สามารถทำการสุ่มได้เพียงครั้งเดียว

2.3.3 Cross-validation

Cross-validation เป็นเทคนิคการตรวจสอบการประเมินผลของการหาค่าที่เหมาะสมของพารามิเตอร์ โดยพื้นฐานของเทคนิคนี้คือการสุ่มตัวอย่าง โดยเริ่มจากแบ่งชุดข้อมูลออกเป็นส่วน ๆ (Fold) และนำข้อมูลบางส่วนที่แบ่งนำไปใช้เพื่อการฝึกสอน และข้อมูลที่เหลือจะนำไปใช้ในการทดสอบ

ในการทำ Cross-validation มีวิธี m –fold cross-validation โดยจะทำการเลือกสุ่มข้อมูลออกเป็น m ชุดเท่ากัน และทำการหาค่าที่เหมาะสมของพารามิเตอร์ m รอบ โดยแต่ละรอบการคำนวณข้อมูลชุดหนึ่งจากข้อมูล m ชุดจะถูกเลือกออกมาเพื่อเป็นข้อมูลทดสอบและข้อมูลอีก $m-1$ ชุดจะถูกใช้เป็นข้อมูลสำหรับการฝึกสอน ยกตัวอย่างเช่น ในการทดลองครั้งแรกข้อมูลชุดที่ 5 เป็น ข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ในการทดลองครั้งสองข้อมูลชุดที่ 4 เป็น ข้อมูลชุดทดสอบและข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำจนกระทั่งข้อมูลทุกชุดได้ถูกนำมาเป็นข้อมูลชุดทดสอบซึ่งมีการทดลองทั้งหมด m ครั้ง ดังรูปที่ 2.4

ครั้งที่ 1 : ข้อมูลชุดสอน	1	2	3	4	ข้อมูลทดสอบ	5
ครั้งที่ 2 : ข้อมูลชุดสอน	1	2	3	5	ข้อมูลทดสอบ	4
ครั้งที่ 3 : ข้อมูลชุดสอน	1	2	4	5	ข้อมูลทดสอบ	3
ครั้งที่ 4 : ข้อมูลชุดสอน	1	3	4	5	ข้อมูลทดสอบ	2
ครั้งที่ 5 : ข้อมูลชุดสอน	2	3	4	5	ข้อมูลทดสอบ	1

รูปที่ 2.5 แสดง 5-fold cross-validation

ข้อดีของ m -fold cross-validation ใช้เพื่อแก้ปัญหา Over-fitting และ อย่างไรก็ตามการใช้ m -fold cross-validation ต้องใช้เวลานานเนื่องจากนิยมกำหนดให้ค่า m มีค่าเท่ากับ 5 หรือ 10 (M. Galar, A. Fernandez, E. Barrenchea, H. Bustince and F. Herrera, 2016) (K. M. Broms, M. B.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Hooten and R. M. Fitzpatrick, 2016) (นิธินันท์ มาตา, แสงดาว นพพิทักษ์ และ ศิโรรัตน์ กุลวงศ์, 2558)

2.4 งานวิจัยที่เกี่ยวข้อง

ปัจจุบันนักวิจัยจำนวนมากนำการเรียนรู้ของเครื่องจักร มาแก้ไขปัญหาด้านต่าง ๆ เช่น การคำนวณเชิงอารมณ์ ชีวสารสนเทศศาสตร์ เคมีสารสนเทศศาสตร์ ก็มีนักวิจัยจำนวนหนึ่งที่ได้นำการเรียนรู้ของเครื่องจักร มาสร้างแบบจำลองเกี่ยวกับน้ำ

(S. Phuphong and C. Surussavadee, 2013) ได้นำเสนอการพยากรณ์ระดับน้ำที่ลุ่มแม่น้ำคลองอู่ตะเภา จังหวัดสงขลา โดยใช้ข้อมูลระดับน้ำท่าย้อนหลัง 12 ชั่วโมง เป็นข้อมูลนำเข้า และใช้อัลกอริทึมโครงข่ายประสาทเทียมมีการกำหนดชั้นซ่อนทั้งหมด 3 ชั้น และมีโหนดแต่ละชั้นซ่อน 10, 5, 1 ตามลำดับ เพื่อพยากรณ์ระดับน้ำล่วงหน้าในอีก 12 ชั่วโมงถัดไป สำหรับการพยากรณ์ 2 สถานี ได้แก่ สถานีบ้านตะเคียนเภา และสถานีบ้านเมืองคง ใช้ข้อมูลระดับน้ำปี 2551 ในการเรียนรู้ ใช้ข้อมูลระดับน้ำปี 2552 ในการทดสอบ ผลจากการพยากรณ์ เกิดความคลาดเคลื่อนของระดับน้ำคิดเป็นร้อยละ 1.92 และร้อยละ 6.67 ของสถานีบ้านตะเคียนเภา และบ้านเมืองคง ตามลำดับ ถึงแม้ว่าโครงข่ายประสาทเทียมสามารถพยากรณ์ระดับน้ำได้แม่นยำ แต่การพยากรณ์ 12 ชั่วโมงล่วงหน้าเป็นเวลาที่น้อยเกินไปสำหรับการเตือนภัยและอพยพไปอยู่ในที่ปลอดภัย รวมทั้งการใช้ข้อมูลนำเข้าเป็นข้อมูลระดับน้ำท่าย้อนหลัง 24 หรือ 48 หรือ 72 ชั่วโมง น่าจะทำให้การพยากรณ์ระดับน้ำแม่นยำมากขึ้น งานวิจัย (ทวีศักดิ์ วังไพศาล และ จักรฤทธิ์ ตรีนาจ, 2555) เป็นงานวิจัยที่ใช้การเรียนรู้ของเครื่องจักรมาสร้างแบบจำลองของน้ำท่า โดยทำการพยากรณ์ระดับน้ำล่วงหน้า 3 วัน ที่สถานีวัดระดับน้ำ M.7 จังหวัดอุบลราชธานี ใช้ข้อมูลระดับน้ำปัจจุบันเป็นข้อมูลนำเข้า และใช้อัลกอริทึมโครงข่ายประสาทเทียม โดยเปรียบเทียบการเรียนรู้ของโครงข่ายประสาทเทียม 2 แบบ คือ Levenberg–Marquardt (LM) และ Bayesian Regularization (BR) ผลจากการศึกษา วิธี LM ให้ผลพยากรณ์ดีกว่า BR โดยกำหนดโหนดในชั้นซ่อนมีค่าเท่ากับ 0.5 เท่าของจำนวนหน่วยในชั้นข้อมูลนำ นอกจากนี้ยังแสดงให้เห็นว่าสถานีน้ำที่มีผลต่อการพยากรณ์คือสถานีน้ำที่อยู่ต้นน้ำที่อยู่ใกล้ที่สุดกับสถานีที่พยากรณ์ งานวิจัย (นิธินันท์ มาตา, แสงดาว นพพิทักษ์ และ ศิโรรัตน์ กุลวงศ์, 2558) นำเสนอการหาปัจจัยที่ส่งผลต่อการพยากรณ์การควบคุมประตูระบายน้ำของอ่างเก็บน้ำห้วยจรเข้มาก จังหวัดบุรีรัมย์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิค ได้แก่ กลุ่มของ Bayes, กลุ่มของ Function, กลุ่มของ Lazy, กลุ่มของ Rules และกลุ่มของ Tree เทคนิคที่มีความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แม่นยำในการพยากรณ์มากที่สุด คือ RF ที่เป็นเทคนิคในกลุ่มของ Tree และปัจจัยที่มีผลต่อการพยากรณ์มากที่สุด คือ ปริมาณน้ำท่าสะสม ปริมาณน้ำฝนสะสม และการรั่วซึม ตามลำดับ และงานวิจัย (P. Markpeng, P. Wongnimmarn, P. Vateekul and K. Sarinnapakorn, 2014) ทำการแก้ปัญหาค่าสูญหายของระดับน้ำสถาบันสารสนเทศทรัพยากรน้ำและการเกษตร (Hydro and Agro Informatics Institute) โดยใช้วิธี DBSCAN ซึ่งวิธีดังกล่าว มีประสิทธิภาพในการแก้ไขค่าสูญหายและผิดปกติของระดับน้ำได้ดีกว่าวิธีการทางสถิติ (ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน)

จากการศึกษางานวิจัยที่กล่าวมาข้างต้น สรุปได้ดังนี้ งานวิจัยทั้ง 4 งาน ไม่มีการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา บริเวณกรุงเทพมหานคร มีงานวิจัย 2 งานใช้โครงข่ายประสาทเทียมในการพยากรณ์ระดับน้ำ ดังนั้นในงานวิจัยนี้จึงต้องการพยากรณ์ระดับน้ำโดยใช้อัลกอริทึมอื่น ๆ ของการเรียนรู้ของเครื่องจักรแทนอัลกอริทึมโครงข่ายประสามเทียม โดยมีข้อมูลนำเข้าเป็นข้อมูลระดับน้ำย้อนหลัง 24 หรือ 48 หรือ 72 ชั่วโมงทำการพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยาล่วงหน้า 1 วัน เพื่อมีเวลาเพียงพอสำหรับการเตือนภัยและอพยพ และมีการนำปัจจัยที่ส่งผลต่อการพยากรณ์ คือ สถานีน้ำที่อยู่ต้นน้ำที่อยู่ใกล้ที่สุดกับสถานีที่พยากรณ์ ตั้งเป็นสมมติฐาน แล้วนำระดับน้ำท่าในแม่น้ำเจ้าพระยาบริเวณกรุงเทพมหานครมาเป็นปัจจัยที่ใช้ในการพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อสร้างแบบจำลองที่สามารถพยากรณ์ระดับน้ำท่าในแม่น้ำเจ้าพระยา โดยมีวิธีดำเนินการวิจัยดังต่อไปนี้

3.1 การเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลระดับน้ำท่าในแม่น้ำเจ้าพระยา ที่กรุงเทพมหานคร ตั้งแต่ปี 2553 ถึงปี 2554 ได้นำข้อมูลมาจาก

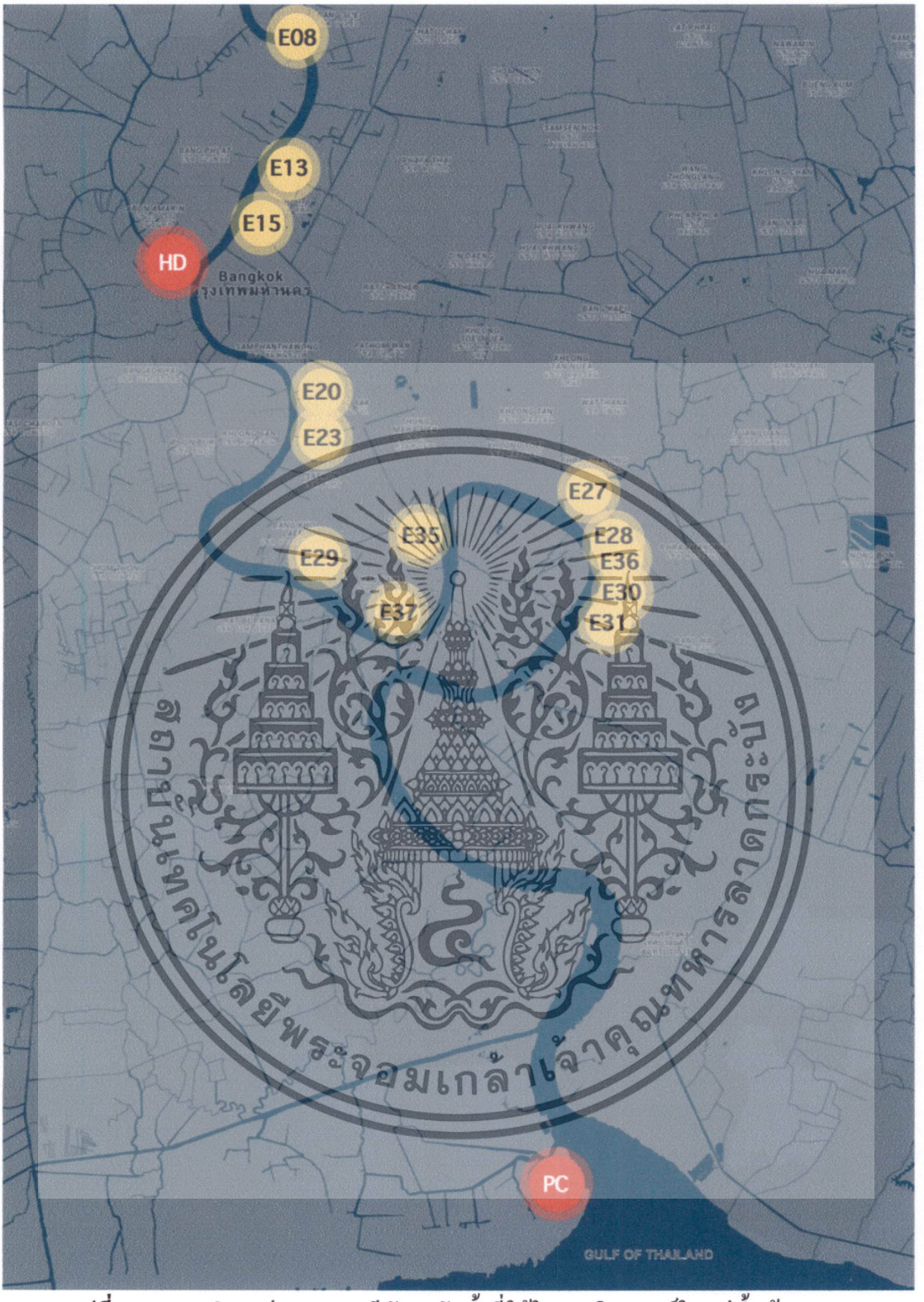
- กรมอุทกศาสตร์ กองทัพเรือ ได้แก่ ระดับน้ำที่กองบัญชาการกองทัพเรือ (HD), ระดับน้ำที่ป้อมพระจุลจอมเกล้า (PC), ระดับน้ำพยากรณ์จาก HTA กองบัญชาการกองทัพเรือ (HD_{HTA}) และ ระดับน้ำพยากรณ์จาก HTA ที่ป้อมพระจุลจอมเกล้า (PC_{HTA})
- สำนักการระบายน้ำ กรุงเทพมหานคร โดยพิจารณาระดับน้ำจาก คลองบางเขนใหม่ ($E08$), คลองบางซื่อ ($E10$), คลองสามเสน ($E13$), คลองเทเวศร์ ($E15$), คลองกรุงเกษม ($E20$), คลองสาทร ($E23$), คลองวัดไทร ($E29$), คลองวัดด่าน ($E37$), คลองช่องนนทรี ($E35$), คลองเตย ($E27$), คลองเจ็ก ($E28$), คลองบางจาก ($E36$), คลองบางอ้อ ($E30$) และคลองบางนา ($E31$)

ในงานวิจัยนี้นำข้อมูลระดับน้ำท่าในแม่น้ำเจ้าพระยามาใช้ทั้งหมด 16 สถานี แต่คลองบางซื่อมีค่า NaN จำนวน 5702 ค่า เพราะเครื่องวัดระดับน้ำที่คลองบางซื่อมีการเสียหาย จึงตัดข้อมูลจากคลองบางซื่อออกในการพิจารณา ทำให้เหลือเพียง 15 สถานีที่เลือกมาใช้ในงานวิจัยได้ดังรูปที่ 3.1 และตารางที่ 3.1 ทำให้มีปัจจัยทั้งหมดมี 17 ปัจจัยได้แก่ $E08$, $E13$, $E15$, HD, HD_{HTA}, $E20$, $E23$, $E29$, $E37$, $E35$, $E27$, $E28$, $E36$, $E30$, $E31$, PC และ PC_{HTA} โดยที่ข้อมูลระดับน้ำท่ามีการบันทึกเป็นรายชั่วโมง ซึ่งจะมีข้อมูลทั้งหมด 17,520 ตัวอย่าง นำมาทำอนุกรมเวลาย้อนหลัง 72 ชั่วโมง จากนั้นค่าข้อมูลใดมีค่า NaN และ 999 (ไม่มีข้อมูลระดับน้ำที่กองบัญชาการกองทัพเรือ เช่น วันหยุดสงกรานต์ วันหยุดปีใหม่ เพราะปิดเครื่องวัดระดับน้ำ) จะทำการลบข้อมูลออกจากข้อมูลตัวอย่างจึงทำให้เหลือข้อมูลทั้งหมด 15,483 ตัวอย่าง จากนั้นจัดการข้อมูลให้เป็นข้อมูลอนุกรมเวลาย้อนหลัง 72 ชั่วโมง

$$x_i = \{x_i(t), x_i(t-1), x_i(t-2), ..., x_i(t-72)\}; i = 1,2,...,17 \quad (3.1)$$

จากนั้นข้อมูลจะถูกนำมาพยากรณ์ล่วงหน้า 24 ชั่วโมง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 แสดงตำแหน่งของสถานีวัดระดับน้ำที่ใช้ในการวิเคราะห์ในแม่น้ำเจ้าพระยา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 สถานีวัดระดับน้ำที่ใช้ในการวิเคราะห์ในแม่น้ำเจ้าพระยา

สถานีวัดระดับน้ำ	ชื่อสถานีน้ำ	รายละเอียด
*E*08	คลองบางเขนใหม่	ระดับน้ำ (ม.รทก.)
*E*13	คลองสามเสน	ระดับน้ำ (ม.รทก.)
*E*15	คลองเทเวศร์	ระดับน้ำ (ม.รทก.)
HD	กองบัญชาการกองทัพเรือ	ระดับน้ำ (ม.), ระดับน้ำจาก HTA (ม.)
*E*20	คลองกรุงเกษม	ระดับน้ำ (ม.รทก.)
*E*23	คลองสาทร	ระดับน้ำ (ม.รทก.)
*E*29	คลองวัดไทร	ระดับน้ำ (ม.รทก.)
*E*37	คลองวัดด่าน	ระดับน้ำ (ม.รทก.)
*E*35	คลองช่องนนทรี	ระดับน้ำ (ม.รทก.)
*E*27	คลองเตย	ระดับน้ำ (ม.รทก.)
*E*28	คลองเจ็ก	ระดับน้ำ (ม.รทก.)
*E*36	คลองบางจาก	ระดับน้ำ (ม.รทก.)
*E*30	คลองบางอ้อ	ระดับน้ำ (ม.รทก.)
*E*31	คลองบางนา	ระดับน้ำ (ม.รทก.)
PC	ป้อมพระจุลจอมเกล้า	ระดับน้ำ (ม.), ระดับน้ำจาก HTA (ม.)

* ม. คือ เมตร; รทก. คือ ระดับน้ำทะเลปานกลาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ระดับน้ำทะเลปานกลาง เป็นการอ้างถึงความสูงของพื้นที่ต่าง ๆ โดยใช้ระดับน้ำทะเลเป็นเกณฑ์ แต่เนื่องจากน้ำทะเลมีขึ้นมีลง จึงต้องใช้ค่าเฉลี่ย ด้วยการเอาค่าระดับสูงสุดกับค่าระดับต่ำสุดมาบวกกันแล้วหารสอง

3.2 วิธีการทดลอง

ในงานวิจัยนี้จะใช้การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของแบบจำลองทั้ง 2 วิธี คือ Split Test และ Cross-validation โดยจะทำการ Split Test ทั้งหมด 5 ครั้ง (M. Galar, A. Fernandez, E. Barrenechea, H. Bustince and F. Herrera, 2016) (B. Chen, R. F. Harrison, K. Pasupa, P. Willett, D. J. Wilton and D. J. Wood, 2006) ในการทำ Split Test แต่ละครั้งจะแบ่งข้อมูลชุดสอนเป็นร้อยละ 30 ของข้อมูลทั้งหมด เพราะข้อมูลทั้งหมดมีจำนวนตัวอย่างถึง 15,000 ตัวอย่าง และมีปัจจัยทั้งหมด 17 ปัจจัย เมื่อนำมาทำอนุกรมเวลาย้อนหลัง 72 ชั่วโมง ทำให้มีแอตทริบิวต์ทั้งหมด 1,241 แอตทริบิวต์ เมื่อทำการสุ่มเลือกข้อมูลชุดสอน ข้อมูลชุดสอนจะมีจำนวนตัวอย่าง 4500 ตัวอย่าง แล้วนำข้อมูลชุดสอนมาทำ 5-fold cross-validation (M. Galar, A. Fernandez, E. Barrenechea, H. Bustince and F. Herrera, 2016) (K. M. Broms, M. B. Hooten and R. M. Fitzpatrick, 2016) จึงมีจำนวนตัวอย่างข้อมูลชุดสอนในแต่ละ fold เท่ากับ 3600 ตัวอย่าง เมื่อทำการเปรียบเทียบกับจำนวนแอตทริบิวต์กับจำนวนตัวอย่าง จำนวนตัวอย่างมากกว่าจำนวนแอตทริบิวต์ประมาณ 3 เท่า อยู่ในกรณีที่ยอมรับได้ เมื่อทำ 5 fold cross-validation ของแต่ละอัลกอริทึม มีรายละเอียดดังตารางที่ 3.2 เพื่อทำการเลือกชุดพารามิเตอร์ที่ดีที่สุด หลังจากนั้นนำข้อมูลชุดสอนทั้งหมดมาสร้างแบบจำลองโดยใช้พารามิเตอร์ที่ดีที่สุด แล้วนำข้อมูลทดสอบมาทดสอบกับแบบจำลอง ทำซ้ำทั้งหมด 5 ครั้ง โดยมีการสุ่มแบ่งข้อมูลที่แตกต่างกัน

ตารางที่ 3.2 การปรับค่าพารามิเตอร์ของแต่ละอัลกอริทึม

Algorithm	Kernel function	Parameter 1	Parameter 2
Regularized Least Square	Linear	$\lambda : 10^{6} - 10^{-6}$	-
	Radial basis function	$\lambda : 10^{6} - 10^{-6}$	$\sigma : 10^{6} - 10^{-6}$
Support Vector Regression	Linear	$C : 10^{6} - 10^{-6}$	-
	Radial basis function	$C : 10^{6} - 10^{-6}$	$\sigma : 10^{6} - 10^{-6}$
k-Nearest Neighbors	-	$k : 1 - 500$	
Random Forest	-	$nTree : 100 - 1000$	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
146469

จากตารางที่ 3.2 พารามิเตอร์ของอัลกอริทึม RLS ได้แก่ค่า λ ซึ่งค่า λ เป็นค่าคงที่ ที่ใช้ในการปรับค่า Regularized ถ้าค่า λ มีค่ามาก ทำให้เกิดปัญหาแบบจำลองนั้นไม่พอดีกับข้อมูล (Under-fitting) ถ้าค่า λ มีค่าน้อย ทำให้เกิดปัญหาพอดีเกินไป (Over-fitting) พารามิเตอร์ของอัลกอริทึม SVR ได้แก่ค่า C ซึ่งค่า C เป็นค่าคงที่เพื่อปรับความสมดุลของ Margin ดังนั้นควรปรับค่า C ให้เหมาะสมกับข้อมูลเพื่อได้ค่าพยากรณ์ใกล้เคียงกับข้อมูลจริงมากที่สุด พารามิเตอร์ของอัลกอริทึม k-NN ได้แก่ค่า k ซึ่งค่า k คือ จำนวนระเบียนของข้อมูลชุดสอนที่อยู่ใกล้กับข้อมูลทดสอบมากที่สุด พารามิเตอร์ของอัลกอริทึม RF ได้แก่ค่า *nTree* ซึ่งค่า *nTree* คือ จำนวนต้นไม้ตัดสินใจ ทั้งค่า k และ ค่า *nTree* ควรปรับให้เหมาะสมกับข้อมูลเพื่อได้ค่าพยากรณ์ใกล้เคียงกับข้อมูลจริงมากที่สุด ถ้าปรับค่ามากเกินไปจะทำให้ใช้เวลานานในการทดลอง พารามิเตอร์ของ RBF Kernel ได้แก่ค่า σ ซึ่งค่า σ คือ ค่าเบี่ยงเบนมาตรฐาน ซึ่งเป็นตัวกำหนดความโค้ง จะอยู่ในรูปแบบของกราฟทรงระฆังคว่ำ (Bell shape) ถ้าค่า σ มีค่ามากพื้นที่ใต้กราฟจะมีมาก ถ้าค่า σ มีค่าน้อยพื้นที่ใต้กราฟจะมีน้อย

3.3 การทดสอบความแม่นยำของผลพยากรณ์ที่ได้

ในงานวิจัยนี้ได้วัดความแม่นยำความของการพยากรณ์ โดยใช้ Root Mean Square Error (RMSE) ดังสมการ (3.1) ถ้า RMSE มีค่ามาก แสดงว่าแบบจำลองพยากรณ์ค่ามีความคลาดเคลื่อนมาก ในทางตรงกันข้ามถ้า RMSE มีค่าน้อย แสดงว่าแบบจำลองพยากรณ์ได้ใกล้เคียงกับค่าจริง

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3.1)$$

โดยที่ y_i คือ ค่าจริง

$\hat{y}_i$ คือ ค่าประมาณจากแบบจำลอง

n คือ จำนวนตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลอง

งานวิจัยนี้ มีการตั้งสมมติดังนี้

- สมมติฐานที่หนึ่ง (H_1): ระดับน้ำของสถานีน้ำที่มีผลต่อการพยากรณ์คือระดับน้ำของสถานีน้ำที่อยู่ต้นน้ำที่อยู่ใกล้ที่สุดกับสถานีที่พยากรณ์ (ทวีศักดิ์ วังไพศาล และ จักรฤทธิ์ ตรีนาจ, 2555)
- สมมติฐานที่สอง (H_2): อิทธิพลการขึ้น-ลงของน้ำทะเลมีผลต่อการพยากรณ์ เพราะกรุงเทพมหานคร ตั้งอยู่บนพื้นที่ราบลุ่มตอนปลายของแม่น้ำเจ้าพระยาและอยู่ภายใต้อิทธิพลการขึ้น-ลงของระดับน้ำทะเล (สำนักการระบายน้ำ, 2558)
- สมมติฐานที่สาม (H_3): การสูบน้ำและระบายน้ำลงแม่น้ำเจ้าพระยา เพื่อใช้ควบคุมระดับน้ำในเขตพื้นที่ เป็นปัจจัยที่มีผลต่อการพยากรณ์

4.1 ผลและวิเคราะห์การเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำท่า

4.1.1 ผลและวิเคราะห์การเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำท่าสถานีกองบัญชาการกองทัพเรือ

การเปรียบเทียบแบบจำลองที่ใช้ในการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือ ลำดับของอัลกอริทึมที่มีประสิทธิภาพมากที่สุดเรียงไปหาอัลกอริทึมที่มีประสิทธิภาพน้อยที่สุด คือ SVR-RBF, RLS-RBF, RF, RLS-Linear, SVR-Linear, k-NN และ HTA มีค่าเฉลี่ยของค่า RMSE แต่ละอัลกอริทึมดังนี้ 0.1555, 0.1564, 0.1795, 0.1862, 0.1883, 0.2569 และ 0.8924 ตามลำดับ การใช้ข้อมูลนำเข้าที่ทำให้การพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือแม่นยำมากสุดคือ HTA รวมกับ ระดับน้ำ 15 สถานี ย้อนหลัง 72 ชั่วโมง โดย SVR-RBF มีค่า RMSE = 0.1159 ดังตารางที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือ

ข้อมูล		ผลลัพธ์ (RMSE)						
		HTA	RLS		SVR		k-NN	RF
ข้อมูลเข้า	ผลลัพธ์		Linear	RBF	Linear	RBF		
HTA								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.8924	0.3213	0.2584	0.3263	0.2651	0.3980	0.2890
ย้อนหลัง 48 ชม.			0.3210	0.2447	0.3270	0.2338	0.3824	0.2825
ย้อนหลัง 72 ชม.			0.3201	0.2214	0.3258	0.2217	0.3777	0.2773
HTA+ ระดับน้ำกองบัญชาการ								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.8924	0.1424	0.1358	0.1434	0.1365	0.2325	0.1511
ย้อนหลัง 48 ชม.			0.1394	0.1304	0.1403	0.1308	0.2317	0.1470
ย้อนหลัง 72 ชม.			0.1396	0.1284	0.1403	0.1276	0.2382	0.1440
ระดับน้ำ 15 สถานี								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.8924	0.1576	0.1402	0.1593	0.1389	0.1975	0.1483
ย้อนหลัง 48 ชม.			0.1430	0.1285	0.1447	0.1284	0.2026	0.1449
ย้อนหลัง 72 ชม.			0.1397	0.1229	0.1431	0.1226	0.2121	0.1417
HTA + ระดับน้ำ 15 สถานี								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.8924	0.1412	0.1274	0.1406	0.1254	0.1958	0.1461
ย้อนหลัง 48 ชม.			0.1347	0.1207	0.1362	0.1192	0.2023	0.1425
ย้อนหลัง 72 ชม.			0.1344	0.1181	0.1348	**0.1159**	0.2124	0.1394
ค่าเฉลี่ยแต่ละอัลกอริทึม		0.8924	0.1862	0.1564	0.1883	**0.1555**	0.2569	0.1795

หมายเหตุ ตัวเลขพิมพ์หนา คือ การพยากรณ์ที่มีประสิทธิภาพมากที่สุด

4.1.2 ผลและวิเคราะห์การเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำท่าสถานีป้อมพระจุลจอมเกล้า

การเปรียบเทียบแบบจำลองที่ใช้ในการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือ ลำดับของอัลกอริทึมที่มีประสิทธิภาพมากที่สุดเรียงไปหาอัลกอริทึมที่มีประสิทธิภาพน้อยที่สุด คือ SVR-RBF, RLS-RBF, RLS-Linear, SVR-Linear, RF, k-NN และ HTA มีค่าเฉลี่ยของค่า RMSE แต่ละอัลกอริทึมดังนี้ 0.1275, 0.1281, 0.1402, 0.1416, 0.1527, 0.2854 และ 0.9755 ตามลำดับ การใช้ข้อมูลนำเข้าที่ทำให้การพยากรณ์ระดับน้ำสถานีป้อมพระจุลจอมเกล้าแม่นยำมากสุดคือ HTA รวมกับ ระดับน้ำ 15 สถานี ย้อนหลัง 72 ชั่วโมง โดย SVR-RBF มีค่า RMSE = 0.1122 ดังตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ผลการเปรียบเทียบแบบจำลองการพยากรณ์ระดับน้ำสถานีป้อมพระจุลจอมเกล้า

ข้อมูล		ผลลัพธ์ (RMSE)						
		HTA	RLS		SVR		k-NN	RF
ข้อมูลเข้า	ผลลัพธ์		Linear	RBF	Linear	RBF		
HTA								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.9755	0.1509	0.1380	0.1512	0.1387	0.3179	0.1659
ย้อนหลัง 48 ชม.			0.1449	0.1324	0.1451	0.1329	0.3213	0.1538
ย้อนหลัง 72 ชม.			0.1413	0.1304	0.1416	0.1318	0.3350	0.1489
HTA + ระดับน้ำป้อมพระจุลฯ								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.9755	0.1305	0.1240	0.1309	0.1238	0.2763	0.1527
ย้อนหลัง 48 ชม.			0.1231	0.1198	0.1233	0.1200	0.2874	0.1414
ย้อนหลัง 72 ชม.			0.1216	0.1188	0.1219	0.1197	0.3035	0.1358
ระดับน้ำ 15 สถานี								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.9755	0.1894	0.1554	0.1900	0.1546	0.2549	0.1716
ย้อนหลัง 48 ชม.			0.1508	0.1357	0.1536	0.1339	0.2653	0.1635
ย้อนหลัง 72 ชม.			0.1441	0.1275	0.1489	0.1256	0.2815	0.1585
HTA + ระดับน้ำ 15 สถานี								
ย้อนหลัง 24 ชม.	ล่วงหน้า 24 ชม.	0.9755	0.1313	0.1231	0.1325	0.1221	0.2453	0.1555
ย้อนหลัง 48 ชม.			0.1272	0.1167	0.1292	0.1151	0.2593	0.1454
ย้อนหลัง 72 ชม.			0.1278	0.1158	0.1309	**0.1122**	0.2776	0.1394
ค่าเฉลี่ยแต่ละอัลกอริทึม		0.9755	0.1402	0.1281	0.1416	**0.1275**	0.2854	0.1527

หมายเหตุ ตัวเลขพิมพ์หนา คือ การพยากรณ์ที่มีประสิทธิภาพมากที่สุด

จากการพยากรณ์ทั้ง 2 สถานี สถานีกองบัญชาการกองทัพเรือ และ สถานีป้อมพระจุลจอมเกล้า การพยากรณ์ที่ดีสุดคือ SVR-RBF การพยากรณ์ระดับน้ำโดยใช้การเรียนรู้ของเครื่องจักร มีประสิทธิภาพมากกว่า HTA

4.2 ผลและการวิเคราะห์ข้อมูลย้อนหลังที่มีผลต่อการพยากรณ์

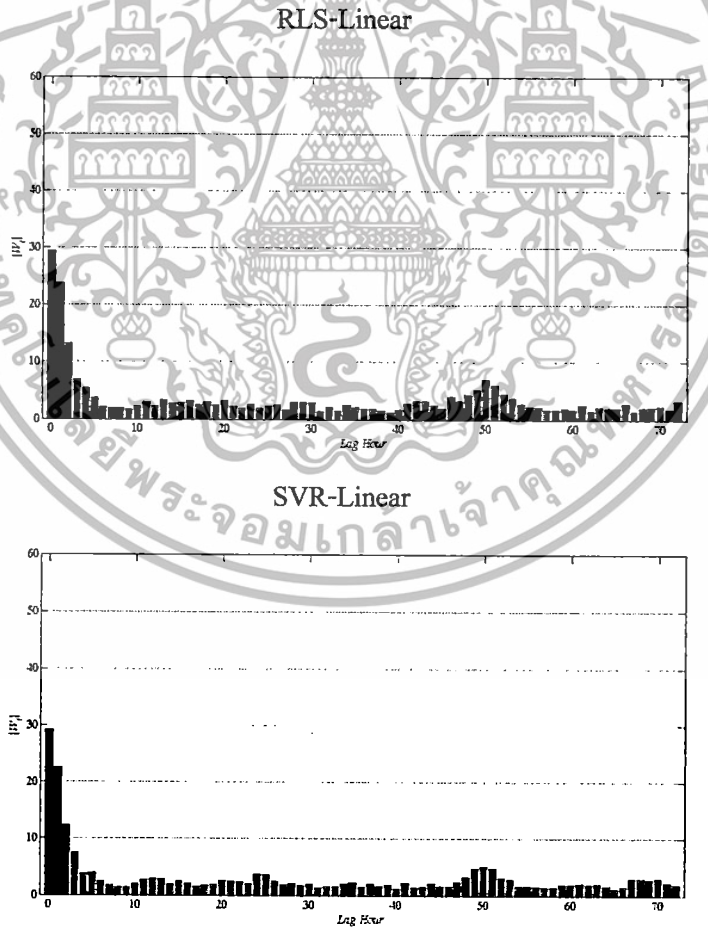
การวิเคราะห์ค่าสัมประสิทธิ์ของ RLS-Linear และ SVR-Linear ทำให้ทราบลำดับความสำคัญของชั่วโมงที่มีผลต่อการพยากรณ์ ถ้าค่าสัมประสิทธิ์มาก ณ ชั่วโมงนั้น แสดงว่าชั่วโมงนั้นมีผลต่อการพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1 ผลและการวิเคราะห์ข้อมูลย้อนหลังที่มีผลต่อการพยากรณ์สถานีกองบัญชาการกองทัพเรือ

จากรูปที่ 4.1 (บน) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม RLS-Linear จากรูปที่ 4.1 (ล่าง) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม SVR-Linear โดยทั้งสองอัลกอริทึมใช้ข้อมูลนำเข้า คือ HTA รวมกับ ระดับน้ำ 15 สถานี ที่ข้อมูลย้อนหลัง 72 ชั่วโมง และทั้งสองกราฟแสดงค่าสัมประสิทธิ์ ของแต่ละชั่วโมง จาก t, $t-1$, ... , $t-70$, $t-71$, $t-72$ เรียงลำดับจากซ้ายไปขวา จะเห็นได้ว่าชั่วโมงที่ส่งผลต่อการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือมากที่สุด คือ ชั่วโมงปัจจุบัน

ค่าสัมประสิทธิ์ ของอัลกอริทึม RLS-Linear และค่าสัมประสิทธิ์ ของอัลกอริทึม SVR-Linear ซึ่งจะมีค่าสัมประสิทธิ์ เป็นแนวโน้มเดียวกันทุก ๆ 24 ชั่วโมง และทั้งสองอัลกอริทึมมี ค่าสัมประสิทธิ์ แตกต่างกันเพียงเล็กน้อย แต่ว่าอัลกอริทึม RLS-Linear สามารถพยากรณ์สถานีกองบัญชาการกองทัพเรือได้แม่นยำมากกว่าอัลกอริทึม SVR-Linear ดังรูปที่ 4.1



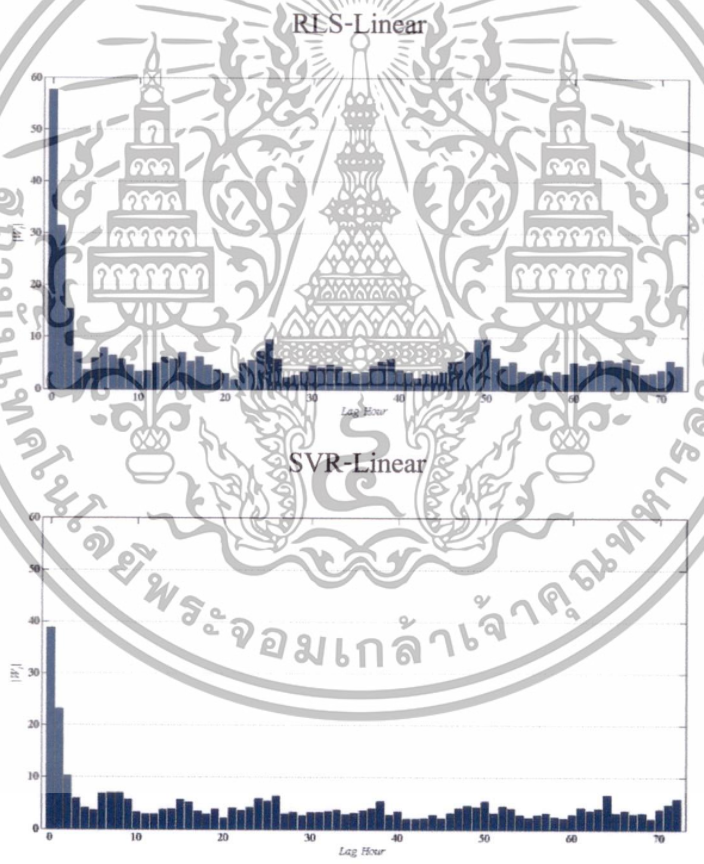
รูปที่ 4.1 แสดงค่าสัมประสิทธิ์ของข้อมูลนำเข้าย้อนหลัง 72 ชั่วโมงสถานีกองบัญชาการกองทัพเรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ผลและการวิเคราะห์ข้อมูลย้อนหลังที่มีผลต่อการพยากรณ์สถานีป้อมพระจุลจอมเกล้า

จากรูปที่ 4.2 (บน) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม RLS-Linear จากรูปที่ 4.2 (ล่าง) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม SVR-Linear โดยทั้งสองอัลกอริทึมใช้ข้อมูลนำเข้า คือ HTA รวมกับ ระดับน้ำ 15 สถานีที่ข้อมูลย้อนหลัง 72 ชั่วโมง และทั้งสองกราฟแสดงค่าสัมประสิทธิ์ของแต่ละชั่วโมง จาก t, $t-1$, ... , $t-70$, $t-71$, $t-72$ เรียงลำดับจากซ้ายไปขวา จะเห็นได้ว่าชั่วโมงที่ส่งผลต่อการพยากรณ์ระดับน้ำสถานีป้อมพระจุลจอมเกล้ามากที่สุด คือ ชั่วโมงปัจจุบัน

ค่าสัมประสิทธิ์ของอัลกอริทึม RLS-Linear และค่าสัมประสิทธิ์ของอัลกอริทึม SVR-Linear ซึ่งจะมีค่าสัมประสิทธิ์ เป็นแนวโน้มเดียวกันทุก ๆ 24 ชั่วโมง และทั้งสองอัลกอริทึมมี ค่าสัมประสิทธิ์ แตกต่างกันเพียงเล็กน้อย แต่ว่าอัลกอริทึม RLS-Linear สามารถพยากรณ์สถานีป้อมพระจุลจอมเกล้าได้แม่นยำมากกว่าอัลกอริทึม SVR-Linear ดังรูปที่ 4.2



รูปที่ 4.2 แสดงค่าสัมประสิทธิ์ของข้อมูลนำเข้าย้อนหลัง 72 ชั่วโมงสถานีป้อมพระจุลจอมเกล้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลและการวิเคราะห์ปัจจัยที่มีผลต่อการพยากรณ์

การวิเคราะห์ค่าสัมประสิทธิ์ของ RLS-Linear และ SVR-Linear ทำให้ทราบลำดับความสำคัญของปัจจัยที่มีผลต่อการพยากรณ์ ถ้าค่าสัมประสิทธิ์มากที่ปัจจัยนั้น แสดงว่าปัจจัยนั้นมีผลต่อการพยากรณ์

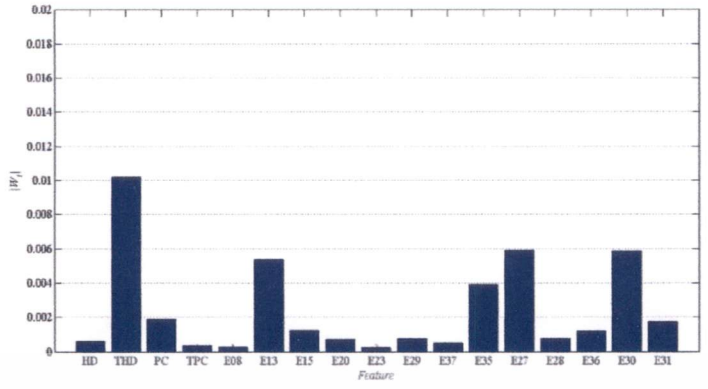
4.3.1 ผลและการวิเคราะห์ปัจจัยที่มีผลต่อการพยากรณ์สถานีกองบัญชาการกองทัพเรือ

จากรูปที่ 4.3 (บน) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม RLS-Linear จากรูปที่ 4.3 (ล่าง) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม SVR-Linear โดยทั้งสองอัลกอริทึมใช้ข้อมูลนำเข้า HTA รวมกับ ระดับน้ำ 15 สถานีที่ข้อมูลย้อนหลัง 72 ชั่วโมง กราฟทั้งสองแสดงค่าสัมประสิทธิ์ของแต่ละปัจจัย ได้แก่ HD, HD_{HTA}, PC, PC_{HTA}, $E08$, $E13$, $E15$, $E20$, $E23$, $E29$, $E37$, $E35$, $E27$, $E28$, $E36$, $E30$ และ $E31$ เรียงลำดับจากซ้ายไปขวา จากการสร้างแบบจำลองโดยใช้อัลกอริทึม RLS-Linear ปัจจัยที่ส่งผลต่อการพยากรณ์ของสถานีกองบัญชาการกองทัพเรือคือ HD_{HTA}, $E13$, $E35$, $E27$ และ $E30$ และจากการสร้างแบบจำลองโดยใช้อัลกอริทึม SVR-Linear ปัจจัยที่ส่งผลต่อการพยากรณ์ของสถานีกองบัญชาการกองทัพเรือคือ HD, $E13$, $E15$, $E30$ และ PC

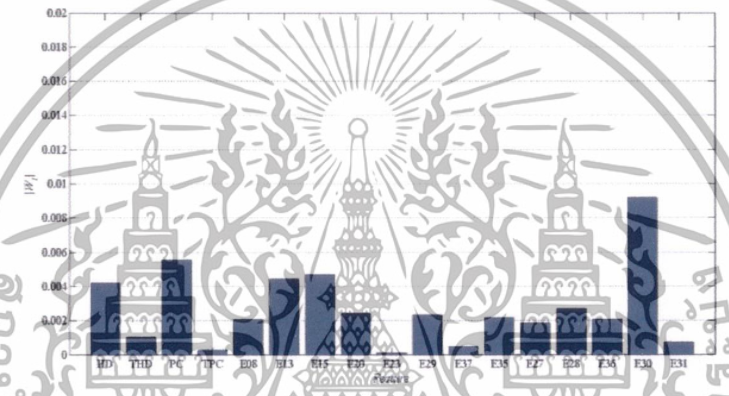
ปัจจัยที่ส่งผลอย่างเห็นได้ชัดต่อการพยากรณ์ของสถานีกองบัญชาการกองทัพเรือทั้งหมดมีดังนี้ HD, HD_{HTA}, $E13$, $E15$, $E35$, $E27$, $E30$ และ PC ปัจจัยที่ $E13$ และ $E15$ ตามสมมติฐานที่ H_1 คือ ระดับน้ำของสถานีน้ำที่อยู่ต้นน้ำที่อยู่ใกล้ที่สุดกับสถานีที่พยากรณ์มีผลต่อการพยากรณ์ ปัจจัยที่ HD, HD_{HTA}, $E30$ และ PC ตามสมมติฐานที่ H_2 คือ อิทธิพลการขึ้น-ลงของน้ำทะเลมีผลต่อการพยากรณ์ ปัจจัยที่ $E27$ ตามสมมติฐานที่ H_3 คือ การสูบน้ำและระบายน้ำ มีผลต่อการพยากรณ์ เพราะว่า $E27$ อยู่ใกล้กับสถานีพระโขนงมาก และสถานีพระโขนงเป็นสถานีสูบน้ำที่ใหญ่ที่สุดสามารถสูบน้ำออกได้ 173 ลูกบาศก์เมตร/วินาที จึงทำให้สถานีใกล้เคียงมีระดับน้ำที่สูงขึ้นด้วย (สำนักการระบายน้ำ, 2558) ปัจจัยที่ $E35$ ตามสมมติฐานที่ H_2 และสมมติฐานที่ H_3 เพราะว่า $E35$ อยู่ใกล้คลองลัดโพธิ์ทำให้ได้รับอิทธิพลการขึ้น-ลงของน้ำทะเล (สำนักงานทรัพย์สินส่วนพระมหากษัตริย์, 2554) และอยู่ใกล้สถานีพระโขนง จึงทำให้การสูบน้ำและระบายน้ำมีผลต่อการพยากรณ์ (สำนักการระบายน้ำ, 2558) ดังรูปที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

RLS-Linear



SVR-Linear



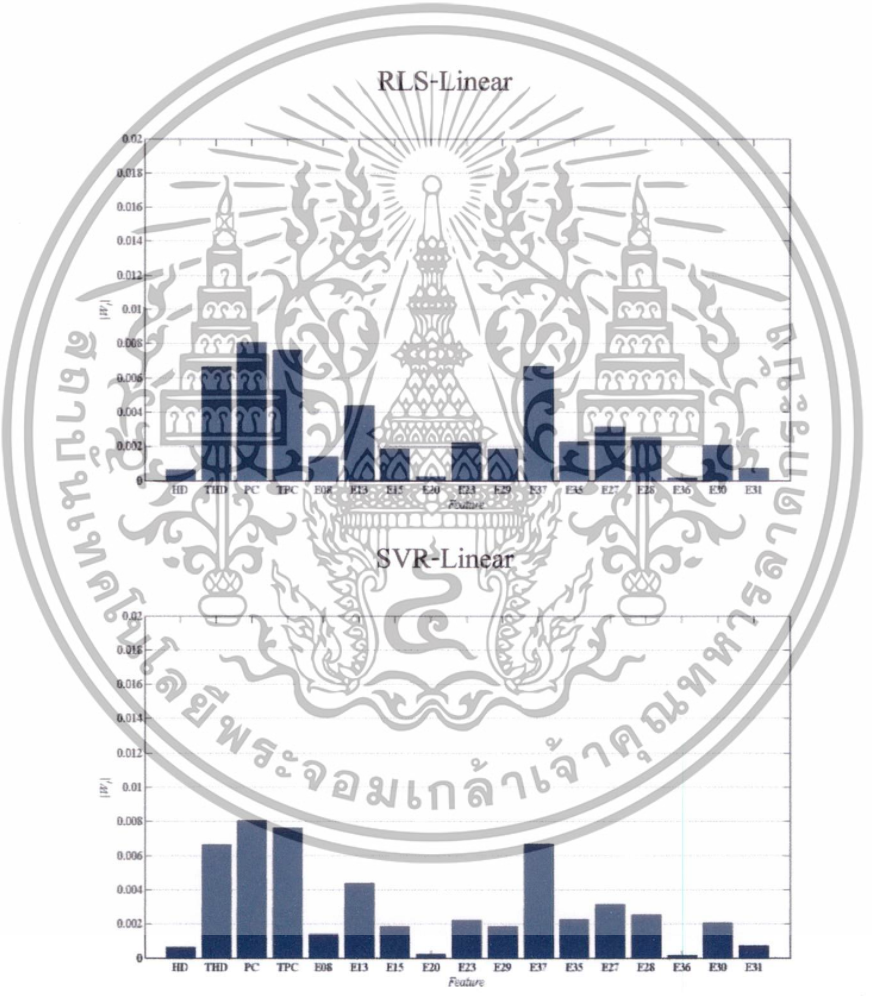
รูปที่ 4.3 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัยของสถานีกองบัญชาการกองทัพเรือ

4.3.2 ผลและการวิเคราะห์ปัจจัยที่มีผลต่อการพยากรณ์สถานีป้อมพระจุลจอมเกล้า

จากรูปที่ 4.4 (บน) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม RLS-Linear จากรูปที่ 4.4 (ล่าง) แสดงค่าสัมประสิทธิ์ของอัลกอริทึม SVR-Linear โดยทั้งสองอัลกอริทึมใช้ข้อมูลนำเข้า HTA รวมกับ ระดับน้ำ 15 สถานีที่ข้อมูลย้อนหลัง 72 ชั่วโมง กราฟทั้งสองแสดงค่าสัมประสิทธิ์ของแต่ละปัจจัย ได้แก่ HD, HD_{HTA}, PC, PC_{HTA}, $E08$, $E13$, $E15$, $E20$, $E23$, $E29$, $E37$, $E35$, $E27$, $E28$, $E36$, $E30$ และ $E31$ เรียงลำดับจากซ้ายไปขวา จากการสร้างแบบจำลองโดยใช้อัลกอริทึม RLS-Linear ปัจจัยที่ส่งผลต่อการพยากรณ์ของสถานีป้อมพระจุลจอมเกล้า คือ HD_{HTA}, PC, PC_{HTA} และ $E37$ และจากการสร้างแบบจำลองโดยใช้อัลกอริทึม SVR-Linear ปัจจัยที่ส่งผลต่อการพยากรณ์ของสถานีป้อมพระจุลจอมเกล้า คือ $E23$, $E35$, $E36$ และ PC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัจจัยที่ส่งผลอย่างเห็นได้ชัดต่อการพยากรณ์ของสถานีป้อมพระจุลจอมเกล้าทั้งหมดมีดังนี้ HD_{HTA}, $E23$, $E37$, $E35$, $E36$, PC_{HTA} และ PC ปัจจัยที่ $E23$, $E37$, $E35$ และ $E36$ ตามสมมติฐานที่ H_1 คือ ระดับน้ำของสถานีน้ำที่อยู่ต้นน้ำที่อยู่ใกล้ที่สุดกับสถานีที่พยากรณ์มีผลต่อการพยากรณ์ เพราะว่า 3 สถานีคือ $E23$, $E37$ และ $E35$ อยู่ใกล้กับคลองลัดโพธิ์ ซึ่งคลองลัดโพธิ์เป็นคลองที่สร้างขึ้นมาเพื่อระบายน้ำให้ลงทะเลเร็วที่สุดและเส้นทางคลองลัดโพธิ์เป็นเส้นทางน้ำที่สั้นกว่าทางปกติ (สำนักงานทรัพย์สินส่วนพระมหากษัตริย์, 2554) ปัจจัยที่ HD_{HTA}, $E30$, PC และ PC_{HTA} ตามสมมติฐานที่ H_2 คือ อิทธิพลการขึ้น-ลงของน้ำทะเลมีผลต่อการพยากรณ์ ดังรูปที่ 4.4



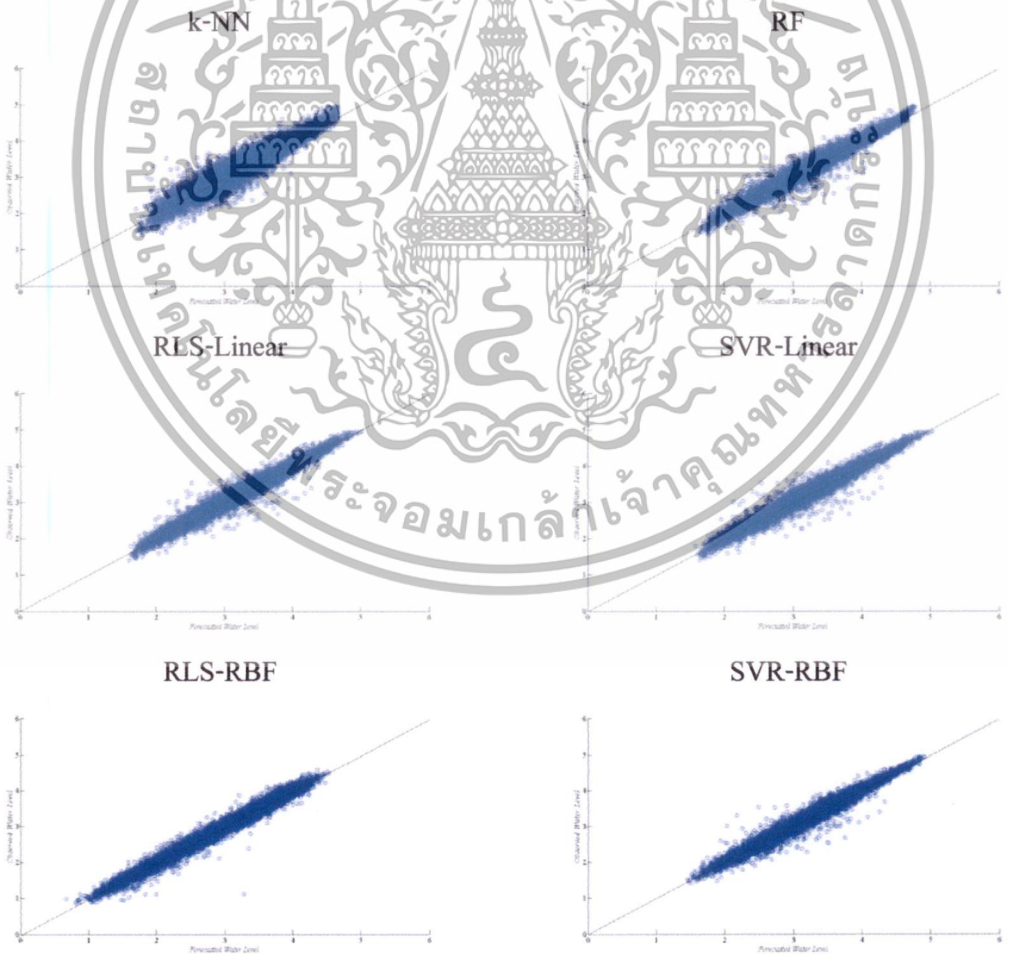
รูปที่ 4.4 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัยของสถานีป้อมพระจุลจอมเกล้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 ผลพยากรณ์ระดับน้ำ

4.4.1 ผลการพยากรณ์ระดับน้ำสถานีกองบัญชาการกองทัพเรือกับระดับน้ำจริง

จากรูปที่ 4.5 เป็นกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีกองบัญชาการกองทัพเรือ โดยมีอัลกอริทึม ดังต่อไปนี้ k-NN, RF, RLS-Linear, SVR-Linear, RLS-RBF และ SVR-RBF ใช้ข้อมูล HTA รวมกับระดับน้ำ 15 สถานี ข้อมูลย้อนหลัง 72 ชั่วโมง เมื่อเปรียบเทียบจากรูปที่ 4.5 อัลกอริทึมที่สามารถพยากรณ์สถานีกองบัญชาการกองทัพเรือมีประสิทธิภาพน้อยที่สุด คือ k-NN เพราะเกิดความคลาดเคลื่อนในการพยากรณ์มากกว่าอัลกอริทึมอื่น ๆ สังเกตได้จากผลพยากรณ์ที่กระจายตัวอยู่นอกเส้นตรงเป็นจำนวนมาก และอัลกอริทึมที่สามารถพยากรณ์สถานีกองบัญชาการกองทัพเรือมีประสิทธิภาพมากที่สุดคือ SVR-RBF เพราะพยากรณ์ได้แม่นยำมากที่สุดมากว่าอัลกอริทึมอื่น ๆ สังเกตได้จากผลพยากรณ์ส่วนใหญ่อยู่บนเส้นตรง มีค่าที่อยู่ห่างจากเส้นตรงเพียงเล็กน้อยเท่านั้น ดังรูปที่ 4.5

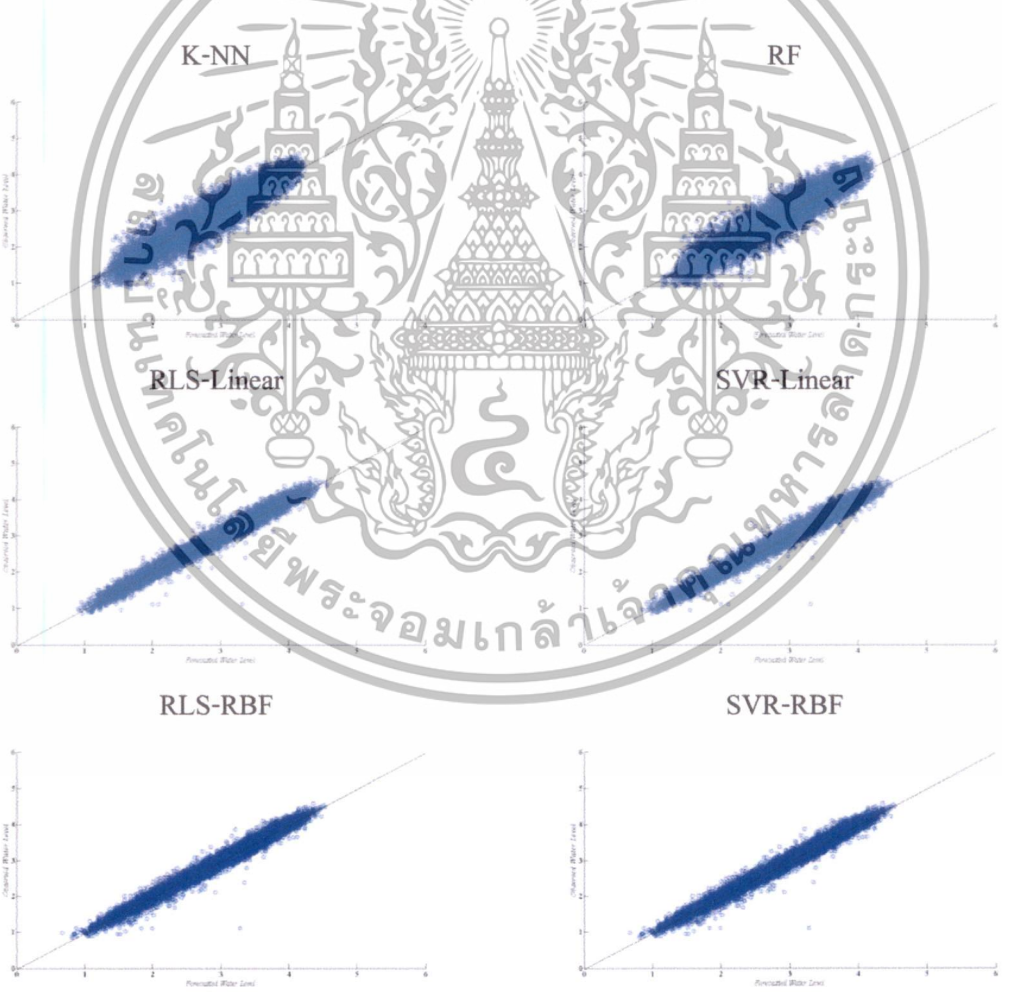


รูปที่ 4.5 แสดงกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีกองบัญชาการกองทัพเรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 ผลการพยากรณ์ระดับน้ำสถานีป้อมพระจุลจอมเกล้ากับระดับน้ำจริง

จากรูปที่ 4.6 เป็นกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีป้อมพระจุลจอมเกล้า โดยมีอัลกอริทึม ดังต่อไปนี้ k-NN, RF, RLS-Linear, SVR-Linear, RLS-RBF และ SVR-RBF ใช้ข้อมูล HTA รวมกับระดับน้ำ 15 สถานี ข้อมูลย้อนหลัง 72 ชั่วโมง เมื่อเปรียบเทียบจากรูปที่ 4.6 อัลกอริทึมที่สามารถพยากรณ์สถานีป้อมพระจุลจอมเกล้ามีประสิทธิภาพน้อยที่สุด คือ k-NN เพราะเกิดความคลาดเคลื่อนในการพยากรณ์มากกว่าอัลกอริทึมอื่น ๆ สังเกตได้จากผลพยากรณ์ที่กระจายตัวอยู่นอกเส้นตรงเป็นจำนวนมาก และอัลกอริทึมที่สามารถพยากรณ์สถานีป้อมพระจุลจอมเกล้ามีประสิทธิภาพมากที่สุดคือ SVR-RBF เพราะพยากรณ์ได้แม่นยำมากที่สุดมากกว่าอัลกอริทึมอื่น ๆ สังเกตได้จากผลพยากรณ์ส่วนใหญ่อยู่บนเส้นตรง มีค่าที่อยู่ห่างจากเส้นตรงเพียงเล็กน้อยเท่านั้น ดังรูปที่ 4.6



รูปที่ 4.6 แสดงกราฟแสดงผลระหว่าง ระดับน้ำจริง กับ ระดับน้ำพยากรณ์ ของสถานีป้อมพระจุลจอมเกล้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาวิธีการสร้างแบบจำลองที่สามารถพยากรณ์ระดับน้ำให้มีประสิทธิภาพดีขึ้นจากการพยากรณ์ระดับน้ำโดยใช้วิธี HTA โดยการทดลอง จะมีข้อมูล 4 แบบ

- ระดับน้ำพยากรณ์จาก HTA 2 สถานี เพียงอย่างเดียว
- ระดับน้ำพยากรณ์จาก HTA รวมกับ ระดับน้ำของสถานีที่ต้องการพยากรณ์
- ระดับน้ำของสถานีวัดระดับน้ำ 15 สถานี เพียงอย่างเดียว
- ระดับน้ำพยากรณ์จาก HTA 2 สถานี รวมกับ ระดับน้ำของสถานีวัดระดับน้ำ 15 สถานี

แล้วนำข้อมูลทั้ง 4 แบบ มาทำอนุกรมเวลาย้อนหลัง 24 48 และ 72 แล้วทำการสร้างแบบจำลองพยากรณ์ระดับน้ำท่า ด้วยอัลกอริทึมของการเรียนรู้ของเครื่องจักร 4 ชนิด ได้แก่ RLS, SVR, k-NN และ RF

จากผลการทดลองสรุปได้ดังนี้

1. ข้อมูลเข้าที่ส่งผลต่อการพยากรณ์ระดับน้ำท่าทั้ง 2 สถานี คือ HTA รวมกับ ระดับน้ำ 15 สถานี ย้อนหลัง 72 ชั่วโมง
2. ชั่วโมงที่ส่งต่อการพยาการณ์ระดับน้ำท่าทั้ง 2 สถานีคือ เวลา ณ ปัจจุบัน (t_0)
3. การพยากรณ์ระดับน้ำท่า ด้วย 4 อัลกอริทึมของการเรียนรู้เครื่องจักรสามารถพยากรณ์ระดับน้ำท่าได้ดีกว่าการใช้ HTA และ อัลกอริทึมที่ทำการพยากรณ์ได้แม่นยำมากที่สุด คือ SVR แสดงว่าแบบจำลองที่ถูกสร้างขึ้นด้วยอัลกอริทึมแบบไม่เชิงเส้น มีประสิทธิภาพดีกว่าแบบจำลองที่สร้างด้วยอัลกอริทึมแบบเชิงเส้น

ข้อเสนอแนะ สามารถสร้างแบบจำลองระดับน้ำท่าในแม่น้ำต่าง ๆ เพื่อช่วยในการจัดการบริหารน้ำ เช่น การจัดสรรน้ำเพื่อพื้นที่การเกษตร การพยากรณ์ระดับน้ำท่าเพื่อป้องกันและเตือนภัยน้ำท่วม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กรมชลประทาน. 2554. **สรุปบทเรียนการบริหารจัดการน้ำ ลุ่มน้ำเจ้าพระยาปี พ.ศ. 2554 และแนวทางในการบรรเทาอุทกภัยโดยการขุดลอกแม่น้ำเจ้าพระยา.** วันที่ค้นข้อมูล 25 ธันวาคม 2558, [Online] Available: http://kmcenter.rid.go.th/kmc12/055.pdf.

สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร. 2556. **พื้นที่การเกษตรที่เหมาะสมกับการเพาะปลูกลุ่มน้ำเจ้าพระยา.** วันที่ค้นข้อมูล 25 ธันวาคม 2558, [Online] Available: http://www.haii.or.th/wiki/index.php/พื้นที่การเกษตรที่เหมาะสมกับการเพาะปลูกลุ่มน้ำเจ้าพระยา.

วิชัย สุภาโสด. 2549. **แนวคิดการพัฒนาทรัพยากรน้ำเพื่อการเกษตร.** วันที่ค้นข้อมูล 25 ธันวาคม 2558, [Online] Available: http://www.rid.go.th/thaicid/_5_article/2549/03Water_Agri.pdf.

เชิดวงศ์ แสงศุภวานิช. 2555. **ระดับน้ำ (Tide and Water level).** วันที่ค้นข้อมูล 25 ธันวาคม 2558, [Online] Available: http://www.imc.src.ku.ac.th/personnel/file_subject/2012115847261.pdf

ทวีศักดิ์ วังไพศาล และ จักรฤทธิ์ ตรีนาจ. 2555. **ปัจจัยที่มีผลต่อความแม่นยำในการพยากรณ์ระดับน้ำหลากที่สถานีวัดระดับน้ำ M.7 โดยใช้แบบจำลองโครงข่ายประสาทเทียม.** การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช.

นิธินันท์ มาตา, แสงดาว นพพิทักษ์ และ ศิโรรัตน์ กุลวงศ์. 2558. **การค้นหาปัจจัยเพื่อสร้างโมเดลสำหรับพยากรณ์การควบคุมประตูระบายน้ำ.** การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ

สำนักการระบายน้ำ. 2558. **แผนปฏิบัติการป้องกันและแก้ไขปัญหาน้ำท่วมกรุงเทพมหานครประจำปี 2558.** วันที่ค้นข้อมูล 3 มีนาคม 2559, [Online] Available: http://203.155.220.119/News_dds/magazine/Plan58/Plan2558_1.pdf

สำนักงานทรัพย์สินส่วนพระมหากษัตริย์. 2554. **แนวพระราชดำริในมหานคร คลองลัดโพธิ์ และสะพานภูมิพล.** วันที่ค้นข้อมูล 4 มีนาคม 2559, [Online] Available: http://www.crownproperty.or.th/public/upload/media/Recommended%20Books/9572d55acb8a4e07a1ef1519662f5a6a.pdf

ปริญญา สงวนสัตย์. 2558. **การเรียนรู้ของเครื่อง (Machine Learnning).** คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์

เอกสิทธิ์ พัชรวงศ์ศักดา. 2557. **การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น.** วิทยาศาสตร์คอมพิวเตอร์ สถาบันเทคโนโลยีนานาชาติสิรินธร มหาวิทยาลัยธรรมศาสตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม (ต่อ)

S. Phuphong and C. Surussavadee. 2013. **An Artificial Neural Network Based Runoff Forecasting Model in the Absence of Precipitation Data: A Case Study of Khlong U-Tapao River Basin, Songkhla Province, Thailand**. Intelligent System, Modelling and Simulation.

P. Markpeng, P. Wongnimmarn, N. Champreeda, P. Vateekul and K. Sarinnapakorn. 2014. **Controlling Quality of Water-Level Data in Thailand** . Intelligent Information and Database Systems.

S. Alex and S. Bernhard. 2003. **A Tutorial on Support Vector Regression** . วันที่ค้นข้อมูล 3 มีนาคม 2559, [Online] Available: http://alex.smola.org/papers/2003/SmoSch03b.pdf

W. Chun-Hsin, H. Jan-Ming and L. D.T. 2004. **Travel-Time Prediction With Support Vector Regression**. IEEE Transactions on Intelligent Transportation Systems.

M. Galar, A. Fernandez, E. Barrenechea, H. Bustince and F. Herrera. 2016. **New Ordering Based Pruning Pruning Metrics for Ensembles of Classifiers in Imbalanced Datasets.** Proceeding of the 9th International Conference on Computer Recognition Systems CORES 2015.

K. M. Broms, M. B. Hooten and R. M. Fitzpatrick. 2016. **Model selection and assessment for multi-species occupancy models.** Ecology.

B. Chen, R. F. Harrison, K. Pasupa, P. Willett, D. J. Wilton and D. J. Wood. 2006. **Virtual Screening Using Binary Kernel Discrimination: Effect of Noisy Training Data and the Optimization of Performance**. Journal of chemical information and modeling.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตอนที่ 4 ข้อเสนอแนะเพื่อปรับปรุงเจตคติที่มีต่อความสูญเปล่าในการผลิต

คำชี้แจง : โปรดแสดงความคิดเห็นและข้อเสนอแนะเพื่อปรับปรุงเจตคติที่มีต่อความสูญเปล่าในการผลิตที่ท่านคิดเห็นว่าเป็นประโยชน์ต่อองค์การหรือหน่วยงาน

ขอขอบพระคุณที่ท่านกรุณาตอบแบบสอบถามในครั้งนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาวสิริเพ็ญ จึงเจริญรัตน์
สถานที่เกิด	จังหวัดชลบุรี
ที่อยู่	81/13 หมู่ 8 ตำบลบ้านสวน อำเภอเมือง จังหวัดชลบุรี
ประวัติการศึกษา	คณะวิทยาศาสตร์
	สาขาวิทยาการคอมพิวเตอร์
	มหาวิทยาลัยบูรพา
ประวัติการทำงาน	
ธ.ค. 50 – ก.พ. 56	บริษัท โควา ซิสเท็ม (ประเทศไทย) จำกัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้