

# การเพิ่มประสิทธิภาพซอฟต์แวร์ตัวเอกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมคอมพิวเตอร์  
มหาวิทยาลัยเทคโนโลยีสุรนารี  
ปีการศึกษา 2559

**THE IMPROVEMENT OF SUPPORT VECTOR  
REGRESSION TO FORECAST  
TIME SERIES**



**A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Master of Engineering in Computer Engineering  
Suranaree University of Technology**

**Academic Year 2016**

## การเพิ่มประสิทธิภาพซอฟต์แวร์ตัวเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วนหนึ่งของการศึกษา  
ตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

(รศ. ดร.กิตติศักดิ์ เกิดประสพ)

ประธานกรรมการ

(รศ. ดร.นิตยา เกิดประสพ)

กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

(ผศ. ดร.ปรเมศวร์ ห่อแก้ว)

กรรมการ

(ศ. ดร.ชูกิจ ลิมปิจำนงค์)

รองอธิการบดีฝ่ายวิชาการและนวัตกรรม

(รศ. ร.อ. ดร.กนต์ธร ชำนิประศาสน์)

คณบดีสำนักวิชาวิศวกรรมศาสตร์

ธีรวัช แก้ววิจิตร : การเพิ่มประสิทธิภาพซอฟต์แวร์เวกเตอร์รีเกรสชันในการพยากรณ์  
อนุกรมเวลา (THE IMPROVEMENT OF SUPPORT VECTOR REGRESSION TO  
FORECAST TIME SERIES) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.นิตยา เกิดประสพ,  
81 หน้า

ในปัจจุบันได้มีความพยายามในการหาเทคนิคใหม่ในการพยากรณ์ เพื่อให้การพยากรณ์มีความแม่นยำและความเร็วเพิ่มขึ้น โดยการคิดค้นเทคนิคใหม่ หรือการนำหลาย ๆ เทคนิคมาผสมกัน ซึ่งงานวิจัยนี้มีจุดประสงค์เพื่อเพิ่มประสิทธิภาพให้กับเทคนิคซอฟต์แวร์เวกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา โดยใช้ค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำให้กับตัวแบบ ซึ่งข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลอนุกรมเวลาทั้งหมด 5 ชุดข้อมูลประกอบไปด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิชเซอร์ ข้อมูลปริมาณการผลิตน้ำมันของวูวินแต่ละเดือน ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองดาร์วิน ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาสีติและดาร์วิน โดยการเปรียบเทียบความแม่นยำของเทคนิคใหม่จะทำการเปรียบเทียบกับเทคนิค 2 แบบคือเทคนิคซอฟต์แวร์เวกเตอร์รีเกรสชันดั้งเดิม และเทคนิคอาร์มา และวัดค่าโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าสหสัมพันธ์ ซึ่งจากผลการเปรียบเทียบพบว่าเทคนิคใหม่สามารถเพิ่มความแม่นยำให้กับเทคนิคซอฟต์แวร์เวกเตอร์รีเกรสชันได้

สาขาวิชา วิศวกรรมคอมพิวเตอร์  
ปีการศึกษา 2559

ลายมือชื่อนักศึกษา \_\_\_\_\_  
ลายมือชื่ออาจารย์ที่ปรึกษา \_\_\_\_\_

TEETAWAT KAEWWIJIT : THE IMPROVEMENT OF SUPPORT  
VECTOR REGRESSION TO FORECAST TIME SERIES.

THESIS ADVISOR : ASSOC. PROF. NITTAYA KERDPRASOP,  
Ph.D., 81 PP.

### SUPPORT VECTOR REGRESSION/FORECAST/TIME SERIES

Currently, there are efforts to find new techniques in forecasting in order to improve precision and speed. The improvement is achieved by using new technique or a combination of techniques.

This research aims to optimize support vector regression in forecasting time series by using the error to increase the accuracy of the model. The five datasets used in time series analysis are the daily temperature of the Fisher River, monthly milk production, the sea level pressure data at Darwin, carbon dioxide concentration at Mauna Loa mountain, and the atmospheric pressure difference between Tahiti and Darwin. The precision of the proposed model is compared against the traditional support vector regression and the ARIMA models using the Mean Squared Error, root mean squared error, Mean Absolute Percentage Error, mean absolute error and the Correlation Coefficient metrics. From the experimental results, the proposed method can improve precision of the support vector regression technique.

School of Computer Engineering

Academic Year 2016

Student's Signature \_\_\_\_\_

Advisor's Signature \_\_\_\_\_

## กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ บุคคล และกลุ่มบุคคลที่ได้  
กรุณาให้คำปรึกษา แนะนำ ช่วยเหลืออย่างดียิ่ง ทั้งในด้านวิชาการ และด้านการดำเนินงานวิจัย  
ดังต่อไปนี้

รองศาสตราจารย์ ดร.นิตยา เกิดประสพ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำปรึกษาใน  
การทำงานวิจัย การทำวิทยานิพนธ์ และตรวจสอบความถูกต้องในการทำวิทยานิพนธ์

รองศาสตราจารย์ ดร.กิตติศักดิ์ เกิดประสพ ที่ให้คำปรึกษาในการทำงานวิจัย การทำ  
วิทยานิพนธ์ และตรวจสอบความถูกต้องในการทำวิทยานิพนธ์

ขอขอบคุณพี่ ๆ และเพื่อน ๆ นักศึกษาทั้งปริญญาโทและปริญญาเอก ที่ช่วยให้คำปรึกษา  
และกำลังใจในการทำวิทยานิพนธ์

สุดท้ายขอกราบขอบพระคุณบิดา มารดา ที่ให้กำเนิด ให้ความรู้ อบรมเลี้ยงดู และส่งเสริม  
ทางการศึกษาเป็นอย่างดี จนผู้วิจัยมีความรู้และความสามารถในการทำวิทยานิพนธ์ และประสบ  
ความสำเร็จในชีวิต

ธีร์ธวัช แก้ววิจิตร

มหาวิทยาลัยเทคโนโลยีสุรนารี

## สารบัญ

	หน้า
บทคัดย่อ (ภาษาไทย).....	ก
บทคัดย่อ (ภาษาอังกฤษ) .....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ช
สารบัญรูป.....	ฉ
<b>บทที่</b>	
<b>1 บทนำ.....</b>	<b>1</b>
1.1 ความสำคัญและที่มาของงานวิจัย.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตของงานวิจัย.....	3
1.4 ประโยชน์ที่จะได้รับ.....	3
<b>2 ปรัชญ่วรรณกรรม.....</b>	<b>4</b>
2.1 อนุกรมเวลา.....	4
2.1.1 ความหมายของอนุกรมเวลา.....	4
2.1.2 ส่วนประกอบของอนุกรมเวลา.....	5
2.1.3 การวิเคราะห์อนุกรมเวลา.....	7
2.1.4 อนุกรมเวลาที่มีลักษณะนิ่ง.....	9
2.2 การวิเคราะห์อนุกรมเวลาด้วยอาร์มา.....	10
2.2.1 ประวัติของอาร์มา.....	10
2.2.2 ตัวแบบอาร์มา.....	10
2.2.3 ตัวแบบอาร์มา.....	11
2.3 การพยากรณ์ด้วยซัพพอร์ตเวกเตอร์รีเกรสชัน.....	12
2.3.1 ประวัติของซัพพอร์ตเวกเตอร์รีเกรสชัน.....	12
2.3.2 สมการซัพพอร์ตเวกเตอร์รีเกรสชัน.....	12

## สารบัญ (ต่อ)

	หน้า	
2.3.3	ท่อเปิดซีลอนสำหรับกำหนดชีพพอร์ตเวกเตอร์.....	12
2.3.4	สมการของลากรานจ์.....	13
2.3.5	เคอร์เนลที่ใช้ในชีพพอร์ตเวกเตอร์รีเกรสชัน.....	14
2.4	เทคนิคอื่นที่ใช้ช่วยในการพยากรณ์อนุกรมเวลา.....	15
2.4.1	เทคนิคไฮบริด.....	15
2.4.2	โครงข่ายประสาทเทียม.....	16
2.4.3	ขั้นตอนวิธีเชิงพันธุกรรม.....	17
2.4.4	การแปลงเวฟเล็ตแบบไม่ต่อเนื่อง.....	18
2.4.5	ขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค.....	19
2.5	เกณฑ์ที่ใช้ในการวัดค่าความแม่นยำของโมเดล.....	20
2.5.1	ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย.....	20
2.5.2	ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย.....	21
2.5.3	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย.....	22
2.5.4	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน.....	23
2.5.5	ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน.....	25
2.5.6	ค่าสหสัมพันธ์.....	26
2.5.7	ค่าคลาดเคลื่อนสัมพัทธ์.....	27
2.5.8	ค่าความผิดพลาดมาตรฐานของการทำนาย.....	28
2.5.9	ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนสมมาตร.....	29
2.6	งานวิจัยที่เกี่ยวข้อง.....	31
3	วิธีดำเนินการวิจัย.....	34
3.1	กรอบแนวคิดของการวิจัย.....	34
3.2	ตัวอย่างการสร้างโมเดลเพื่อการพยากรณ์.....	37
3.2.1	ชุดข้อมูลอนุกรมเวลา.....	37
3.2.2	ขั้นตอนสร้างตัวแบบ SVR จากชุดข้อมูล.....	38



## สารบัญ (ต่อ)

	หน้า
3.2.3 ขั้นตอนคำนวณหาค่าความคลาดเคลื่อนของตัวแบบ SVR .....	38
3.2.4 ขั้นตอนสร้างตัวแบบ SVR จากค่าความคลาดเคลื่อน.....	39
3.2.5 ขั้นตอนสร้างตัวแบบผสม HSVR ด้วยตัวแบบ SVR จากข้อมูล และตัวแบบ SVR จากค่าความคลาดเคลื่อน	39
3.3 เครื่องมือที่ใช้ในการวิจัย.....	40
<b>4 การทดสอบและอภิปรายผล.....</b>	<b>41</b>
4.1 ข้อมูลที่ใช้ในงานวิจัย.....	41
4.2 การทดสอบประสิทธิภาพของเทคนิคซ์พอร์ตเวกเตอร์รีเกรสชันที่เพิ่ม ความแม่นยำด้วยค่าความคลาดเคลื่อน	43
4.3 ผลการทดสอบประสิทธิภาพ.....	43
4.4 อภิปรายผล.....	48
<b>5 สรุปผลการวิจัยและข้อเสนอแนะ.....</b>	<b>49</b>
5.1 สรุปผลการวิจัย.....	49
5.2 ปัญหาและข้อเสนอแนะ.....	50
<b>รายการอ้างอิง.....</b>	<b>51</b>
ภาคผนวก	
ภาคผนวก ก. รหัสต้นฉบับ โปรแกรมภาษา R.....	53
ภาคผนวก ข. บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่.....	67
ประวัติผู้เขียน.....	81

## สารบัญตาราง

ตารางที่	หน้า
2.1	แสดงการพยากรณ์ด้วยค่าเฉลี่ยเคลื่อนที่..... 8
2.2	แสดงการพยากรณ์ด้วยค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก..... 8
2.3	แสดงสมการเคอร์เนลที่นิยมใช้..... 15
2.4	แสดงข้อมูลสมมติที่ใช้ในการวัดค่าความแม่นยำของโมเดล..... 20
2.5	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ RMSE 21
2.6	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MAE 22
2.7	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MSE 23
2.8	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ NMSE 24
2.9	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MAPE 25
2.10	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ R 26
2.11	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ RE 27
2.12	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ SEP 28
2.13	การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ SMAPE 29
2.14	แสดงการเปรียบเทียบเกณฑ์ที่ใช้ในการวัดค่าความแม่นยำของโมเดลจำนวน 2 โมเดล 30
2.15	สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง..... 33
3.1	แสดงชุดข้อมูลตัวอย่างในการสร้างโมเดลเพื่อพยากรณ์..... 37
3.2	แสดงค่าที่ได้จากการพยากรณ์ข้อมูลตัวอย่าง..... 38

## สารบัญตาราง (ต่อ)

ตารางที่	หน้า
3.3 แสดงค่าความคลาดเคลื่อนจากตัวแบบ SVR ที่สร้างได้.....	38
3.4 แสดงค่าที่ได้จากการพยากรณ์ค่าความคลาดเคลื่อน.....	39
3.5 แสดงค่าพยากรณ์ที่ได้จากตัวแบบผสม.....	40
4.1 แสดงจำนวนชุดข้อมูลที่ใช้.....	43
4.2 แสดงการเปรียบเทียบค่าราคาที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลฝึกสอน	44
4.3 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลฝึกสอน.....	44
4.4 แสดงการเปรียบเทียบค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนจากข้อมูลฝึกสอน	44
4.5 แสดงการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลฝึกสอน.....	44
4.6 แสดงการเปรียบเทียบค่าสหสัมพันธ์จากข้อมูลฝึกสอน.....	45
4.7 แสดงการเปรียบเทียบค่าราคาที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลทดสอบ	45
4.8 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลทดสอบ.....	46
4.9 แสดงการเปรียบเทียบค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนจากข้อมูลทดสอบ	46
4.10 แสดงการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ย จากข้อมูลทดสอบ.....	47
4.11 แสดงการเปรียบเทียบค่าสหสัมพันธ์จากข้อมูลทดสอบ.....	48

## สารบัญรูป

รูปที่		หน้า
2.1	แสดงข้อมูลจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2545 - 2555	4
2.2	แสดงข้อมูลที่มีค่าแนวโน้ม (ก) และข้อมูลที่ไม่มีค่าแนวโน้ม (ข).....	5
2.3	แสดงข้อมูลที่มีการเปลี่ยนแปลงตามฤดูกาล.....	6
2.4	แสดงข้อมูลที่เป็นวัฏจักร.....	6
2.5	แสดงข้อมูลที่มีการแปรผันแบบผิปกคิ.....	7
2.6	แสดงการเปรียบเทียบข้อมูลที่มีลักษณะนิ่ง (ก) และข้อมูลที่มีลักษณะไม่นิ่ง (ข).	9
2.7	แสดงลักษณะของท่อเปิดซีลอน.....	12
2.8	แสดงโมเดลโครงข่ายประสาทเทียม.....	16
3.1	กรอบแนวคิดและขั้นตอนการวิจัย.....	35
4.1	แสดงกราฟข้อมูลทั้ง 5 ชุดที่ใช้ในงานวิจัยนี้.....	42
4.2	แสดงกราฟ % Improvement จากเกณฑ์ทดสอบ 5 เกณฑ์.....	48

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของงานวิจัย

การทำนายตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง "การบอกเหตุการณ์หรือความเป็นไปที่จะเกิดในเบื้องหน้า" ซึ่งการทำนายนั้นได้มีมาตั้งแต่สมัยอดีต โดยการทำนายจะถูกใช้โดยผู้ทำพิธี ซึ่งอาจจะเป็นหมอดู พระ นักบวช คนทรง เป็นต้น โดยสิ่งที่ทำนายนั้นมักจะเป็นเรื่องเกี่ยวกับ ชีวิต ความรัก ความเป็นอยู่ ผลลัพธ์การพนันต่าง ๆ เป็นต้น โดยผลลัพธ์ในการทำนายนั้นอาจจะถูกต้อง หรือไม่ถูกต้องก็ได้ เมื่อเวลาผ่านไปการทำนายได้มีการพัฒนาขึ้นโดยอาศัยหลักการต่าง ๆ เข้ามาช่วยเพื่อเพิ่มความแม่นยำในการทำนายเรียกว่าการพยากรณ์ โดยตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง "ทำนายหรือคาดการณ์โดยอาศัยหลักวิชา" ซึ่งหลักวิชาที่ว่านี้คือเทคนิคต่าง ๆ ที่มาช่วยในการพยากรณ์ โดยทางสถิติและคอมพิวเตอร์คือการนำข้อมูลต่าง ๆ ที่มีความเกี่ยวข้องกับสิ่งที่ต้องการพยากรณ์มาช่วยในการหาค่าพยากรณ์ ซึ่งค่าที่นิยมใช้ในการพยากรณ์จะเป็นค่าข้อมูลที่ต้องการพยากรณ์ย้อนหลัง โดยลักษณะของข้อมูลที่ใส่จะอยู่ในรูปแบบของอนุกรมเวลา (Time series) และมีการใช้เทคนิคต่าง ๆ ในการพยากรณ์ โดยเทคนิคที่นิยมใช้คือ เทคนิคสมการเชิงเส้นทั่วไป (Generalized Linear) เทคนิคอาร์มา (ARIMA or Autoregressive Integrated Moving Average) เทคนิคการถดถอย (Regression) เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) เทคนิคเครือข่ายประสาทเทียม (Neural Network) ซึ่งเทคนิคที่ได้กล่าวมานั้นเป็นเทคนิคที่ให้ผลลัพธ์ที่มีความแม่นยำสูง แต่ก็ยังมีความต้องการที่จะให้การพยากรณ์ที่ได้มีความแม่นยำที่สูงขึ้น ซึ่งในปัจจุบันได้มีนักวิจัยพัฒนาเทคนิคใหม่ ๆ เพื่อให้ได้ผลลัพธ์ที่ดีกว่าวิธีเดิม โดยมีการนำเทคนิคหลาย ๆ วิธีมาใช้ร่วมกัน เช่นการนำเทคนิควิธีเชิงพันธุกรรม (Genetic Algorithm) มาช่วยในการหาค่าพารามิเตอร์ในโมเดล (Model) เพื่อให้ได้โมเดลที่ดีที่สุด เป็นต้น

อีกเทคนิคหนึ่งที่น่าสนใจ และสามารถนำมาใช้ช่วยในการเพิ่มความแม่นยำคือการนำค่าความคลาดเคลื่อนที่ได้จากโมเดลมาพิจารณาในการปรับปรุงโมเดล โดยได้มีงานวิจัยที่ใช้เทคนิคอาร์มาในการสร้างโมเดล และวัดค่าความคลาดเคลื่อน ต่อจากนั้นนำเทคนิคเครือข่ายประสาทเทียมมาสร้างโมเดลจากค่าความคลาดเคลื่อนที่ได้จากโมเดลอาร์มา ผลลัพธ์ที่ได้จากทั้งสองขั้นตอนจะถูกนำมาสร้างเป็นโมเดลใหม่ ซึ่งแนวคิดของงานวิจัยที่ใช้เทคนิคแบบผสมนี้คือ โมเดลประกอบด้วย

ส่วนที่เป็นเชิงเส้นและส่วนที่ไม่เป็นเชิงเส้น โดยใช้เทคนิคอาร์มาใช้ในการพยากรณ์ส่วนที่เป็นเชิงเส้น และเทคนิคเครือข่ายประสาทเทียมใช้ในการพยากรณ์ส่วนที่ไม่เป็นเชิงเส้น โดยตั้งชื่อเทคนิคแบบผสมนี้ว่าไฮบริด (Hybrid) ซึ่งเทคนิคนี้ได้มีการพัฒนาต่อโดยการเปลี่ยนแปลงเทคนิคในส่วนที่ไม่เป็นเชิงเส้น เช่น การใช้เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันแทนการใช้เทคนิคเครือข่ายประสาทเทียม หรือ การนำเทคนิควิธีเชิงพันธุกรรมมาช่วยในการเพิ่มความสามารถของตัวแบบไฮบริด เป็นต้น

ที่กล่าวมาข้างต้นพบว่าปัจจุบันได้มีความพยายามในการหาเทคนิคใหม่ ๆ เพื่อนำมาใช้ในการเพิ่มความแม่นยำในการพยากรณ์อนุกรมเวลา โดยในงานวิจัยนี้ผู้วิจัยได้นำเสนอเทคนิคใหม่ โดยนำเทคนิคไฮบริดมาดัดแปลงใหม่ โดยมีสมมติฐานที่ว่าในการสร้างโมเดลฝึกสอนสิ่งที่ได้เมื่อนำมาเทียบกับค่าจริงคือค่าความคลาดเคลื่อน ซึ่งค่าความคลาดเคลื่อนที่ได้นั้นเป็นค่าความสัมพันธ์ที่ขาดหายไปในการพยากรณ์ครั้งแรก ดังนั้นเพื่อที่จะทำให้การพยากรณ์มีความแม่นยำมากขึ้นจึงได้ทำการนำค่าความคลาดเคลื่อนที่ได้มาทำการพยากรณ์อีกครั้งเพื่อเติมความสัมพันธ์ที่ขาดหายไป ซึ่งเทคนิคที่ใช้ในการทดลองคือเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน และได้มีการเปรียบเทียบความแม่นยำกับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม และเทคนิคอาร์มา โดยเกณฑ์ที่ใช้ในการทดสอบความแม่นยำ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error : MSE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error : MAPE) และค่าสหสัมพันธ์ (Correlation coefficient : R)

## 1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาวิธีการเพิ่มความแม่นยำในการพยากรณ์ให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม โดยการนำค่าความคลาดเคลื่อนมาช่วยในการสร้างตัวแบบ
2. เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ของเทคนิคที่นำเสนอกับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม และเทคนิคอาร์มา ด้วยเกณฑ์การเปรียบเทียบ 5 เกณฑ์ได้แก่ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน และค่าสหสัมพันธ์

### 1.3 ขอบเขตของงานวิจัย

1. ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่อยู่ในรูปแบบของอนุกรมเวลา โดยเป็นข้อมูลจาก Data Market (<https://datamarket.com/data/>) จำนวน 3 ชุดข้อมูล และข้อมูลจาก Duke University ([http://www2.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www2.stat.duke.edu/~mw/ts_data_sets.html)) จำนวน 2 ชุดข้อมูล ซึ่งข้อมูลที่นำมาใช้มีทั้งหมด 5 ชุด ประกอบไปด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพีชเชอร์ ข้อมูลปริมาณการผลิตน้ำมันของวัวในแต่ละเดือน ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วิน ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและคาร์วิน

2. การเปรียบเทียบความแม่นยำของเทคนิคใหม่จะทำการเปรียบเทียบกับเทคนิค 2 แบบคือ เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม และเทคนิคอาร์มา

3. การเปรียบเทียบความแม่นยำใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย เนื่องจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นค่าที่นิยมใช้ในการทดสอบความแม่นยำ ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนเป็นค่าที่สามารถนำไปเปรียบเทียบค่าความคลาดเคลื่อนได้ชัดเจนเนื่องจากมีการแปลงค่าให้อยู่ในรูปเปอร์เซ็นต์ ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจะบ่งบอกถึงขนาดของความคลาดเคลื่อนรวมได้ และค่าสหสัมพันธ์สามารถบอกถึงความสัมพันธ์ระหว่างค่าที่พยากรณ์กับค่าจริงได้

### 1.4 ประโยชน์ที่จะได้รับ

จากการศึกษาและพัฒนาเทคนิคที่ใช้ในการพยากรณ์อนุกรมเวลา ผู้วิจัยคาดว่าเทคนิคที่พัฒนาขึ้นจะเกิดประโยชน์ในประเด็นต่อไปนี้

1. เพิ่มความแม่นยำในการพยากรณ์ให้กับการพยากรณ์ด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม
2. เป็นเทคนิคทางเลือกใหม่ในการนำไปใช้พยากรณ์ข้อมูลอนุกรมเวลาอื่น ๆ
3. สามารถนำเทคนิคใหม่นี้ไปต่อยอด โดยผนวกกับเทคนิคพื้นฐานอื่น ๆ เพื่อเพิ่มความแม่นยำการพยากรณ์ได้ในอนาคต

## บทที่ 2

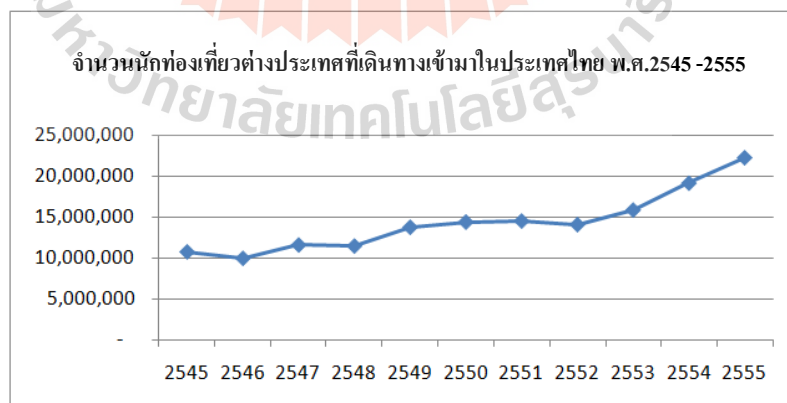
### ปริทัศน์วรรณกรรม

ปริทัศน์วรรณกรรมนี้ประกอบด้วยบททบทวนวรรณกรรมเกี่ยวกับอนุกรมเวลาและงานวิจัยที่เกี่ยวข้อง ซึ่งประกอบด้วยลักษณะข้อมูลอนุกรมเวลา การพยากรณ์อนุกรมเวลาด้วยเทคนิคอาร์มีนา เทคนิคซ์พอร์เตอร์เรกเรชัน และเทคนิคอื่น ๆ เทคนิคที่ใช้ในการวัดความแม่นยำของโมเดล และงานวิจัยที่เกี่ยวข้องกับการพยากรณ์อนุกรมเวลา

#### 2.1 อนุกรมเวลา

##### 2.1.1 ความหมายของอนุกรมเวลา

อนุกรมเวลา (Time Series) หมายถึงค่าข้อมูลของตัวแปรที่สนใจที่มีการบันทึกข้อมูลตามลำดับเวลา ตัวอย่างของชุดข้อมูลที่เป็นอนุกรมเวลา ได้แก่ ข้อมูลจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2545 - 2555 (ข้อมูลจากสำนักงานสถิติแห่งชาติ <http://service.nso.go.th/nso/nsopublish/BaseStat/basestat.html>) ซึ่งแสดงดังรูปที่ 2.1 โดยเส้นในแนวแกนแนวนอนแทนเวลาในหน่วยปีตั้งแต่ปี พ.ศ. 2545 ถึง 2555 และเส้นแกนในแนวตั้งแทนจำนวนนักท่องเที่ยว (คน)



รูปที่ 2.1 แสดงข้อมูลจำนวนนักท่องเที่ยวต่างประเทศที่เดินทางเข้ามาในประเทศไทย พ.ศ. 2545 - 2555

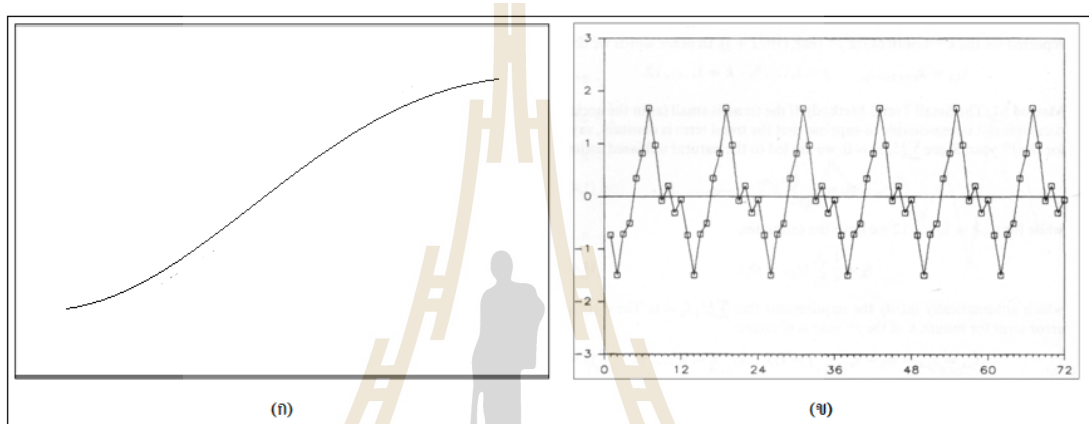


## 2.1.2 ส่วนประกอบของอนุกรมเวลา

อนุกรมเวลาสามารถแบ่งออกได้เป็น 4 ส่วน คือ แนวโน้ม การเปลี่ยนแปลงตามฤดูกาล วัฏจักร และการแปรผันแบบผิดปกติ (มุกดา แม่นมินทร์, 2549, หน้า 2) ซึ่งสามารถอธิบายได้ดังนี้

### 2.1.2.1 ค่าแนวโน้ม (Trend)

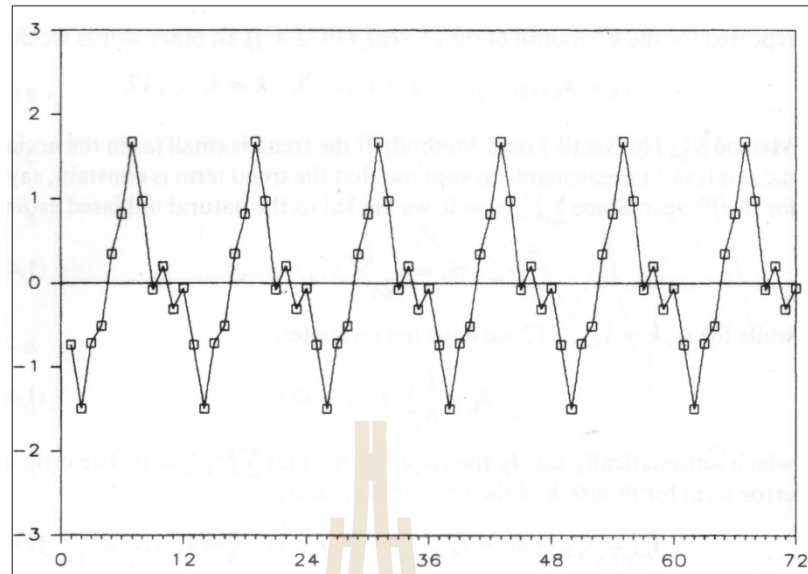
แนวโน้มเป็นการเคลื่อนไหวของอนุกรมเวลา โดยมีการเปลี่ยนแปลงขึ้นหรือลงซ้ำ ๆ ตลอดคาบเวลาที่เราสสนใจ โดยข้อมูลมีลักษณะดังรูปที่ 2.2



รูปที่ 2.2 แสดงการเปรียบเทียบข้อมูลที่มีค่าแนวโน้ม (ก) และข้อมูลที่ไม่มีค่าแนวโน้ม (ข)

### 2.1.2.2 การเปลี่ยนแปลงตามฤดูกาล (Seasonal)

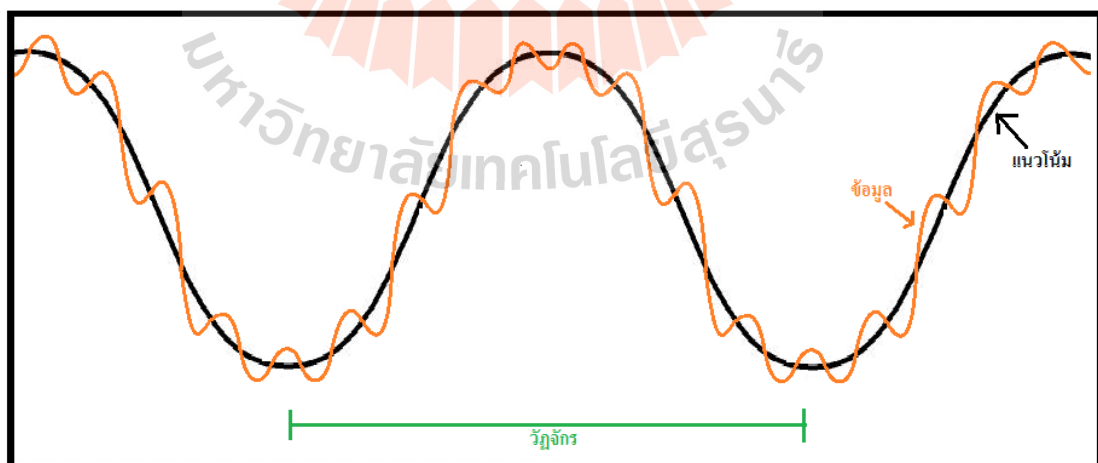
ฤดูกาลเป็นการเปลี่ยนแปลงในคาบเวลาใด ๆ โดยจะมีลักษณะการเปลี่ยนแปลงซ้ำ ๆ ที่คล้ายกัน เช่น ข้อมูลที่เป็นรายปีจะมีลักษณะการเปลี่ยนแปลงตามฤดูกาลในแต่ละเดือนดังตัวอย่างข้อมูลในรูปที่ 2.3 โดยแกนตั้งคือค่าที่สนใจเพื่อใช้ในการพยากรณ์ และแกนนอนคือคาบเวลา จากรูปที่ 2.3 จะเห็นได้ว่าในคาบเวลาที่ 1 - 12 มีลักษณะคล้ายกับคาบเวลาที่ 13 - 24 และมีการเปลี่ยนแปลงแบบซ้ำ ๆ กันในข้อมูลเวลาถัด ๆ ไป



รูปที่ 2.3 แสดงข้อมูลที่มีการเปลี่ยนแปลงตามฤดูกาล (Brockwell and Davis, 2013, page 22)

### 2.1.2.3 วัฏจักร (Cycle)

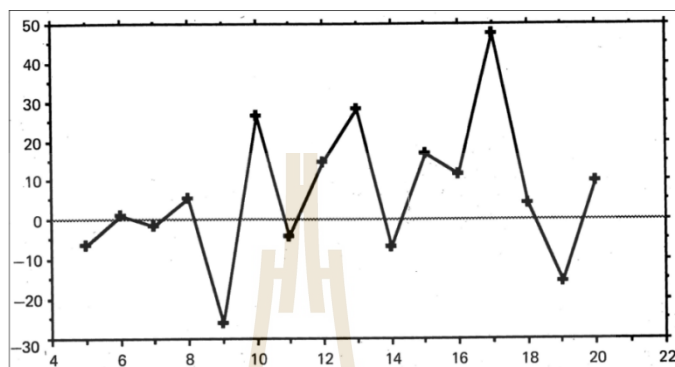
วัฏจักรเป็นการเปลี่ยนแปลงที่มีลักษณะซ้ำ ๆ คล้ายกับการเปลี่ยนแปลงตามฤดูกาล แต่จะมีการเปลี่ยนแปลงอย่างช้า ๆ ซึ่งเกิดจากการเปลี่ยนแปลงของค่าแนวโน้มที่เพิ่มขึ้นหรือลดลง ซึ่งจากรูปที่ 2.4 จะเห็นได้ว่าข้อมูลมีแนวโน้มที่มีรูปแบบซ้ำ ๆ กัน



รูปที่ 2.4 แสดงข้อมูลที่เป็นวัฏจักร

### 2.1.2.4 การแปรผันแบบผิดปกติ (Irregular fluctuations)

การแปรผันแบบผิดปกติเป็นการเคลื่อนไหวที่ผิดปกติของอนุกรมเวลา จากสิ่งอื่นที่ไม่เกี่ยวกับค่าแนวโน้ม การเปลี่ยนแปลงตามฤดูกาล หรือวัฏจักร เช่น แผ่นดินไหว สงคราม การจลาจล เป็นต้น โดยข้อมูลมีลักษณะดังรูปที่ 2.5



รูปที่ 2.5 แสดงข้อมูลที่มีการแปรผันแบบผิดปกติ (Gareth Janacek and Louise Swift, 1993, p.19)

### 2.1.3 การวิเคราะห์หอนุกรมเวลา

การวิเคราะห์หอนุกรมเวลาเป็นการศึกษาลักษณะความเคลื่อนไหวหรือการเปลี่ยนแปลงของข้อมูลที่น่าสนใจในช่วงเวลาใด ๆ เพื่อใช้ในการคาดคะเนแนวโน้มหรือพยากรณ์ค่าของข้อมูลที่จะเกิดขึ้นในอนาคต โดยอาศัยเทคนิคการพยากรณ์แบบต่าง ๆ มาช่วย โดยเรียนรู้จากรูปแบบข้อมูลในอดีต ทั้งนี้สามารถแยกเทคนิคการวิเคราะห์หอนุกรมเวลาได้เป็น 3 ประเภทใหญ่ ๆ คือ

#### 2.1.3.1 การวิเคราะห์ด้วยค่าเฉลี่ยเคลื่อนที่

การวิเคราะห์ด้วยค่าเฉลี่ยเคลื่อนที่ (moving average) เป็นการพยากรณ์ด้วยค่าก่อนหน้า ซึ่งสามารถพยากรณ์ได้ในกรณีที่มีแนวโน้มเพียงเล็กน้อยหรือไม่มีแนวโน้ม มีสมการดังสมการที่ 2.1 และสามารถแสดงการพยากรณ์ได้ดังตารางที่ 2.1 (Hatchett และคณะ, 2010)

$$\text{BaSis}_t(N) = \frac{1}{N} \sum_{i=1}^N \text{Basis}_{t-i} \quad (2.1)$$

เมื่อ  $\text{BaSis}_t(N)$  คือค่าพยากรณ์ค่าเฉลี่ยเคลื่อนที่ ณ เวลา  $t$

$N$  คือจำนวนข้อมูลล้าหลัง

$\text{Basis}$  คือค่าข้อมูลล้าหลัง

ตารางที่ 2.1 แสดงการพยากรณ์ด้วยค่าเฉลี่ยเคลื่อนที่

เดือน	ราคาสินค้า	3 month Moving Average
มกราคม	12	
กุมภาพันธ์	13	
มีนาคม	15	
เมษายน	17	$(15+13+12)/3 = 13.34$
พฤษภาคม	21	$(17+15+13)/3 = 15$
มิถุนายน	24	$(21+17+15)/3 = 17.67$

### 2.1.3.2 การวิเคราะห์ด้วยค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก

การวิเคราะห์ด้วยค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weighted moving average) เป็นการพยากรณ์ด้วยค่าก่อนหน้า โดยนิยมใช้เมื่อข้อมูลมีแนวโน้มมาเกี่ยวข้อง ซึ่งส่งผลให้ยิ่งค่าที่ล่าช้ามากเท่าใดความสำคัญจะยิ่งลดลงไปเท่านั้น โดยค่าถ่วงน้ำหนักเป็นค่าที่ผู้พยากรณ์กำหนดเอง มีสมการดังสมการที่ 2.2 และสามารถแสดงการพยากรณ์ได้ดังตารางที่ 2.2 (Kapgate,2014)

### Weighted Moving Average Forecast

$$= \frac{\sum(\text{weight for period } n) \cdot (\text{demand in period } n)}{\sum \text{weights}} \quad (2.2)$$

เมื่อ **Weighted Moving Average Forecast** คือค่าพยากรณ์ค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก

**n** คือจำนวนข้อมูลล่าช้า

ตารางที่ 2.2 แสดงการพยากรณ์ด้วยค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก

เดือน	ราคาสินค้า	3 month Weighted Moving Average (7-3-1)
มกราคม	12	
กุมภาพันธ์	13	
มีนาคม	15	
เมษายน	17	$((15*7)+(13*3)+(12*1))/11 = 14.18$
พฤษภาคม	21	$((17*7)+(15*3)+(13*1))/11 = 16.09$
มิถุนายน	24	$((21*7)+(17*3)+(15*1))/11 = 19.36$

### 2.1.3.3 การวิเคราะห์ด้วยการปรับเรียบแบบเอ็กซ์โพเนนเชียล

การวิเคราะห์ด้วยการปรับเรียบแบบเอ็กซ์โพเนนเชียลเป็นการพยากรณ์โดยใช้ค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแต่ปรับให้ค่าถ่วงน้ำหนักมีค่าลดลงแบบเอ็กซ์โพเนนเชียล ซึ่งต้องกำหนดค่าคงที่ของการปรับให้เรียบ โดยมีค่าตั้งแต่ 0 ถึง 1 มีสมการดังสมการที่ 2.3 (Hunter, 1986)

$$\hat{y}_{t+1} = \lambda y_t + (1 - \lambda)\hat{y}_t \quad (2.3)$$

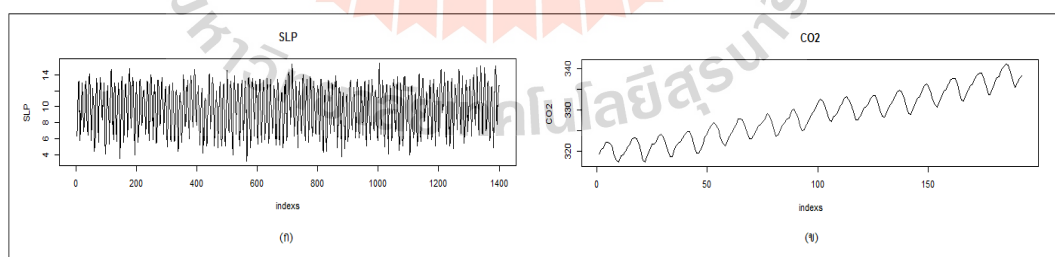
เมื่อ  $\hat{y}_t$  คือค่าพยากรณ์ใหม่ ณ เวลา  $t$

$\lambda$  คือค่าคงที่ของการปรับให้เรียบ

$y_t$  คือค่าจริง ณ เวลา  $t$

### 2.1.4 อนุกรมเวลาที่มีลักษณะนิ่ง (Stationary time series)

อนุกรมเวลาที่มีลักษณะนิ่งคืออนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่ตลอดคาบเวลา ซึ่งเป็นอนุกรมเวลาที่เหมาะสำหรับการนำมาใช้ในการสร้างโมเดลที่ใช้ในการพยากรณ์ เนื่องจากข้อมูลมีความสมดุลและเสถียรภาพ โดยข้อดีของอนุกรมเวลาที่มีลักษณะนิ่งคือเมื่อเกิดการเปลี่ยนแปลงที่ผิดปกติ โมเดลที่สร้างจากข้อมูลอนุกรมเวลาที่มีลักษณะนิ่งจะได้ค่าพยากรณ์ที่เข้าสู่สภาพปกติได้อย่างรวดเร็ว ซึ่งวิธีที่นิยมในการใช้ทดสอบความนิ่งของอนุกรมเวลา คือวิธี Augmented Dickey-Fuller (ADF) ซึ่งถูกเสนอโดย Dickey และ Fuller ในปี 1979 (Dickey and Fuller, 1979) แสดงดังรูปที่ 2.6



รูปที่ 2.6 แสดงการเปรียบเทียบข้อมูลที่มีลักษณะนิ่ง (ก) และข้อมูลที่มีลักษณะไม่นิ่ง (ข)

## 2.2 การวิเคราะห์หอนุกรมเวลาด้วยอาร์มา

### 2.2.1 ประวัติของอาร์มา

อาร์มา (ARIMA) เป็นเทคนิคการวิเคราะห์หอนุกรมเวลาที่เสนอโดยบ็อกซ์และ เจนกินส์ในปี 1976 (Box and Jenkins, 1976) โดยเทคนิคแรกที่ถูกเสนอเรียกว่าอาร์มา (ARMA) โดยจะแบ่งการวิเคราะห์หอนุกรมออกเป็น 2 ส่วนคือ Autoregressive (AR) และ Moving average (MA) โดยข้อมูลที่จะนำมาสร้างโมเดลอาร์มาจำเป็นต้องมีลักษณะนิ่ง ต่อมาจึงได้มีการพัฒนาเทคนิคอาร์มาขึ้นมาโดยพัฒนาให้เทคนิค ARMA สามารถทำงานได้กับข้อมูลที่มีลักษณะไม่นิ่งได้ เทคนิคอาร์มาเป็นเทคนิคที่ได้รับความนิยมมากในการวิเคราะห์หอนุกรมเวลาเนื่องจากให้ผลการพยากรณ์ที่มีความแม่นยำสูง

### 2.2.2 ตัวแบบอาร์มา

#### 2.2.2.1 ส่วน Autoregressive (AR) หรือ AR(p)

รูปแบบ AR เป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้า โดยกำหนดว่าที่ค่าสังเกต  $Y_t$  ที่เวลา  $t$  ใด ๆ จะขึ้นกับค่าสังเกตที่เวลา  $t-1, t-2, t-3, \dots, t-p$  โดยที่ค่า  $p$  คือค่าคาบเวลาที่ล่าหลัง ซึ่งสามารถแสดงเป็นสมการได้ดังสมการที่ 2.4

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (2.4)$$

เมื่อ  $Y_t$  คือค่าสังเกตที่เวลา  $t$  ใด ๆ

$\theta_0$  คือค่าคงที่

$\phi_1, \phi_2, \dots, \phi_p$  คือเซตของพารามิเตอร์ถ่วงน้ำหนัก

$\varepsilon_t$  คือค่าคลาดเคลื่อนที่เวลา  $t$

ถ้าต้องการพยากรณ์ข้อมูลในวันที่ 30 และกำหนดให้ค่า  $p$  เป็น 5 ดังนั้นสมการที่ได้คือ

$$Y_{30} = \theta_0 + \phi_1 Y_{29} + \phi_2 Y_{28} + \phi_3 Y_{27} + \phi_4 Y_{26} + \phi_5 Y_{25} + \varepsilon_{30}$$

#### 2.2.2.2 ส่วน Moving average (MA) หรือ MA(q)

รูปแบบ MA เป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าคลาดเคลื่อนก่อนหน้า โดยกำหนดว่าค่าสังเกต  $Y_t$  ที่เวลา  $t$  ใด ๆ จะขึ้นกับค่าคลาดเคลื่อนที่เวลา  $t-1, t-2, t-3, \dots, t-q$  โดยที่ค่า  $q$  คือค่าคาบเวลาที่ล่าหลัง ซึ่งสามารถแสดงเป็นสมการได้ดังสมการที่ 2.5

$$Y_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.5)$$

เมื่อ  $Y_t$  คือค่าสังเกตที่เวลา  $t$  ใด ๆ  
 $\theta_0$  คือค่าคงที่  
 $\theta_1, \theta_2, \dots, \theta_q$  คือเซตของพารามิเตอร์ถ่วงน้ำหนัก  
 $\varepsilon_t$  คือค่าคลาดเคลื่อนที่เวลา  $t$

ถ้าต้องการพยากรณ์ข้อมูลในวันที่ 30 และกำหนดให้ค่า  $q$  เป็น 5 ดังนั้น

สมการที่ได้คือ

$$Y_{30} = \theta_0 + \varepsilon_{30} - \theta_1 \varepsilon_{29} - \theta_2 \varepsilon_{28} - \theta_3 \varepsilon_{27} - \theta_4 \varepsilon_{26} - \theta_5 \varepsilon_{25}$$

โดยเมื่อทำการรวม AR(p) และ MA(q) จะได้เทคนิคที่เรียกว่า

ARMA(p,q) โดยมีการคำนวณดังสมการที่ 2.6

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.6)$$

ถ้าต้องการพยากรณ์ข้อมูลในวันที่ 30 และกำหนดให้ค่า  $p$  เป็น 5 ค่า  $q$

เป็น 5 ดังนั้นสมการที่ได้คือ

$$Y_{30} = \theta_0 + \phi_1 Y_{29} + \phi_2 Y_{28} + \phi_3 Y_{27} + \phi_4 Y_{26} + \phi_5 Y_{25} + \varepsilon_{30} - \theta_1 \varepsilon_{29} - \theta_2 \varepsilon_{28} - \theta_3 \varepsilon_{27} - \theta_4 \varepsilon_{26} - \theta_5 \varepsilon_{25}$$

### 2.2.3 ตัวแบบอาร์มา

ตัวแบบอาร์มา หรือ ARIMA(p,d,q) เป็นเทคนิคที่พัฒนามาจากอาร์มา โดยทำให้สามารถใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะนิ่งได้โดยทำการแปลงข้อมูลอนุกรมเวลาเดิม  $Y_t$  ซึ่งเป็นข้อมูลที่ไม่นิ่งให้เป็นอนุกรมเวลาค่าใหม่  $Z_t$  ซึ่งเป็นการหาค่าผลต่างระหว่างค่าสังเกตในอนุกรมเวลาเดิม ดังสมการที่ 2.7

$$Z_t = \nabla^d Y_t \quad (2.7)$$

ถ้าให้  $d = 1$  จะได้  $Z_t = \nabla^1 Y_t = Y_t - Y_{t-1}$

ถ้าให้  $d = 2$  จะได้  $Z_t = \nabla^2 Y_t = \nabla^1 Y_t - \nabla^1 Y_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$

## 2.3 การพยากรณ์ด้วยซัพพอร์ตเวกเตอร์รีเกรสชัน

### 2.3.1 ประวัติของซัพพอร์ตเวกเตอร์รีเกรสชัน

ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) หรือ SVR (Smola et al., 2004) เป็นวิธีการหนึ่งที่ใช้พยากรณ์ค่าที่เป็นเลขจำนวนจริง เทคนิค SVR ได้รับความนิยมอย่างมาก และมีการใช้อย่างแพร่หลาย เนื่องจากค่าที่พยากรณ์นั้นมีความแม่นยำสูง โดยวิธีการซัพพอร์ตเวกเตอร์รีเกรสชันนั้นเสนอโดย Vapnik ในปี 1995 เป็นการดัดแปลงมาจากวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยแทนที่จะใช้สมการไฮเปอร์เพลนทำการแบ่งข้อมูลออกเป็น 2 กลุ่มซัพพอร์ตเวกเตอร์รีเกรสชันนั้นจะใช้สมการไฮเปอร์เพลนเป็นสมการที่ใช้ในการพยากรณ์ค่าแทน

### 2.3.2 สมการซัพพอร์ตเวกเตอร์รีเกรสชัน

สิ่งที่ต้องการคือพยากรณ์ค่าจากตัวแปรที่กำหนด โดยการฝึกจะใช้ข้อมูลในอยู่ในรูป  $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$  โดยที่  $X$  คือค่าข้อมูลนำเข้า โดยเป้าหมายที่ต้องการคือหาค่าฟังก์ชัน  $f(x)$  ซึ่งใช้เป็นตัวแทนในการพยากรณ์ซึ่งมีค่าดังสมการที่ 2.8

$$f(x) = \langle w, x \rangle + b \quad \text{โดยที่ } w \in X, b \in \mathbb{R} \quad (2.8)$$

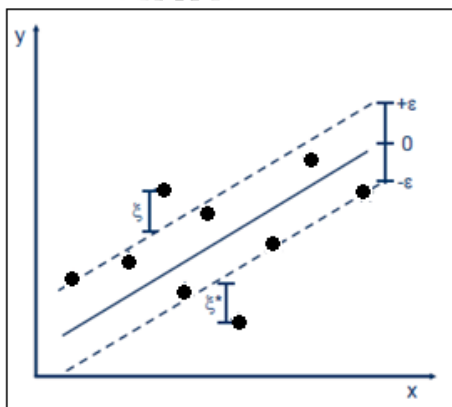
โดยที่  $\langle \cdot, \cdot \rangle$  คือการทำ dot product

$w$  คือค่าน้ำหนักที่ทำการคูณอยู่กับตัวแปร  $x$  ใด ๆ

$b$  คือค่าไบแอสซึ่งเป็นค่าคงที่

### 2.3.3 ท่อเอปซิลอนสำหรับกำหนดซัพพอร์ตเวกเตอร์

ท่อเอปซิลอน (Epsilon tube) เป็นท่อที่ใช้ในการหาค่าที่เป็นซัพพอร์ตเวกเตอร์ ซึ่งเป็นค่าที่สำคัญในการสร้างสมการไฮเปอร์เพลน โดยค่าที่สนใจและให้เป็นซัพพอร์ตเวกเตอร์คือค่าที่อยู่นอกท่อเอปซิลอนดังแสดงในรูปที่ 2.7



รูปที่ 2.7 แสดงลักษณะของท่อเอปซิลอน



โดยที่ค่า  $\xi$  (ไซ) เป็นค่าระยะห่างระหว่างท่อเปิดซีลอนและข้อมูลตัวอย่าง ซึ่งมีค่า ดังสมการที่ 2.9

$$|\xi|_{\varepsilon} = \begin{cases} 0 & \text{ถ้า } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{ในกรณีอื่น} \end{cases} \quad (2.9)$$

ซึ่งค่าสมการไฮเปอร์เพลนที่ดีที่สุดต้องเป็นค่าที่  $w$  น้อยที่สุดดังสมการที่ 2.10

$$\text{ค่าต่ำสุดของ } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2.10)$$

$$\text{โดยที่ } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

โดยที่  $C$  คือค่าคอส (Cost)

### 2.3.4 สมการของลากรานจ์

จากสมการ 2.10 สามารถหาค่า  $w$  ที่มีขนาดเล็กที่สุดโดยใช้สมการของลากรานจ์ (Lagrange function) โดยจะมีการเพิ่มตัวแปรในสมการซึ่งเรียกว่าตัวคูณลากรานจ์ (Lagrange multipliers) ซึ่งให้ค่าดังสมการที่ 2.11

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \end{aligned} \quad (2.11)$$

โดยที่ค่า  $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$  เป็นค่าตัวคูณลากรานจ์

โดยกำหนดให้ค่าตัวคูณลากรานจ์มีค่ามากกว่าศูนย์ดังสมการที่ 2.12

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0 \quad (2.12)$$

โดยที่  $\alpha_i^{(*)}$  เป็นค่าจาก  $\alpha_i$  และ  $\alpha_i^*$

$\eta_i^{(*)}$  เป็นค่าจาก  $\eta_i$  และ  $\eta_i^*$

โดยวิธีการหาค่าที่มีขนาดเล็กที่สุดนั้นทำได้โดยการนำสมการของลากรานจ์มาหาค่าอนุพันธ์เทียบกับตัวแปรที่ต้องการหาค่าซึ่งจากสมการที่ 2.11 จะได้ผลลัพธ์ดังสมการที่ 2.13, 2.14 และ 2.15

$$\partial_b L = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (2.13)$$

$$\partial_w L = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \quad (2.14)$$

$$\partial_{\xi_i^{(*)}} L = C - \alpha_i^* - \eta_i^* = 0 \quad (2.15)$$

โดยจากสมการที่ 2.15 ได้แปลงเป็น  $\eta_i^{(*)} = C - \alpha_i^{(*)}$  จากนั้นนำสมการที่ 2.13, 2.14 และ 2.15 ไปแทนในสมการที่ 2.11 จะได้ผลลัพธ์ดังสมการที่ 2.16

$$\text{ค่าสูงสุด} \begin{cases} \frac{1}{2} \sum_{i=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) < x_i, x_j > \\ -\varepsilon \sum_{i=1}^l (\alpha_i - \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{cases} \quad (2.16)$$

$$\text{โดยที่} \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad \text{และ} \quad \alpha_i, \alpha_i^* \in [0, C]$$

จากสมการที่ 2.14 สามารถแปลงได้สมการ 2.17

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (2.17)$$

เมื่อแทนค่า  $w$  ใน  $F(x)$  จะได้ผลลัพธ์ดังสมการที่ 2.18

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) < x_i, x > + b \quad (2.18)$$

ดังนั้นหากสามารถหาค่าตัวคูณลากรานจ์ได้ก็จะสามารถหาสมการของไฮเปอร์เพลนเพื่อใช้ในการพยากรณ์ได้

### 2.3.5 เคอร์เนลที่ใช้ในซัพพอร์ตเวกเตอร์รีเกรสชัน

การใช้เคอร์เนลเป็นอีกขั้นหนึ่งในการทำซัพพอร์ตเวกเตอร์รีเกรสชัน โดยสามารถทำให้แก้ปัญหาที่ไม่เป็นเส้นตรงได้ ซึ่งทำให้ค่าพยากรณ์ที่ได้แม่นยำยิ่งขึ้น โดยการทำให้อยู่ในมิติที่สูงขึ้น โดยทำการแปลง  $x_i$  เป็น  $\phi$  ยกตัวอย่างเช่น ต้องการแปลง  $(x_1, x_2)$  จาก 2 มิติไปเป็น 3 มิติ ซึ่งจะอยู่ในรูป  $\phi(x_1, x_2)$  โดยกำหนดให้  $\phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$  เป็นต้น โดยสมการเคอร์เนลที่นิยมใช้แสดงดังตารางที่ 2.3 ซึ่งเคอร์เนลที่นิยมใช้ในการสร้างโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันเพื่อพยากรณ์อนุกรมเวลาคือ Radial Basis Function

ตารางที่ 2.3 แสดงสมการเคอร์เนลที่นิยมใช้

Kernel	$K(x, x_i)$
Radial Basis Function	$\exp(-\gamma \ x - x_i\ ^2), \gamma > 0$
Inverse multiquadratic	$\frac{1}{\sqrt{\ x - x_i\  + \eta}}$
Polynomial of degree d	$((x^T \cdot x_i) + \eta)^d$
Sigmoidal	$\tanh(\gamma(x^T \cdot x_i) + \eta), \gamma > 0$
Linear	$x^T \cdot x_i$

## 2.4 เทคนิคอื่นที่ใช้ช่วยในการพยากรณ์อนุกรมเวลา

### 2.4.1 เทคนิคไฮบริด

เทคนิคไฮบริด (hybrid) หรือวิธีการผสมผสาน (Zhang, 2003) เป็นแนวคิดใหม่เพื่อเพิ่มความแม่นยำในการพยากรณ์ โดยมีแนวคิดที่ว่าข้อมูลอนุกรมเวลานั้นมีส่วนประกอบของข้อมูลที่มีลักษณะเป็นเชิงเส้นและไม่เป็นเชิงเส้นผสมกันอยู่ โดยมีสมการดังสมการที่ 2.19

$$y_t = L_t + N_t \quad (2.19)$$

โดยที่  $y_t$  คือค่าสังเกตที่เวลา  $t$

$L_t$  คือส่วนประกอบที่มีลักษณะเป็นเชิงเส้นที่เวลา  $t$

$N_t$  คือส่วนประกอบที่มีลักษณะไม่เป็นเชิงเส้นที่เวลา  $t$

ในขั้นตอนแรกได้ทำการสร้างโมเดลโดยใช้เทคนิคอาร์มาเนื่องจากเทคนิคอาร์มาสามารถวิเคราะห์ได้ดีในข้อมูลที่มีลักษณะเป็นเชิงเส้น โดยมีสมการดังสมการที่ 2.20

$$e_t = y_t - \hat{L}_t \quad (2.20)$$

โดยที่  $e_t$  คือค่าคลาดเคลื่อนที่เวลา  $t$

$y_t$  คือค่าสังเกตที่เวลา  $t$

$\hat{L}_t$  คือค่าที่ได้จากการพยากรณ์โดยใช้เทคนิคอาร์มาที่เวลา  $t$

ในขั้นตอนที่ 2 ได้นำเทคนิคเครือข่ายประสาทเทียมมาใช้ในการทำนายค่าความคลาดเคลื่อนเนื่องจากเทคนิคเครือข่ายประสาทเทียมสามารถวิเคราะห์ได้ดีในข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น โดยแสดงดังสมการที่ 2.21 และ 2.22

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \quad (2.21)$$

$$e_t = \hat{N}_t + \varepsilon_t \quad (2.22)$$

โดยที่  $f$  คือฟังก์ชันของค่าที่ไม่เป็นเชิงเส้นที่พยากรณ์โดยเทคนิคเครือข่ายประสาทเทียมที่เวลา  $t$

$\varepsilon_t$  คือค่าความคลาดเคลื่อนสุ่ม

$\hat{N}_t$  คือค่าที่ได้จากการพยากรณ์โดยใช้เทคนิคเครือข่ายประสาทเทียมที่เวลา  $t$

ได้สมการการพยากรณ์ดังสมการที่ 2.23

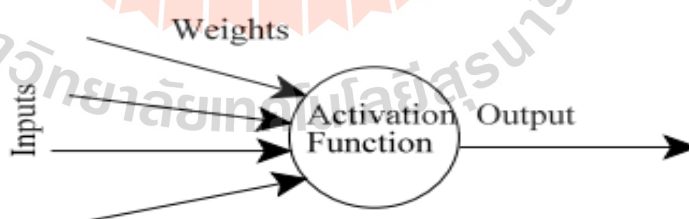
$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (2.23)$$

จากสมการที่ 2.23 ค่า  $\hat{y}_t$  คือค่าพยากรณ์ที่ได้จากการพิจารณาข้อมูลที่มีลักษณะเป็นเชิงเส้น  $\hat{L}_t$  และข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น  $\hat{N}_t$  ซึ่งจากการทดลองพบว่าสามารถเพิ่มความแม่นยำได้

## 2.4.2 โครงข่ายประสาทเทียม

### 2.4.2.1 ประวัติ

โครงข่ายประสาทเทียม (Artificial neural network : ANN) เป็นโครงข่ายที่ใช้ในการประมวลผลข้อมูล โดยเลียนแบบการทำงานของระบบโครงข่ายประสาทของสิ่งมีชีวิตโดยมีลักษณะดังรูปที่ 2.8



รูปที่ 2.8 แสดงโมเดลโครงข่ายประสาทเทียม (Gershenson, 2003)

จากรูปที่ 2.8 ส่วนประกอบหลักของโครงข่ายประสาทเทียมคือ ข้อมูลเข้า (Input :  $p$ ) ค่าน้ำหนัก (Weights :  $w$ ) ฟังก์ชันถ่ายโอน (Transfer Function or Activation Function) และข้อมูลออก (Output :  $y$ ) โดยมีหลักการทำงานคือ รับข้อมูลนำเข้าจากนั้นนำมาคูณกับค่าน้ำหนักแล้วนำไปผ่านฟังก์ชันถ่ายโอน ซึ่งผลลัพธ์ที่ได้คือข้อมูลออก โดยมีสมการดังสมการที่ 2.24

$$y = f(wp) \quad (2.24)$$

แต่โดยปกติแล้วจะมีค่าคงที่มาบวกเพิ่มให้กับฟังก์ชัน โดยเรียกค่านี้ว่า ไบแอส (Bias : b) ดังนั้นสมการที่ได้จึงเป็นดังสมการที่ 2.25

$$y = f(wp) + b \quad (2.25)$$

## 2.4.2.2 ประเภทของโครงข่ายประสาทเทียม

### 2.4.2.2.1 การเรียนรู้แบบมีผู้ฝึกสอน

เป็นการเรียนรู้ซึ่งมีการตรวจสอบผลลัพธ์กับคำตอบเพื่อให้โครงข่ายทำการปรับตัว โดยหากผลลัพธ์ที่ได้มาไม่ตรงกับคำตอบ โครงข่ายประสาทเทียมจะทำการปรับตัวเพื่อให้ผลลัพธ์ที่ได้มีค่าดีขึ้น ตัวอย่างเช่น ต้องการ โครงข่ายประสาทเทียมที่สามารถแยกแยะชนิดต่าง ๆ จากการสำรวจดาวอังคาร โดยข้อมูลที่ใช้ในการพยากรณ์ประกอบด้วย สี ความแข็ง จุดเยือกแข็ง จุดหลอมเหลว เป็นต้น โดยถ้าเป็นการเรียนรู้แบบมีผู้ฝึกสอนในการฝึกสอนจะมีผลลัพธ์ซึ่งเป็นแรงแชชนิดต่าง ๆ ซึ่งแต่ละชนิดจะมีลักษณะที่แตกต่างกัน โดยมีการแบ่งกลุ่มตามผลลัพธ์ที่ให้ไป จากนั้นเมื่อฝึกสอนเสร็จก็จะสามารถนำไปใช้แยกชนิดของแร่ได้ ซึ่งข้อดีของการเรียนรู้แบบมีผู้ฝึกสอนคือมีความแม่นยำสูงเนื่องจากผ่านการฝึกสอนมาแล้ว

### 2.4.2.2.2 การเรียนรู้แบบไม่มีผู้ฝึกสอน

เป็นการเรียนรู้โดยที่ไม่มีคำตอบให้เปรียบเทียบ โดยโครงข่ายจะทำการจัดเรียงโครงสร้างตามผลลัพธ์ที่ได้ ซึ่งโครงข่ายจะทำการจัดหมวดหมู่ของผลลัพธ์ด้วยตัวเองได้ ตัวอย่างเช่น ต้องการ โครงข่ายประสาทเทียมที่สามารถแยกแยะชนิดต่าง ๆ จากการสำรวจดาวอังคารเช่นเดียวกับหัวข้อการเรียนรู้แบบมีผู้ฝึกสอน แต่แตกต่างตรงที่ไม่มีผลลัพธ์ให้เปรียบเทียบ ซึ่งโครงข่ายประสาทเทียมจะทำการแยกแยะและจัดกลุ่มตามความเหมาะสมด้วยตัวมันเอง เมื่อมีค่าใหม่เข้ามาระบบก็จะทำการปรับปรุงโครงข่ายไปเรื่อย ๆ ซึ่งข้อดีของการเรียนรู้แบบไม่มีผู้ฝึกสอนคือสามารถแยกข้อมูลที่ไม่เคยพบมาก่อนได้ เนื่องจากในบางกรณีเราไม่สามารถกำหนดผลลัพธ์ที่แน่นอนก่อนได้ เช่นอาจมีแร่ชนิดใหม่ที่ไม่เคยค้นพบมาก่อนในการสำรวจ

## 2.4.3 ขั้นตอนวิธีเชิงพันธุกรรม

ขั้นตอนวิธีเชิงพันธุกรรมเป็นเทคนิคทางปัญญาประดิษฐ์อย่างหนึ่งที่ใช้ในการเพิ่มความแม่นยำให้กับโมเดล โดยการใช้ทฤษฎีการวิวัฒนาการทางธรรมชาติ เพื่อช่วยในการหาคำตอบที่เหมาะสมที่สุด โดยมีวิธีการดังนี้

### 2.4.3.1 แปลงให้อยู่ในรูปแบบโครโมโซม

ปัญหาที่ต้องการหาคำตอบจะต้องถูกแปลงให้อยู่ในรูปแบบโครโมโซม (Chromosome Encoding) โดยรูปแบบที่ธรรมดาที่สุดคือเป็นเลขฐานสอง เช่น Chromosome A = 0111000110110110 เป็นต้น

### 2.4.3.2 การสืบพันธุ์

การสืบพันธุ์ (Reproduction) เป็นการสร้างกลุ่มประชากรใหม่โดยการคัดเลือกจากประชากรเดิมโดยการประเมินด้วยฟังก์ชันความเหมาะสม (Fitness Function) โดยเป็นแนวคิดมาจากการคัดเลือกสายพันธุ์ ซึ่งสายพันธุ์ที่มีความเหมาะสมมากกว่าจะมีโอกาสรอดและสืบพันธุ์มากกว่า โดยวิธีการที่ใช้ในการคัดเลือกลักษณะนั้นมียุทธวิธีหลายวิธี เช่น Roulette Wheel, Tournament, Ranking เป็นต้น

### 2.4.3.3 การผสมยีน

การผสมยีน (Crossover) เป็นการนำโครโมโซมของประชากรที่ได้มาจากการสืบพันธุ์มาจับคู่กันเพื่อทำการผสมยีนเพื่อให้ได้ประชากรใหม่ขึ้นมา โดยวิธีการที่ง่ายที่สุดคือการสุ่มตำแหน่งการผสมยีนจากนั้นนำโครโมโซมที่อยู่หน้าตำแหน่งการผสมยีนจากยีนพ่อและโครโมโซมที่อยู่หลังตำแหน่งการผสมยีนจากยีนแม่ ซึ่งผลลัพธ์ที่ได้คือยีนลูกคนแรก จากนั้นทำการสลับโดยการให้โครโมโซมที่อยู่หน้าตำแหน่งการผสมยีนจากยีนแม่และโครโมโซมที่อยู่หลังตำแหน่งการผสมยีนจากยีนพ่อจะได้เป็นลูกคนที่สอง

### 2.4.3.4 การกลายพันธุ์

การกลายพันธุ์ (Mutation) เป็นเป็นกระบวนการต่อจากการผสมยีนโดยการนำยีนที่ได้มาทำการเปลี่ยนแปลงเพียงเล็กน้อยเช่นการเปลี่ยนค่าเพียงตัวเดียวจาก 0 เป็น 1 หรือจาก 1 เป็น 0 โดยการสุ่ม โดยสิ่งที่ได้จากการกลายพันธุ์คือทำให้ได้ลักษณะใหม่ ๆ ที่ไม่พบในรุ่นพ่อแม่ ซึ่งอาจมีประโยชน์ในการหาค่าที่เหมาะสม

### 2.4.3.5 การทำงานของขั้นตอนวิธีเชิงพันธุกรรม

ขั้นตอนวิธีเชิงพันธุกรรมเป็นเป็นวิธีที่ใช้หาคำตอบที่เหมาะสมโดยการใช้วิธีการที่ได้กล่าวมาข้างต้นและหากยังไม่ได้ค่าที่เหมาะสมก็จะทำการวนซ้ำใหม่จนกว่าจะได้คำตอบที่เหมาะสม หรือจนกว่าจะครบรอบที่กำหนด

## 2.4.4 การแปลงเวฟเล็ตแบบไม่ต่อเนื่อง

การแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform : DWT) เป็นวิธีการหนึ่งในการแบ่งสัญญาณออกเป็นสัญญาณความถี่สูงและสัญญาณความถี่ต่ำ โดยการนำสัญญาณไปผ่านตัวกรองความถี่สูงและความถี่ต่ำ โดยหลักการที่ว่าสัญญาณทุกสัญญาณใน

ธรรมชาติเกิดจากการรวมกันของสัญญาณย่อย ๆ หลากหลายความถี่ จากนั้นทำการรวมสัญญาณ โดยผลลัพธ์ที่ได้จะทำให้สัญญาณรบกวนที่มีลดลงโดยที่ยังคงรูปแบบของสัญญาณเดิมไว้ได้

#### 2.4.5 ขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค

ขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization : PSO) (Eberhart and Kennedy, 1995) เป็นขั้นตอนวิธีการหาค่าที่เหมาะสมที่สุดของปัญหา ซึ่งวิธีการนี้ได้แรงบันดาลใจมาจากการสังเกตการเคลื่อนที่ของฝูงนกในการออกหาอาหาร โดยเหล่านกจะมีการส่งสัญญาณสื่อสารกันเพื่อบอกถึงแหล่งที่มีอาหารและทำการเคลื่อนที่ไปยังแหล่งอาหารนั้น จากแนวคิดดังกล่าววิธีขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคจึงได้ถูกสร้างขึ้นมา โดยมีการใช้ออนุภาค (Particle) จำนวนมากเคลื่อนที่ทำการค้นหาผลลัพธ์ของปัญหาในพื้นที่ค้นหา (Search Space) ซึ่งตัวอนุภาคนี้ก็เปรียบได้เหมือนกับนกและผลลัพธ์ก็คือแหล่งอาหาร โดยอนุภาคแต่ละตัวจะถูกสุ่มตำแหน่งเริ่มต้นในการค้นหา จากนั้นจะทำการค้นหาผลลัพธ์ โดยผลลัพธ์ของแต่ละอนุภาคหาเจอจะเรียกว่าตำแหน่งที่ดีที่สุดส่วนบุคคล (Personal Best Position) ซึ่งเมื่อแต่ละอนุภาคทำการค้นหาเสร็จสิ้นจะมีการเทียบตำแหน่งที่ดีที่สุดส่วนบุคคลของแต่ละอนุภาคและเลือกคำตอบที่ดีที่สุดเรียกว่าตำแหน่งที่ดีที่สุดสากล (Global Best Position) และนำค่านี้มาใช้งาน ซึ่งวิธีนี้จะช่วยแก้ปัญหาในส่วนของคำตอบที่เป็นเฉพาะถิ่น (Local) ได้



## 2.5 เกณฑ์ที่ใช้ในการวัดความแม่นยำของโมเดล

เกณฑ์ที่ใช้ในการวัดค่าความแม่นยำของโมเดลมีด้วยกันหลากหลายวิธี โดยในการอธิบายวิธีคำนวณค่าความคลาดเคลื่อนที่จะบ่งชี้ถึงความแม่นยำของโมเดลจะใช้ข้อมูลสมมติ 10 รายการที่กำหนดค่าจริงของข้อมูลและค่าที่ได้จากการพยากรณ์ของโมเดลตามตารางที่ 2.4

ตารางที่ 2.4 แสดงข้อมูลสมมติที่ใช้ในการวัดค่าความแม่นยำของโมเดล

ค่าจริง	ค่าพยากรณ์
2	2.2
3	2.9
5	5.1
6	6.2
7	6.8
9	9
10	10.1
11	10.9
13	12.5
15	15.3

### 2.5.1 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) เป็นวิธีที่การวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการดังสมการที่ 2.26 ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่า RMSE ที่ได้มีค่าน้อย แสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (2.26)$$

โดยที่ RMSE คือค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$n$  คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา  $t$  ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า RMSE ได้ดังตารางที่ 2.5



ตารางที่ 2.5 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ RMSE

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$(Y_t - \hat{Y}_t)^2$
2	2.2	0.04
3	2.9	0.01
5	5.1	0.01
6	6.2	0.04
7	6.8	0.04
9	9	0
10	10.1	0.01
11	10.9	0.01
13	12.5	0.25
15	15.3	0.09

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{10}(0.04 + 0.01 + 0.01 + 0.04 + 0.04 + 0 + 0.01 + 0.01 + 0.25 + 0.09)} \\
 &= \sqrt{\frac{1}{10}(0.5)} \\
 &= \sqrt{(0.05)} \\
 &= 0.2236
 \end{aligned}$$

### 2.5.2 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Deviation : MAD) หรือ (Mean Absolute Error : MAE) เป็นวิธีที่การวัดค่าความคลาดเคลื่อนที่นิยมอีกวิธีหนึ่ง ซึ่งวิธีนี้จะช่วยบอกถึงขนาดของความคลาดเคลื่อนรวมได้ โดยมีสมการดังสมการที่ 2.27 ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (2.27)$$

โดยที่ MAE คือค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

n คือจำนวนข้อมูลที่ให้

$Y_t$  คือค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า MAE ได้ดังตารางที่ 2.6

ตารางที่ 2.6 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MAE

ค่าจริง ( $Y_t$ )	ค่าพยากรณ์ ( $\hat{Y}_t$ )	$ Y_t - \hat{Y}_t $
2	2.2	0.2
3	2.9	0.1
5	5.1	0.1
6	6.2	0.2
7	6.8	0.2
9	9	0
10	10.1	0.1
11	10.9	0.1
13	12.5	0.5
15	15.3	0.3
$MAE = \frac{1}{10}(0.2 + 0.1 + 0.1 + 0.2 + 0.2 + 0 + 0.1 + 0.1 + 0.5 + 0.3)$ $= \frac{1}{10}(1.8)$ $= 0.18$		

### 2.5.3 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error : MSE) เป็นการวัดค่าความคลาดเคลื่อน โดยการนำค่าความคลาดเคลื่อนมายกกำลังแล้วนำไปหาค่าเฉลี่ย ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมากโดยมีสมการดังสมการที่ 2.28

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (2.28)$$

โดยที่ MSE คือค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

n คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า MSE ได้ดังตารางที่ 2.7

ตารางที่ 2.7 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MSE

ค่าจริง ( $Y_t$ )	ค่าพยากรณ์ ( $\hat{Y}_t$ )	$(Y_t - \hat{Y}_t)^2$
2	2.2	0.04
3	2.9	0.01
5	5.1	0.01
6	6.2	0.04
7	6.8	0.04
9	9	0
10	10.1	0.01
11	10.9	0.01
13	12.5	0.25
15	15.3	0.09

$$MSE = \frac{1}{10} (0.04 + 0.01 + 0.01 + 0.04 + 0.04 + 0 + 0.01 + 0.01 + 0.25 + 0.09)$$

$$= \frac{1}{10} (0.5)$$

$$= 0.05$$

#### 2.5.4 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน (Normalized Mean Square Error : NMSE) เป็นค่าที่ตัดแปลงมาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก โดยมีสมการดังสมการที่ 2.29

$$NMSE = \frac{1}{(\delta^2 n)} * \sum_{i=1}^n (a_i - p_i)^2 \quad (2.29)$$

$$\delta^2 = \frac{1}{(n - 1)} * \sum_{i=1}^n (a_i - \bar{a})^2$$

โดยที่ NMSE คือค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน

- $n$  คือจำนวนข้อมูลที่ใช้  
 $a_i$  คือค่าจริงที่เวลา  $t$  ใด ๆ  
 $\bar{a}$  คือค่าเฉลี่ยของค่าจริง  
 $p_i$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า NMSE ได้ดังตารางที่ 2.8

ตารางที่ 2.8 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ NMSE

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$(a_i - \bar{a})^2$	$(a_i - p_i)^2$
2	2.2	37.21	0.04
3	2.9	26.01	0.01
5	5.1	9.61	0.01
6	6.2	4.41	0.04
7	6.8	1.21	0.04
9	9	0.81	0
10	10.1	3.61	0.01
11	10.9	8.41	0.01
13	12.5	24.01	0.25
15	15.3	47.61	0.09

$$\begin{aligned}
 \delta^2 &= \frac{1}{(9)} * (37.21 + 26.01 + 9.61 + 4.41 + 1.28 + 0.81 + 3.61 + 8.41 + 24.01 + 47.61) \\
 &= \frac{1}{(9)} * 162.9 \\
 &= 18.1 \\
 \text{NMSE} &= \frac{1}{(18.1^2 * 10)} * (0.5) \\
 &= 0.0001526
 \end{aligned}$$

### 2.5.5 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน

ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error : MAPE) เป็นค่าความคลาดเคลื่อนที่แปลงให้อยู่ในรูปแบบของเปอร์เซ็นต์ ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก ซึ่งยิ่ง โดยมีสมการดังสมการที่ 2.30

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 * \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (2.30)$$

โดยที่ MAPE คือค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน

n คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า MAPE ได้ดังตารางที่ 2.9

ตารางที่ 2.9 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ MAPE

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$\left  100 * \frac{Y_t - \hat{Y}_t}{Y_t} \right $
2	2.2	10
3	2.9	3.33
5	5.1	2
6	6.2	3.33
7	6.8	2.85
9	9	0
10	10.1	1
11	10.9	0.91
13	12.5	3.85
15	15.3	2
$MAPE = \frac{1}{10} (10 + 3.33 + 2 + 3.33 + 2.85 + 0 + 1 + 0.91 + 3.85 + 2)$ $= \frac{1}{10} (29.28)$ $= 2.928$		

### 2.5.6 ค่าสหสัมพันธ์

ค่าสหสัมพันธ์ (Correlation Coefficient : R) เป็นค่าที่ใช้ในการบอกความสัมพันธ์ระหว่างค่าสองค่า ในที่นี้คือค่าจริงและค่าจากการพยากรณ์ ในการวัดค่าความแม่นยำ จากวิธีการนี้ยังค่าที่ได้มีค่ามากแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก โดยมีค่าอยู่ระหว่าง 1 และ -1 ซึ่งมีสมการดังสมการที่ 2.31

$$R = \frac{\sum_{i=1}^n (a_i * p_i)}{\sqrt{\sum_{i=1}^n a_i^2} * \sqrt{\sum_{i=1}^n p_i^2}} \quad (2.31)$$

โดยที่ R คือค่าสหสัมพันธ์  
 n คือจำนวนข้อมูลที่ใช่  
 $a_i$  คือค่าจริงที่เวลา t ใด ๆ  
 $p_i$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า R ได้ดังตารางที่ 2.10

ตารางที่ 2.10 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ R

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$a_i^2$	$p_i^2$	$a_i * p_i$
2	2.2	4	4.84	4.4
3	2.9	9	8.41	8.7
5	5.1	25	26.01	25.5
6	6.2	36	38.44	37.2
7	6.8	49	46.24	47.6
9	9	81	81	81
10	10.1	100	102.01	101
11	10.9	121	118.81	119.9
13	12.5	169	156.25	162.5
15	15.3	225	234.09	229.5
$R = \frac{4.4 + 8.7 + 25.5 + 37.2 + 47.6 + 81 + 101 + 119.9 + 162.5 + 229.5}{\sqrt{819} * \sqrt{816.1}}$ $= \frac{817.3}{28.62 * 28.58}$ $= 0.9992$				

### 2.5.7 ค่าคลาดเคลื่อนสัมพัทธ์

ค่าคลาดเคลื่อนสัมพัทธ์ (Relative Error : RE) เป็นวิธีหาค่าความคลาดเคลื่อนพื้นฐาน โดยการเทียบค่าความคลาดเคลื่อนกับค่าจริงในแต่ละค่าที่ทำการพยากรณ์และแปลงให้อยู่ในรูปเปอร์เซ็นต์ ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก โดยมีสมการดังสมการที่ 2.32

$$RE = \frac{|\hat{Y}_t - Y_t|}{Y_t} * 100 \quad (2.32)$$

โดยที่ RE คือค่าคลาดเคลื่อนสัมพัทธ์

$Y_t$  คือค่าจริงที่เวลา t ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า RE ได้ดังตารางที่ 2.11

ตารางที่ 2.11 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ RE

ค่าจริง ( $Y_t$ )	ค่าพยากรณ์ ( $\hat{Y}_t$ )	$\frac{ \hat{Y}_t - Y_t }{Y_t} * 100$
2	2.2	10
3	2.9	3.33
5	5.1	2
6	6.2	3.33
7	6.8	2.85
9	9	0
10	10.1	1
11	10.9	0.91
13	12.5	3.85
15	15.3	2

### 2.5.8 ค่าความผิดพลาดมาตรฐานของการทำนาย

ค่าความผิดพลาดมาตรฐานของการทำนาย (Standard Error of Prediction : SEP) เป็นค่าที่ใช้ในการทดสอบค่าความผิดพลาด ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าโมเดลที่ได้จะมีความแม่นยำมาก โดยมีสมการดังสมการที่ 2.33

$$SEP = \frac{100}{\bar{X}} \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_t - Y_t)^2}{n}} \quad (2.33)$$

โดยที่ SEP คือค่าความผิดพลาดมาตรฐานของการทำนาย  
 n คือจำนวนข้อมูลที่ใช้  
 $Y_t$  คือค่าจริงที่เวลา t ใด ๆ  
 $\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา t ใด ๆ  
 $\bar{X}$  คือค่าเฉลี่ยค่าจริงที่เวลา t ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า SEP ได้ดังตารางที่ 2.12

ตารางที่ 2.12 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ SEP

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$(\hat{Y}_t - Y_t)^2$
2	2.2	0.04
3	2.9	0.01
5	5.1	0.01
6	6.2	0.04
7	6.8	0.04
9	9	0
10	10.1	0.01
11	10.9	0.01
13	12.5	0.25
15	15.3	0.09
$SEP = \frac{100}{8.1} * 0.2236$ $= 2.7605$		



### 2.5.9 ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนสมมาตร

ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนสมมาตร (Symmetric Mean Absolute Percentage Error : SMAPE) เป็นมาตรวัดอีกชนิดหนึ่งที่น่าสนใจ โดยมีสมการดังสมการที่ 2.34

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{(|Y_t| + |\hat{Y}_t|)/2} \quad (2.34)$$

โดยที่ SMAPE คือค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนสมมาตร

$n$  คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา  $t$  ใด ๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใด ๆ

จากข้อมูลสมมติตามตารางที่ 2.4 จะคำนวณค่า SMAPE ได้ดังตารางที่ 2.13

ตารางที่ 2.13 การคำนวณความคลาดเคลื่อนของการพยากรณ์ข้อมูล 10 รายการด้วยเกณฑ์ SMAPE

ค่าจริง( $Y_t$ )	ค่าพยากรณ์( $\hat{Y}_t$ )	$ Y_t - \hat{Y}_t $	$( Y_t  +  \hat{Y}_t )/2$	$\frac{ Y_t - \hat{Y}_t }{( Y_t  +  \hat{Y}_t )/2}$
2	2.2	0.2	2.1	0.04
3	2.9	0.1	2.95	0.01
5	5.1	0.1	5.05	0.01
6	6.2	0.2	6.1	0.04
7	6.8	0.2	6.9	0.04
9	9	0	9	0
10	10.1	0.1	10.05	0.01
11	10.9	0.1	10.95	0.01
13	12.5	0.5	12.75	0.25
15	15.3	0.3	15.15	0.09
$SMAPE = \frac{1}{10} \left( \frac{0.2}{2.1} + \frac{0.1}{2.95} + \frac{0.1}{5.05} + \frac{0.2}{6.1} + \frac{0.2}{6.9} + \frac{0}{9} + \frac{0.1}{10.05} + \frac{0.1}{10.95} + \frac{0.5}{12.75} + \frac{0.3}{15.15} \right)$ $= \frac{1}{10} (0.2888)$ $= 0.02888$				

จากเกณฑ์ที่ใช้ในการวัดค่าความแม่นยำทั้งหมดจากชุดข้อมูลสมมติจำนวน 10 รายการ สามารถสรุปเปรียบเทียบความแม่นยำในการพยากรณ์ของโมเดลจำนวน 2 โมเดลและแสดงผลการเปรียบเทียบได้ดังตารางที่ 2.14

จากตารางที่ 2.14 ได้ทำการเปลี่ยนแปลงรายการข้อมูลที่ 1 จากโมเดล A ค่า 2.2 เป็น โมเดล B ค่า 2.5 ซึ่งทำให้โมเดล B พยากรณ์มีความผิดพลาดมากกว่า จากการใช้เกณฑ์ทดสอบแบบต่าง ๆ พบว่า RMSE MAE MSE NMSE MAPE SEP SMAPE โมเดล B ให้ค่าที่สูงกว่าโมเดล A และ R โมเดล B ให้ค่าที่ต่ำกว่าโมเดล A ซึ่งผลลัพธ์ที่ได้สอดคล้องกับข้อมูล

ตารางที่ 2.14 แสดงการเปรียบเทียบเกณฑ์ที่ใช้ในการวัดค่าความแม่นยำของโมเดลจำนวน 2 โมเดล

รายการข้อมูลที่	ค่าจริง	ค่าพยากรณ์จากโมเดล A	ค่าพยากรณ์จากโมเดล B
1	2	2.2	2.5
2	3	2.9	2.9
3	5	5.1	5.1
4	6	6.2	6.2
5	7	6.8	6.8
6	9	9	9
7	10	10.1	10.1
8	11	10.9	10.9
9	13	12.5	12.5
10	15	15.3	15.3
	RMSE	0.2236	0.2265
	MAE	0.18	0.21
	MSE	0.05	0.071
	NMSE	0.002762	0.003927
	MAPE	2.928	4.428
	R	0.9997	0.9995
	SEP	2.7605	3.2896
	SMAPE	0.02888	0.04158

## 2.6 งานวิจัยที่เกี่ยวข้อง

ในหัวข้อนี้จะอธิบายถึงงานวิจัยอื่น ๆ ที่เกี่ยวข้องกับการพยากรณ์อนุกรมเวลาด้วยค่าความคลาดเคลื่อน โดยมีนักวิจัยได้ทำการนำเสนอเทคนิควิธีการพยากรณ์แบบใหม่ เพื่อเพิ่มค่าความแม่นยำให้กับวิธีการเดิม โดยผู้วิจัยได้ศึกษางานเหล่านี้และสามารถสรุปได้ดังนี้

Zhang (2003) ได้เสนอเทคนิคใหม่โดยรวมเทคนิคอาร์มาและเครือข่ายประสาทเทียม โดยมีแนวคิดที่ว่าข้อมูลอนุกรมเวลานั้นมีส่วนประกอบของข้อมูลที่มีลักษณะเป็นเชิงเส้นและไม่เป็นเชิงเส้นผสมกันอยู่ โดยกำหนดให้เทคนิคแรกที่ใช้คือเทคนิคอาร์มาเนื่องจากเทคนิคอาร์มาสามารถวิเคราะห์ได้ดีในข้อมูลที่มีลักษณะเป็นเชิงเส้น จากนั้นนำค่าความคลาดเคลื่อนที่ได้จากการพยากรณ์ของอาร์มาไปใช้ในการสร้างโมเดลด้วยเทคนิคเครือข่ายประสาทเทียมเนื่องจากเป็นเทคนิคที่สามารถวิเคราะห์ได้ดีในข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น จากนั้นนำทั้ง 2 โมเดลมาใช้ในการพยากรณ์โดยเรียกวิธีการใหม่นี้ว่าเทคนิคไฮบริด และวัดประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

Chen และ Wang (2007) เสนอเทคนิคที่ดัดแปลงมาจากเทคนิคไฮบริดของ Zhang (2003) ซึ่งเป็นการรวมเทคนิคของอาร์มา และซัพพอร์ตเวกเตอร์รีเกรสชัน โดยกำหนดให้ตัวแบบแรกใช้เทคนิคอาร์มาจากนั้นนำค่าความคลาดเคลื่อนจากโมเดลอาร์มานำมาใช้ในโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ และได้มีการใช้เจเนติกอัลกอริทึมในการช่วยหาค่าพารามิเตอร์ของซัพพอร์ตเวกเตอร์รีเกรสชัน โดยเรียกเทคนิคที่เสนอใหม่ว่า GA-SVM และวัดประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน และค่าสหสัมพันธ์

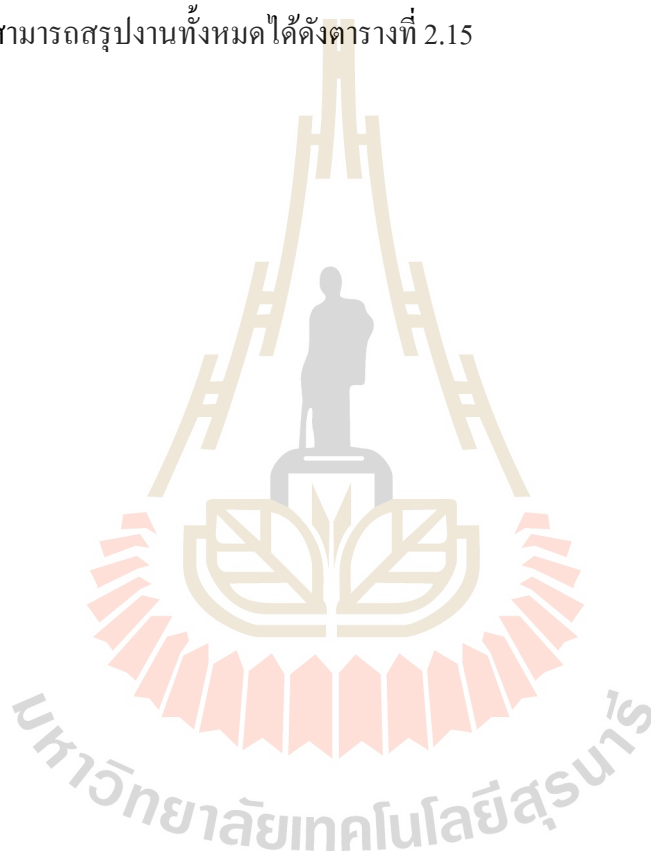
Wang และคณะ (2012) เสนอเทคนิคที่ดัดแปลงมาจากเทคนิคไฮบริดของ Zhang (2003) โดยเริ่มต้นการพยากรณ์โดยใช้เทคนิคอาร์มาแล้วนำค่าความคลาดเคลื่อนที่ได้มาสร้างตัวแบบอาร์มาอีกครั้ง จากนั้นนำตัวแบบทั้งสองไปใช้ในการพยากรณ์ และวัดประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนสัมพัทธ์ ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าความผิดพลาดมาตรฐานของการทำนาย

Khandelwal และคณะ (2015) ได้นำวิธีของ Zhang (2003) มาใช้ในการพยากรณ์ร่วมกับวิธี Discrete Wavelet Transform (DWT) โดยการใช้ DWT ในการแบ่งข้อมูลออกเป็นข้อมูลความถี่สูง และความถี่ต่ำ จากนั้นนำวิธีการไฮบริดมาใช้ในการพยากรณ์ และวัดประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน

de Oliveira และ Ludermir (2016) ได้เสนอวิธีการนำขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาคมาช่วยในการหาผลลัพธ์ที่ดีที่สุด และนำเทคนิคไฮบริดของ Zhang (2003) มาใช้

ในการพยากรณ์และวัดประสิทธิภาพโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนสมมาตร

แนวคิดของวิทยานิพนธ์นี้ได้้นำแนวคิดของ Zhang (2003) มาดัดแปลงเช่นเดียวกับงานวิจัยอื่น ๆ ที่กล่าวข้างต้น แต่มีข้อแตกต่างโดยการเปลี่ยนแนวคิดที่ว่าข้อมูลประกอบด้วยข้อมูลที่มีลักษณะเป็นเชิงเส้นรวมกับข้อมูลที่มีลักษณะไม่เป็นเชิงเส้นไปเป็นความสัมพันธ์ที่ขาดหายไป โดยการนำค่าความคลาดเคลื่อนที่ได้จากสมมติฐานที่คาดว่าจะจะเป็นความสัมพันธ์ที่ขาดหายไปจากโมเดลดั้งเดิมนำไปสร้าง โมเดลด้วยเทคนิคเดิมอีกครั้งเพื่อเติมความสัมพันธ์ที่ขาดหายไปให้กับโมเดลแรก ซึ่งสามารถสรุปงานทั้งหมดได้ดังตารางที่ 2.15



ตารางที่ 2.15 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง

กระบวนการทำงาน	งานวิจัยที่เกี่ยวข้อง					
	ก	ข	ค	ง	จ	ฉ*
เทคนิคที่ใช้ในการพยากรณ์อนุกรมเวลา						
ARIMA	X			X	X	
SARIMA		X	X			
ANN	X			X	X	
SVR		X			X	X
GA		X				
DWT				X		
PSO					X	
เกณฑ์วัดค่าความคลาดเคลื่อนของการพยากรณ์						
MSE (Mean Squared Error)	X			X	X	X
MAE (Mean Absolute Error)	X					X
NMSE (Normalized Mean Square Error)		X				
MAPE (Mean Absolute Percentage Error)		X	X	X		X
R (Correlation Coefficient)		X				X
RE (Relative Error)			X			
RMSE (Root Mean Squared Error)			X			X
SEP (Standard Error of Prediction)			X			
SMAPE (Symmetric Mean Absolute Percentage Error)					X	

ก หมายถึง งานวิจัยของ Zhang (2003)

ข หมายถึง งานวิจัยของ Chen และ Wang (2007)

ค หมายถึง งานวิจัยของ Wang และคณะ (2012)

ง หมายถึง งานวิจัยของ Khandelwal และคณะ (2015)

จ หมายถึง งานวิจัยของ de Oliveira และ Ludermir (2016)

ฉ หมายถึง งานวิจัยของวิทยานิพนธ์ฉบับนี้

## บทที่ 3

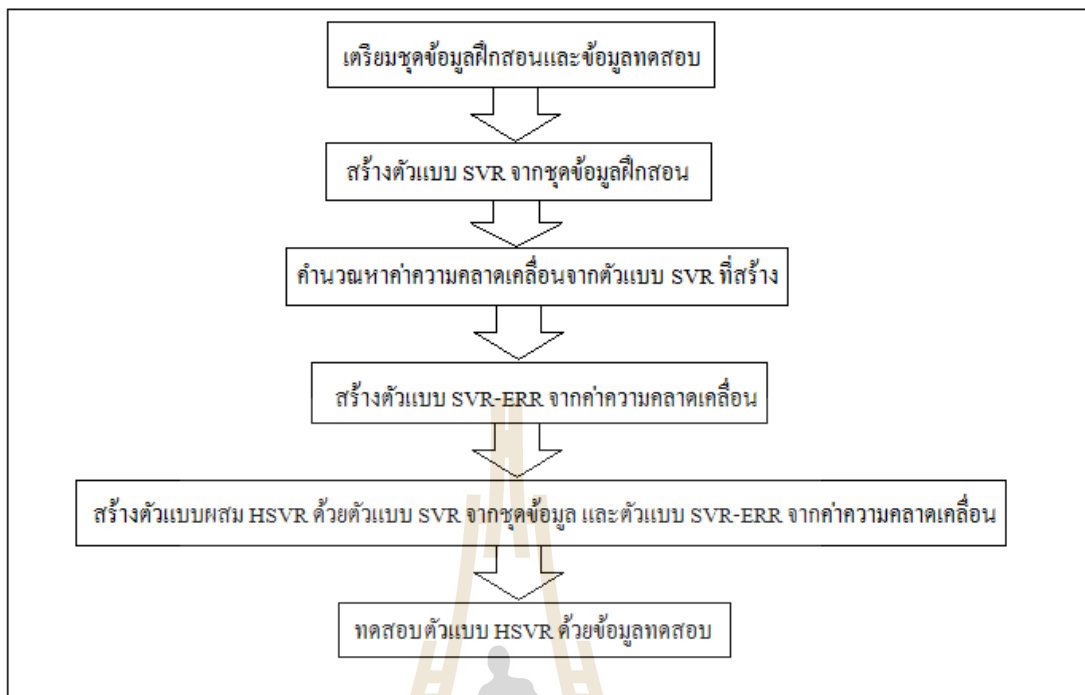
### วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาเทคนิคที่ใช้เพิ่มความแม่นยำให้การพยากรณ์อนุกรมเวลา โดยใช้ค่าความคลาดเคลื่อนช่วยในการเพิ่มความแม่นยำ ซึ่งเทคนิคที่นำมาใช้ในการพยากรณ์คือเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน โดยในบทนี้จะกล่าวถึงกรอบแนวคิดของการวิจัย เครื่องมือที่ใช้ในการวิจัย กระบวนการต่าง ๆ ในการวิจัย ซึ่งมีรายละเอียดดังนี้

#### 3.1 กรอบแนวคิดของการวิจัย

แนวคิดของงานวิจัยนี้เกิดจากความต้องการที่จะเพิ่มความแม่นยำให้การพยากรณ์อนุกรมเวลา โดยการใช้นิเทศซัพพอร์ตเวกเตอร์รีเกรสชัน และเพิ่มความแม่นยำโดยใช้ค่าความคลาดเคลื่อนจากการพยากรณ์ครั้งแรกมาเป็นข้อมูลให้กับการสร้างโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันในครั้งใหม่ เพื่อเป็นตัวแทนสำหรับการประมาณการความบกพร่องของการทำนายและชดเชยความบกพร่องนี้ในขั้นสุดท้ายของตัวแบบผสม ซึ่งแนวคิดในการวิจัยมีขั้นตอนดังนี้

1. นำชุดข้อมูลอนุกรมเวลามาใช้ในการสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน
  2. นำตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันที่ได้มาพยากรณ์ข้อมูลหาค่าความคลาดเคลื่อน โดยเทียบกับชุดข้อมูลอนุกรมเวลาที่ใช้ในการสร้างตัวแบบ
  3. นำค่าความคลาดเคลื่อนที่ได้มาสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันสำหรับทำหน้าที่ประมาณการความบกพร่องของการทำนายในครั้งแรก
  4. สร้างตัวแบบผสมโดยใช้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันจากชุดข้อมูล และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันจากค่าความคลาดเคลื่อน
- ซึ่งสามารถแสดงกรอบแนวคิดได้ดังรูปที่ 3.1



รูปที่ 3.1 กรอบแนวคิดและขั้นตอนการวิจัย

รูปที่ 3.1 ได้แสดงรายละเอียดของกรอบแนวคิดและขั้นตอนในการวิจัย ซึ่งสามารถอธิบายรายละเอียดในขั้นตอนต่าง ๆ ได้ดังต่อไปนี้

#### ขั้นตอนที่ 1 การเตรียมข้อมูลโดยแบ่งเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ

ผู้วิจัยได้ทำการแบ่งชุดข้อมูลอนุกรมเวลาออกเป็น 2 ส่วน ในส่วนแรกเป็นข้อมูลฝึกสอน โดยใช้อัตราส่วน 70% ของข้อมูลทั้งหมด และส่วนที่สองเป็นข้อมูลทดสอบ โดยใช้อัตราส่วน 30% ของข้อมูลทั้งหมด เช่นข้อมูลที่มีทั้งหมด 100 ข้อมูล ข้อมูลที่ 1 - 70 จะใช้เป็นข้อมูลฝึกสอนและ 71 - 100 จะใช้เป็นข้อมูลทดสอบ โดยข้อมูลฝึกสอนถูกนำมาใช้ในขั้นตอนที่ 2 และ 3 ส่วนข้อมูลทดสอบถูกนำมาใช้ในขั้นตอนที่ 4

#### ขั้นตอนที่ 2 การสร้างโมเดลเบื้องต้นด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน

ในการสร้างโมเดลหรือตัวแบบ SVR เพื่อการพยากรณ์อนุกรมเวลาด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันนั้นจำเป็นต้องกำหนดสมการที่ใช้ เนื่องจากข้อมูลมีลักษณะเป็นอนุกรมเวลาดังนั้นตัวแปรที่ใช้ช่วยในการพยากรณ์คือค่า วัน เดือน ปี โดยกำหนดการพยากรณ์ดังสมการที่ 3.1

$$Y_t = f(x) = \langle w, x \rangle + b \quad (3.1)$$

โดยที่ตัวแปร  $Y_t$  คือค่าที่ได้จากการพยากรณ์ด้วยโมเดล SVR

$x$  คือข้อมูลนำเข้าซึ่งแต่ละชุดข้อมูลประกอบด้วยข้อมูล วัน เดือน ปี และค่าเป้าหมายของการพยากรณ์( $Y$ )

$b$  คือค่าไบแอสของโมเดล SVR

ในงานวิจัยนี้ผู้วิจัยได้กำหนดค่าให้ใช้เคอร์เนลของซัพพอร์ตเวกเตอร์รีเกรสชันเป็นเรเดียลเบสิสฟังก์ชัน ใช้พารามิเตอร์ค่าแกมมาเป็นสัดส่วนของ  $1/\text{จำนวนมิติของข้อมูล}$  ค่าคอสเป็น 1 และค่าเอปซิลอนเป็น 0.1 เพื่อใช้ในการเรียนรู้สำหรับสร้างโมเดลพยากรณ์ โดยใช้ข้อมูลฝึกสอนในการสร้างโมเดล SVR และเมื่อประเมินความแม่นยำของโมเดลด้วยการนำค่าที่ได้จากการพยากรณ์และค่าจริงจากข้อมูลฝึกสอนมาทำการหาผลต่าง สิ่งที่ได้ก็คือค่าความคลาดเคลื่อน ซึ่งหากค่าจริง ( $Y$ ) มากกว่าค่าพยากรณ์ ( $Y_t$ ) จะทำให้ค่าความคลาดเคลื่อนมีค่าเป็นบวก แต่หากค่าจริงน้อยกว่าค่าพยากรณ์ค่าความคลาดเคลื่อนจะมีค่าเป็นลบ การคำนวณค่าความคลาดเคลื่อนแสดงได้ดังสมการที่ 3.2

$$\text{error} = Y - Y_t \quad (3.2)$$

**ขั้นตอนที่ 3 การนำค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำของการพยากรณ์**

การนำค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำนั้นทำได้โดยการนำค่าความคลาดเคลื่อนมาสร้างโมเดลพยากรณ์ความบกพร่องของการทำนายโดยค่าความคลาดเคลื่อนที่ได้ยังคงมีลักษณะเป็นอนุกรมเวลา โดยมีสมการการพยากรณ์ของโมเดล SVR-ERR แสดงได้ดังสมการที่ 3.3

$$Y_{\text{error}} = f(x) = \langle w, x \rangle + b \quad (3.3)$$

โดยที่ตัวแปร  $Y_{\text{error}}$  คือค่าที่ได้จากการพยากรณ์ความคลาดเคลื่อนด้วยโมเดล SVR-ERR  
 $x$  คือข้อมูลนำเข้าซึ่งแต่ละชุดข้อมูลประกอบด้วยข้อมูล วัน เดือน ปี และค่าความคลาดเคลื่อน (error)

$b$  คือค่าไบแอสของโมเดล SVR-ERR

โดยค่าอื่น ๆ ของพารามิเตอร์ที่ใช้ในการสร้างโมเดลเป็นค่าเดียวกับที่ใช้สร้างโมเดลด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันในขั้นตอนที่ 2 โดยผลลัพธ์ที่ได้นั้นจะเป็นโมเดลที่ใช้คาดหมายค่าความคลาดเคลื่อน ซึ่งจากสมมติฐานที่ว่าค่าความคลาดเคลื่อนคือค่าความสัมพันธ์ที่ขาดหายไป ดังนั้นเพื่อเพิ่มความแม่นยำให้โมเดลที่สร้างด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิมจึงได้นำโมเดลที่ได้จากค่าความคลาดเคลื่อนมาช่วยในการพยากรณ์ ซึ่งโมเดลใหม่ที่ได้แสดงดังสมการที่ 3.4 โดยเรียกโมเดลนี้ว่าโมเดลผสมหรือ HSVR model



$$Y_p = Y_t + Y_{\text{error}} \quad (3.4)$$

โดยที่ตัวแปร  $Y_p$  คือค่าที่ได้จากการพยากรณ์ด้วยโมเดลผสมหรือ HSVR model  
 $Y_t$  คือค่าที่ได้จากการพยากรณ์ด้วยโมเดล SVR  
 $Y_{\text{error}}$  คือค่าความคลาดเคลื่อนที่พยากรณ์โดยโมเดล SVR-ERR ใช้ช่วยเพิ่มความแม่นยำให้กับโมเดล HSVR

#### ขั้นตอนที่ 4 การทดสอบความแม่นยำของโมเดล

ในการทดสอบความแม่นยำของโมเดลนั้นข้อมูลที่ใช้คือข้อมูลทดสอบที่เตรียมไว้ตั้งแต่ขั้นตอนแรก โดยใช้โมเดลผสม HSVR ในการพยากรณ์ค่า ทดสอบความแม่นยำใช้การวัดค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (MAPE) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าสหสัมพันธ์ (R) โดยทำการเปรียบเทียบเทคนิคทั้งหมด 3 ชนิดคือ เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน (HSVR) และเทคนิคอาร์มา

### 3.2 ตัวอย่างการสร้างโมเดลเพื่อการพยากรณ์

#### 3.2.1 ชุดข้อมูลอนุกรมเวลา

ตารางที่ 3.1 เป็นตารางแสดงชุดข้อมูลตัวอย่าง ซึ่งประกอบไปด้วย id คือลำดับของข้อมูล month คือเดือนที่ผลิตนม year คือปีที่ผลิตนม milk คือปริมาณน้ำนมและค่านี้เป็นค่าที่ต้องการพยากรณ์

ตารางที่ 3.1 แสดงชุดข้อมูลตัวอย่างในการสร้างโมเดลเพื่อพยากรณ์

id	month	year	milk
1	1	1962	589
2	2	1962	561
3	3	1962	640
4	4	1962	656
5	5	1962	727
6	6	1962	697
7	7	1962	640
8	8	1962	599
9	9	1962	568
10	10	1962	577

### 3.2.2 ขั้นตอนสร้างตัวแบบ SVR จากชุดข้อมูล

จากข้อมูลในตารางที่ 3.1 เมื่อนำมาสร้างเป็นตัวแบบ SVR ผลลัพธ์ของการพยากรณ์ด้วยตัวแบบ SVR แสดงดังตารางที่ 3.2 โดย forecast (milk) คือค่าที่พยากรณ์ได้ และ milk คือค่าจริงที่ต้องการพยากรณ์

ตารางที่ 3.2 แสดงค่าที่ได้จากการพยากรณ์ข้อมูลตัวอย่าง

id	month	year	milk	forecast (milk)
1	1	1962	589	618.029
2	2	1962	561	618.635
3	3	1962	640	620.5
4	4	1962	656	621.701
5	5	1962	727	621.949
6	6	1962	697	621.701
7	7	1962	640	620.489
8	8	1962	599	618.5
9	9	1962	568	617.422
10	10	1962	577	617.758

### 3.2.3 ขั้นตอนคำนวณหาค่าความคลาดเคลื่อนของตัวแบบ SVR

สามารถหาค่าความคลาดเคลื่อนของการพยากรณ์จากตัวแบบ SVR ที่สร้างได้โดยการนำค่าจริงลบด้วยค่าที่ได้จากการพยากรณ์ผลลัพธ์ที่ได้คือ error แสดงในตารางที่ 3.3

ตารางที่ 3.3 แสดงค่าความคลาดเคลื่อนจากตัวแบบ SVR ที่สร้างได้

id	month	year	milk	forecast (milk)	error
1	1	1962	589	618.029	-29.0288
2	2	1962	561	618.635	-57.6353
3	3	1962	640	620.500	19.5000
4	4	1962	656	621.701	34.2987
5	5	1962	727	621.949	105.0506
6	6	1962	697	621.701	75.2991
7	7	1962	640	620.489	19.5111
8	8	1962	599	618.500	-19.5000
9	9	1962	568	617.422	-49.4223
10	10	1962	577	617.758	-40.7581

### 3.2.4 ขั้นตอนสร้างตัวแบบ SVR-ERR จากค่าความคลาดเคลื่อน

จากค่าความคลาดเคลื่อนที่คำนวณได้ดังตารางที่ 3.3 สามารถนำค่าคลาดเคลื่อนมาสร้างตัวแบบ SVR-ERR เพื่อประมาณการความบกพร่องหรือ error ของโมเดล SVR ผลของการพยากรณ์ความบกพร่องโมเดล SVR แสดงด้วยค่าในคอลัมน์ forecast (error) ของตารางที่ 3.4

ตารางที่ 3.4 แสดงค่าที่ได้จากการพยากรณ์ค่าความคลาดเคลื่อน

id	month	year	milk	forecast (milk)	error	forecast(error)
1	1	1962	589	618.029	-29.0288	-1.4712
2	2	1962	561	618.635	-57.6353	-0.8647
3	3	1962	640	620.5	19.5	1.0000
4	4	1962	656	621.701	34.2987	2.2013
5	5	1962	727	621.949	105.0506	2.4494
6	6	1962	697	621.701	75.2991	2.2009
7	7	1962	640	620.489	19.5111	0.9889
8	8	1962	599	618.5	-19.5	-1.0000
9	9	1962	568	617.422	-49.4223	-2.0777
10	10	1962	577	617.758	-40.7581	-1.7419

### 3.2.5 ขั้นตอนสร้างตัวแบบผสม HSVR ด้วยตัวแบบ SVR จากข้อมูล และตัวแบบ SVR-ERR จากค่าความคลาดเคลื่อน

สามารถสร้างตัวแบบผสม HSVR ได้โดยการนำค่าที่ได้จากการนำตัวแบบที่ได้จากข้อมูลตัวอย่างร่วมกับตัวแบบจากค่าความคลาดเคลื่อน ทำให้ได้ผลลัพธ์ดังตารางที่ 3.5 โดยจะเห็นว่าค่าที่ได้จากการพยากรณ์ในคอลัมน์ proposed method มีค่าใกล้เคียงค่าจริงในคอลัมน์ milk มากขึ้น

ตารางที่ 3.5 แสดงค่าพยากรณ์ที่ได้จากตัวแบบผสม

id	month	year	milk	forecast (milk)	error	forecast (error)	proposed method
1	1	1962	589	618.029	-29.0288	-1.4712	616.5578
2	2	1962	561	618.635	-57.6353	-0.8647	617.7703
3	3	1962	640	620.5	19.5	1.0000	621.5
4	4	1962	656	621.701	34.2987	2.2013	623.9023
5	5	1962	727	621.949	105.0506	2.4494	624.3984
6	6	1962	697	621.701	75.2991	2.2009	623.9019
7	7	1962	640	620.489	19.5111	0.9889	621.4779
8	8	1962	599	618.5	-19.5	-1.0000	617.5
9	9	1962	568	617.422	-49.4223	-2.0777	615.3443
10	10	1962	577	617.758	-40.7581	-1.7419	616.0161

### 3.3 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการพัฒนางานวิจัยนี้ ประกอบด้วย

1) เครื่องคอมพิวเตอร์สำหรับการพัฒนา มีรายละเอียดดังนี้

หน่วยประมวลผลกลาง : AMD A10

หน่วยความจำหลัก : 8 GB

หน่วยความจำสำรอง : 1 TB

2) ระบบปฏิบัติการและโปรแกรมประยุกต์สำหรับการพัฒนา ประกอบไปด้วย

ระบบปฏิบัติการ : Windows 8.1 Pro 64 bits

เครื่องมือที่ใช้ในการพัฒนา : R Language, RStudio

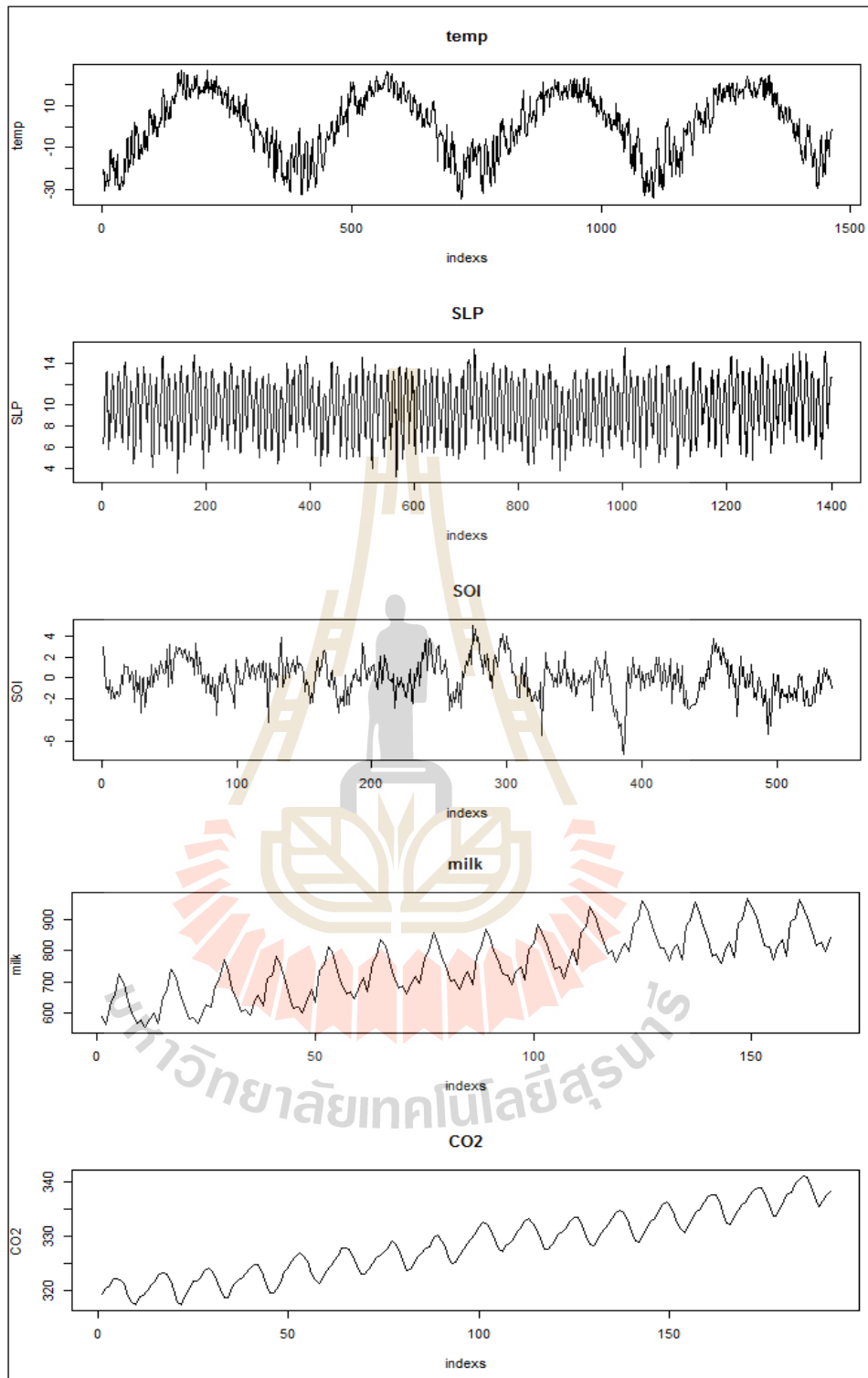
## บทที่ 4

### การทดสอบและอภิปรายผล

การทดสอบประสิทธิภาพของเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน จะทดสอบโดยใช้เกณฑ์ 5 เกณฑ์ประกอบด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error : MSE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error : MAPE) และค่าสหสัมพันธ์ (Correlation coefficient : R) โดยการเปรียบเทียบประสิทธิภาพของ HSVR จะเปรียบเทียบกับอีก 2 เทคนิค คือ เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน และเทคนิคอาร์มา โดยเนื้อหาในบทนี้ประกอบด้วยข้อมูลที่ใช้ในงานวิจัย การทดสอบประสิทธิภาพของเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน (HSVR) ผลการทดสอบและอภิปรายผล

#### 4.1 ข้อมูลที่ใช้ในงานวิจัย

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่อยู่ในรูปแบบของอนุกรมเวลา โดยเป็นข้อมูลจาก Data Market (<https://datamarket.com/data/>) จำนวน 3 ชุดข้อมูล และข้อมูลจาก Duke University ([http://www2.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www2.stat.duke.edu/~mw/ts_data_sets.html)) จำนวน 2 ชุดข้อมูล รวมข้อมูลที่นำมาใช้มีทั้งหมด 5 ชุด ประกอบไปด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพีชเชอร์ (temp) ข้อมูลปริมาณการผลิตนมของวัวในแต่ละเดือน (milk) ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วิน (SLP) ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา (CO<sub>2</sub>) และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีติและคาร์วิน (SOI) ลักษณะข้อมูลทั้งหมดแสดงได้ดังรูปที่ 4.1 ในการสร้างโมเดลและทดสอบโมเดลใช้การแบ่งแต่ละชุดข้อมูลออกเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ ซึ่งแบ่งข้อมูลในอัตราส่วนร้อยละ 70 และ 30 ตามลำดับ จำนวนข้อมูลในแต่ละชุด แสดงดังตารางที่ 4.1



รูปที่ 4.1 แสดงกราฟข้อมูลทั้ง 5 ชุดที่ใช้ในงานวิจัยนี้

ตารางที่ 4.1 แสดงจำนวนชุดข้อมูลที่ใช้

ชุดข้อมูล	จำนวนข้อมูลฝึกสอน	จำนวนข้อมูลทดสอบ	ข้อมูลทั้งหมด
temp	1022	439	1461
milk	118	50	168
SLP	979	421	1400
CO2	136	56	192
SOI	378	162	540

#### 4.2 การทดสอบประสิทธิภาพของเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน

วิธีการทดสอบประสิทธิภาพของเทคนิค HSVR ใช้วิธีการทดสอบด้วยชุดข้อมูลที่แยกไว้เป็นการเฉพาะ โดยข้อมูลที่ใช่แบ่งเป็นข้อมูลฝึกสอน 70% และข้อมูลทดสอบ 30% หลังจากสร้างโมเดลด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันจากข้อมูลฝึกสอนแล้ว จากนั้นทำการหาความคลาดเคลื่อนโดยการเปรียบเทียบค่าที่ได้จากการพยากรณ์และค่าจริงของข้อมูล แล้วทำการสร้างโมเดล HSVR โดยใช้ค่าความคลาดเคลื่อนประกอบการทำนาย จากนั้นนำโมเดล HSVR ไปพยากรณ์ข้อมูลทดสอบ ผลการพยากรณ์ของ HSVR จะเปรียบเทียบกับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันและเทคนิคอาร์มา โดยใช้เกณฑ์เปรียบเทียบ 5 เกณฑ์ ได้แก่ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสหสัมพันธ์

#### 4.3 ผลการทดสอบประสิทธิภาพ

จากการทดสอบโมเดล HSVR เปรียบเทียบกับโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันและโมเดลอาร์มาได้ผลลัพธ์ดังนี้ จากตารางที่ 4.2 ถึง 4.6 พบว่าเมื่อนำค่าความคลาดเคลื่อนมาช่วยในการทำนายทำให้โมเดลเข้าใจค่าของข้อมูลฝึกสอนมากขึ้น โดยทุกข้อมูลให้ผลลัพธ์ที่ดีขึ้นทั้งหมด จากตารางที่ 4.3 ชุดข้อมูลอุณหภูมิรายวันของแม่น้ำฟิชเซอร์ (temp) และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาสีติและดาร์วิน (SOI) ไม่สามารถทดสอบประสิทธิภาพด้วยค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนได้เนื่องจากข้อมูลมีค่าเข้าใกล้ศูนย์

ตารางที่ 4.2 แสดงการเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลฝึกสอน

data	SVR	Proposed method	Improvement	% Improvement
temp	5.15720	5.11310	0.04410	0.855115179
milk	22.4174	21.4962	0.92120	4.109307948
SLP	1.0115	1.0099	0.00160	0.158180919
co2	0.4751	0.434	0.04110	8.650810356
SOI	1.5845	1.5772	0.00730	0.460713159

ตารางที่ 4.3 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลฝึกสอน

data	SVR	Proposed method	Improvement	% Improvement
temp	3.96800	3.90040	0.06760	1.703629032
milk	16.9748	15.1543	1.82050	10.72472135
SLP	0.7916	0.7864	0.00520	0.656897423
co2	0.3925	0.3262	0.06630	16.89171975
SOI	1.2274	1.2123	0.01510	1.23024279

ตารางที่ 4.4 แสดงการเปรียบเทียบค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนจากข้อมูลฝึกสอน

data	SVR	Proposed method	Improvement	% Improvement
milk	2.39944	2.14672	0.25272	10.53246
SLP	9.38009	9.33268	0.04741	0.505432
co2	0.12036	0.1001	0.02026	16.83283

ตารางที่ 4.5 แสดงการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลฝึกสอน

data	SVR	Proposed method	Improvement	% Improvement
temp	26.59735	26.14377	0.45358	1.705357865
milk	502.5391	462.08746	40.45166	8.049454936
SLP	1.02319	1.01998	0.00321	0.313724724
co2	0.22573	0.18839	0.03734	16.54188632
SOI	2.51064	2.48766	0.02298	0.915304464



ตารางที่ 4.6 แสดงการเปรียบเทียบค่าสหสัมพันธ์จากข้อมูลฝึกสอน

data	SVR	Proposed method	Improvement	% Improvement
temp	0.94100	0.94173	0.00073	0.77517
milk	0.99951	0.99955	0.00004	0.04002
SLP	0.99502	0.99504	0.00002	0.02010
co2	0.9999989	0.9999991	0.0000002	0.0002
SOI	0.27569	0.28568	0.00999	34.96920

จากตารางที่ 4.7 เป็นการเปรียบเทียบโดยใช้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งผลลัพธ์ที่ได้ คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์เรกเรชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล และเทคนิคอาร์มาให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 4 ชุดข้อมูล และเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิซเซอร์ให้ผลดีขึ้น 1.34% ข้อมูลปริมาณการผลิตน้ำมันของวุ้นในแต่ละเดือนให้ผลดีขึ้น 17.15% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 3.86% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 4.31% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและคาร์วินให้ผลแย่ลง 1.99%

ตารางที่ 4.7 แสดงการเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลทดสอบ

data	ARIMA	SVR	Proposed method	Improvement	% Improvement
temp	15.31096	10.56466	10.42317	0.14149	1.33928
milk	85.42048	60.28028	49.94115	10.33913	17.15176
SLP	2.548502	1.921235	1.846996	0.07424	3.86413
co2	2.930608	6.510226	6.229361	0.28087	4.31421
SOI	1.984341	1.918531	1.956628	-0.03810	-1.98574

จากตารางที่ 4.8 เป็นการเปรียบเทียบโดยใช้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งผลลัพธ์ที่ได้คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล และเทคนิคอาร์มาให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 4 ชุดข้อมูล และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิซเซอร์ให้ผลดีขึ้น 1.25% ข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือนให้ผลดีขึ้น 18.57% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 5.52% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 5.60% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีติและคาร์วินให้ผลแย่งลง 2.60%

ตารางที่ 4.8 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลทดสอบ

data	ARIMA	SVR	Proposed method	Improvement	% Improvement
temp	12.73791	9.07392	8.96033	0.11359	1.25183
milk	66.36006	48.35438	39.37634	8.97804	18.56717
SLP	2.155819	1.57495	1.48809	0.08686	5.51510
co2	2.42601	5.58294	5.27019	0.31275	5.60189
SOI	1.530834	1.45858	1.49647	-0.03789	-2.59773

จากตารางที่ 4.9 เป็นการเปรียบเทียบโดยใช้ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน ซึ่งผลลัพธ์ที่ได้คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 2 ชุดข้อมูลและเทคนิคอาร์มาให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล โดยพบว่าข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือนให้ผลดีขึ้น 18.14% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 3.56% และข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 5.61%

ตารางที่ 4.9 แสดงการเปรียบเทียบค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อนจากข้อมูลทดสอบ

data	ARIMA	SVR	Proposed method	Improvement	% Improvement
milk	7.40727	5.50175	4.50366	0.99809	18.14131867
SLP	24.47317	15.3652	14.81896	0.54624	3.555046469
co2	0.72331	1.65733	1.56431	0.09302	5.612642021

จากตารางที่ 4.10 เป็นการเปรียบเทียบโดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งผลลัพธ์ที่ได้คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล และเทคนิคอาร์มาให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 4 ชุดข้อมูล และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิค ซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิซเซอร์ให้ผลดีขึ้น 1.25% ข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือนให้ผลดีขึ้น 18.57% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 5.52% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 5.60% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีติและคาร์วินให้ผลแย่ง 2.60%

ตารางที่ 4.10 แสดงการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลทดสอบ

data	ARIMA	SVR	Proposed method	Improvement	% Improvement
temp	12.73791	9.07392	8.96033	0.11359	1.25183
milk	66.36006	48.35438	39.37634	8.97804	18.56717
SLP	2.155819	1.57495	1.48809	0.08686	5.51510
co2	2.42601	5.58294	5.27019	0.31275	5.60189
SOI	1.530834	1.45858	1.49647	-0.03789	-2.59773

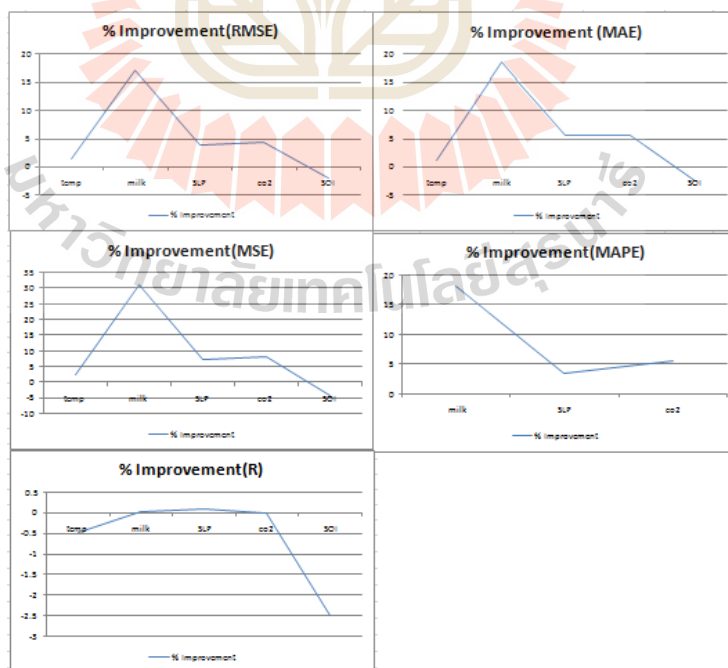
จากตารางที่ 4.11 เป็นการเปรียบเทียบโดยใช้ค่าสหสัมพันธ์ ซึ่งผลลัพธ์ที่ได้คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 2 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 2 ชุดข้อมูล และเทคนิคอาร์มาให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 2 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิค ซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิซเซอร์ให้ผลแย่ง 0.48% ข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือนให้ผลดีขึ้น 0.007% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 0.087% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 0.00005% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีติและคาร์วินให้ผลแย่ง 2.50%

ตารางที่ 4.11 แสดงการเปรียบเทียบค่าสหสัมพันธ์จากข้อมูลทดสอบ

data	ARIMA	SVR	Proposed method	Improvement	% Improvement
temp	-0.04061	0.76070	0.75706	-0.00364	-0.480807334
milk	0.99752	0.99882	0.99889	0.00007	0.007007779
SLP	0.96984	0.9884	0.98926	0.00086	0.086933668
co2	0.99996	0.9999494	0.9999499	0.0000005	0.00005
SOI	0.02541	0.24257	0.23666	-0.00591	-2.497253444

#### 4.4 อภิปรายผล

จากการทดสอบประสิทธิภาพพบว่าเทคนิคการเพิ่มประสิทธิภาพให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม โดยการนำค่าความคลาดเคลื่อนมาช่วยในการสร้างตัวแบบสามารถทำได้ เนื่องจากการนำค่าความคลาดเคลื่อนมาใช้ในการสร้างโมเดลจะทำให้โมเดลที่ได้นั้นเข้าใจในตัวข้อมูลฝึกสอนมากขึ้น ซึ่งทำให้ผลลัพธ์ที่ได้เมื่อนำไปใช้กับข้อมูลทดสอบได้ผลลัพธ์ที่ดีขึ้น แต่ในกรณีของข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาสีติและคาร์วินที่ให้ผลลัพธ์แย่ง อาจเนื่องมาจากเกิดการ โอเวอร์ฟิตติ้ง (Overfitting) ซึ่งเป็นเหตุการณ์ที่โมเดลสามารถให้ผลลัพธ์ได้ดีมากในข้อมูลฝึกสอนแต่ให้ผลลัพธ์ที่แย่งกับข้อมูลทดสอบดังแสดงในรูปที่ 4.2



รูปที่ 4.2 แสดงกราฟ % Improvement จากเกณฑ์ทดสอบ 5 เกณฑ์

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

ในปัจจุบันได้มีเทคนิคที่ใช้ในการพยากรณ์มากมาย ซึ่งแต่ละเทคนิคมีทั้งข้อดี และข้อเสียแตกต่างกัน โดยมีการพัฒนาเทคนิคใหม่ ๆ โดยการคิดค้นเทคนิคใหม่ การนำเทคนิคเดิมมาพัฒนาให้ดีขึ้น หรือการนำเทคนิคหลาย ๆ อย่างมาใช้ร่วมกัน

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันดั้งเดิม โดยการนำค่าความคลาดเคลื่อนมาช่วยในการสร้างตัวแบบ เพื่อเป็นเทคนิคใหม่ในการพยากรณ์ โดยมีสมมติฐานที่ว่าในการสร้างโมเดลฝึกสอนสิ่งที่ได้เมื่อนำมาเทียบกับค่าจริงคือค่าความคลาดเคลื่อน ซึ่งค่าความคลาดเคลื่อนที่ได้นั้นเป็นค่าความสัมพันธ์ที่ขาดหายไปในการพยากรณ์ครั้งแรก ดังนั้นเพื่อที่จะทำให้การพยากรณ์มีความแม่นยำมากขึ้นจึงได้ทำการนำค่าความคลาดเคลื่อนที่ได้มาทำการพยากรณ์อีกครั้งเพื่อเติมความสัมพันธ์ที่ขาดหายไป ซึ่งเทคนิคที่ใช้ในการทดลองคือเทคนิคซัพพอร์ตเวกเตอร์เรกเรชัน และได้มีการเปรียบเทียบความแม่นยำกับเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันดั้งเดิม และเทคนิคอาร์มา โดยเกณฑ์ที่ใช้ในการทดสอบความแม่นยำ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error : MSE) ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน (Mean Absolute Percentage Error : MAPE) และค่าสหสัมพันธ์ (Correlation coefficient : R)

#### 5.1 สรุปผลการวิจัย

จากการวิจัยสามารถสรุปผลการทดลองได้ดังนี้

##### 1) ข้อมูลที่ใช้

ข้อมูลที่ใช้เป็นข้อมูลที่อยู่ในรูปของอนุกรมเวลา โดยชุดข้อมูลที่ใช้มีทั้งหมด 5 ชุด ข้อมูลประกอบด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพิซเซอร์ (temp) ข้อมูลปริมาณการผลิตนมของวัวในแต่ละเดือน (milk) ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วิน (SLP) ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา (CO2) และข้อมูลค่าดัชนีที่คำนวณจากค่าความ

กคอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาสีติและคาร์วิน (SOI) โดยทำการแบ่งข้อมูลเป็น 2 ชุด คือข้อมูลฝึกสอน และข้อมูลทดสอบ ซึ่งแบ่งในอัตราส่วนร้อยละ 70 และ 30 ตามลำดับ

## 2) การเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน

เทคนิคที่ใช้ในการเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน คือ การนำค่าความคลาดเคลื่อนที่ได้จากโมเดลที่สร้างด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน และข้อมูลฝึกสอนมาสร้างโมเดลอีกครั้ง โดยใช้เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน ซึ่งสิ่งที่ได้คือโมเดล 2 โมเดล โดยเมื่อนำไปใช้กับข้อมูลทดสอบผลลัพธ์ที่ได้คือค่าผลรวมของค่าพยากรณ์ที่ได้จากโมเดลทั้ง 2 โมเดล

## 3) การประเมินประสิทธิภาพ

ในการทดสอบความแม่นยำของโมเดลนั้นข้อมูลที่ใช้คือข้อมูลทดสอบ โดยใช้โมเดลผสมในการพยากรณ์ค่า ซึ่งวัดค่าโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ค่าสัมบูรณ์ของเปอร์เซ็นต์ของความคลาดเคลื่อน และค่าสหสัมพันธ์ โดยทำการเปรียบเทียบเทคนิคทั้งหมด 3 เทคนิค คือ เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน และเทคนิคอาร์มา จากการทดสอบพบว่าเทคนิคใหม่ให้ค่าดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันมีค่าดีที่สุด 1 ชุดข้อมูล และเทคนิคอาร์มามีค่าดีที่สุด 1 ชุดข้อมูล แต่เมื่อเทียบเฉพาะเทคนิคใหม่ และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน พบว่าเทคนิคใหม่ให้ผลดีกว่า 4 ชุดข้อมูลจาก 5 ชุดข้อมูล

จากการวิจัยพบว่าเทคนิคใหม่สามารถเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันได้ โดยเทคนิคใหม่จะทำให้โมเดลที่ได้มีค่าเข้าใกล้ข้อมูลฝึกสอนมากขึ้น ทำให้การพยากรณ์มีประสิทธิภาพมากขึ้น

## 5.2 ปัญหาและข้อเสนอแนะ

เนื่องจากเทคนิคใหม่นี้เป็นการสร้างโมเดลที่ทำให้ค่าการพยากรณ์มีค่าเข้าใกล้ข้อมูลทดสอบมากขึ้น ซึ่งอาจทำให้เกิดปัญหาโอเวอร์ฟิตติ้งขึ้น ได้ดังตัวอย่างชุดข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาสีติและคาร์วิน โดยจะเห็นได้ว่าโมเดลจากข้อมูลฝึกสอนให้ค่าสูงขึ้นแต่กลับให้ค่าต่ำลงจากข้อมูลทดสอบ

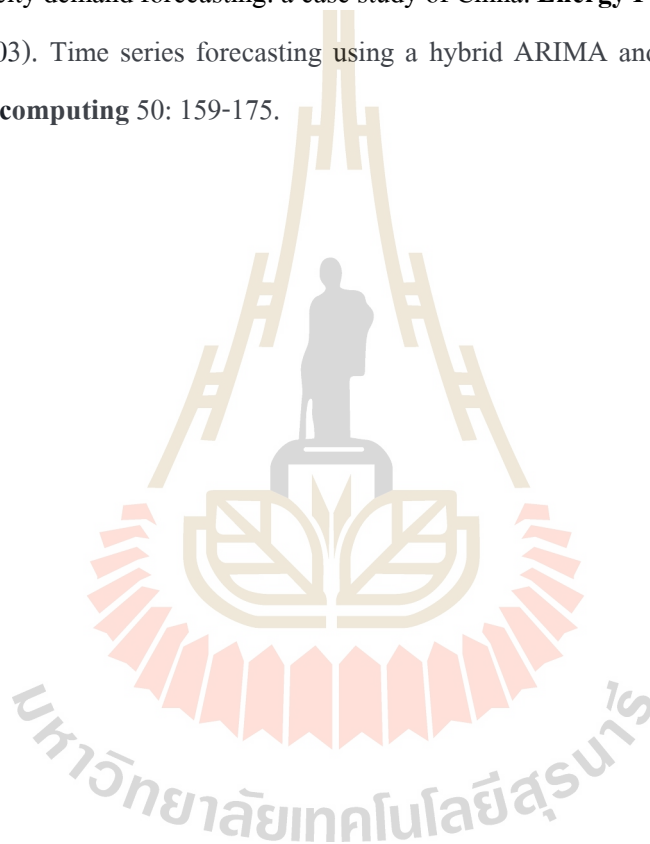
เทคนิคใหม่นี้เป็นการนำเสนอวิธีการใหม่ ซึ่งสามารถนำไปใช้พัฒนาต่อยอดเพื่อหาเทคนิคการพยากรณ์ที่ดียิ่งขึ้นกว่านี้ได้



## รายการอ้างอิง

- มุกดา แม้นมินทร์ (2549). อนุกรมเวลาและการพยากรณ์ Time series and forecasting. กรุงเทพฯ :  
ประกายพริก.
- Hatchett, R. B., Brorsen, B. W., & Anderson, K. B. (2010). Optimal length of moving average to  
forecast futures basis. **Journal of Agricultural and Resource Economics**, 18-33.
- Kapgate, D. (2014). Weighted moving average forecast model based prediction service broker  
algorithm for cloud computing. **International Journal of Computer Science and  
Mobile Computing**, 3(2), 71-79.
- Hunter, J. S. (1986). The exponentially weighted moving average. **J. Quality Technol.**, 18(4),  
203-210.
- Box, G. & Jenkins, G.M. (1976). **Time series analysis, control, and forecasting**. San Francisco,  
CA: Holden.
- Brockwell, P.J. & Davis, R.A. (2013) **Time series: theory and methods**. Springer Science &  
Business Media.
- Chen, K. & Wang, C. (2007). A hybrid SARIMA and support vector machines in forecasting the  
production values of the machinery industry in Taiwan. **Expert Systems with  
Applications** 32.1: 254-264.
- de Oliveira, J. & Ludermir, T. (2016). A hybrid evolutionary decomposition system for time  
series forecasting. **Neurocomputing** 180: 27-34.
- Dickey, D. & Fuller, W. (1979). Distribution of the estimators for autoregressive time series with  
a unit root. **Journal of the American Statistical Association** 74.366a: 427-431.
- Eberhart, R. C., & Kennedy, J. (1995). A new optimizer using particle swarm theory.  
**Proceedings of the Sixth International Symposium on Micro Machine and Human  
Science** 1: 39-43.
- Gershenson, C. (2003). Artificial neural networks for beginners. **arXiv preprint** cs/0308031.
- Janacek G. & Swift L. (1993). **Time Series: Forecasting, Simulation, Applications**. Ellis  
Horwood.

- Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. **Procedia Computer Science** 48: 173-179.
- Smola, A. & Schölkopf, B. (2004). A tutorial on support vector regression. **Statistics and computing** 14.3: 199-222.
- Wang, Y., et al. (2012) Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. **Energy Policy** 48: 284-294.
- Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing** 50: 159-175.







ภาคผนวก ก

รหัสต้นฉบับของโปรแกรมภาษา R

**%สร้างฟังก์ชัน RMSE, MAE, MAPE, MSE, R**

```
rmse <- function(error)
{
  sqrt(mean(error^2))
}
mae <-function(error)
{
  mean(abs(error))
}
mape <-function(error,t)
{
  mean(abs(error/t))*100
}
mse <-function(error)
{
  mean(error^2)
}
r <-function(a,p)
{
  sum(a*p)/( sqrt(sum(a*a))*sqrt(sum(p*p)) )
}
```

**%สร้างโมเดลจากชุดข้อมูล temp**

**library(e1071)** %ใช้ library ชื่อ e1071 ในการสร้างโมเดลซ์พพอร์ตเวกเตอร์รีเกรสชัน

**%สร้างโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันจากข้อมูลฝึกสอน**

```
train<-read.csv(file="temptrain.csv") %โหลดข้อมูลฝึกสอนใส่ตัวแปร train
```

```
myformular<-temp ~ Date+month+year %กำหนดตัวแปรที่ใช้ในการพยากรณ์
```

```
model <- svm(myformular,data=train) %สร้างโมเดลซัพพอร์ตเวกเตอร์รีเกรสชัน
```

**%สร้างชุดข้อมูลใหม่เพื่อใช้ในการพยากรณ์(เป็นข้อมูลฝึกสอน)**

```
newdata = data.frame(Date=train$Date,month=train$month,year=train$year)
```

**%พยากรณ์ข้อมูลฝึกสอนด้วยโมเดลซัพพอร์ตเวกเตอร์รีเกรสชัน**

```
predictedY <- predict(model,newdata=newdata)
```

```
error <- train$temp - predictedY %หาค่าความคลาดเคลื่อน
```

```
svrmse <- rmse(error) %แปลงค่าความคลาดเคลื่อนเป็น RMSE
```

```
svrmae <- mae(error) %แปลงค่าความคลาดเคลื่อนเป็น MAE
```

```
svrmape <- mape(error,train$temp) %แปลงค่าความคลาดเคลื่อนเป็น MAPE
```

```
svrmse <- mse(error) %แปลงค่าความคลาดเคลื่อนเป็น MSE
```

```
svrr <- r(train$temp , predictedY) %แปลงค่าความคลาดเคลื่อนเป็น R
```

**%สร้างชุดข้อมูลใหม่เพื่อใช้ในการสร้างโมเดลจากค่าความคลาดเคลื่อน(เป็นข้อมูลฝึกสอน)**

```
newdata3 = data.frame(error=error,Date=train$Date,month=train$month,year=train$year)
```

```
myformular2<-error ~ Date+month+year %กำหนดตัวแปรที่ใช้ในการพยากรณ์
```

```
model2 <- svm(myformular2,data=newdata3) %สร้างโมเดลจากค่าความคลาดเคลื่อน
```

**%สร้างชุดข้อมูลใหม่เพื่อใช้ในการพยากรณ์(เป็นข้อมูลฝึกสอน)**

```
newda = data.frame(Date=train$Date,month=train$month,year=train$year)
```

**%พยากรณ์ข้อมูลฝึกสอนด้วยโมเดลค่าความคลาดเคลื่อน**

```

predicted<- predict(model2,newdata=newda)

trainerror <- train$temp - (predictedY+predicted)      %หาค่าความคลาดเคลื่อนโดยใช้ทั้ง2โมเดล

trainrmse <- rmse(trainerror)      %แปลงค่าความคลาดเคลื่อนเป็น RMSE

trainmae <- mae(trainerror)      %แปลงค่าความคลาดเคลื่อนเป็น MAE

trainmape <- mape(trainerror,train$temp) %แปลงค่าความคลาดเคลื่อนเป็น MAPE

trainmse <- mse(trainerror)      %แปลงค่าความคลาดเคลื่อนเป็น MSE

trainr <- r(train$temp , (predictedY+predicted)) %แปลงค่าความคลาดเคลื่อนเป็น R

%ทดสอบโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันด้วยข้อมูลทดสอบ

test<-read.csv(file="temptest.csv")      %โหลดข้อมูลทดสอบใส่ตัวแปร test

%สร้างชุดข้อมูลใหม่เพื่อใช้ในการพยากรณ์ด้วยโมเดลซัพพอร์ตเวกเตอร์รีเกรสชัน(เป็นข้อมูลทดสอบ)

newdata2 = data.frame(Date=test$Date,month=test$month,year=test$year)

%พยากรณ์ด้วยโมเดลซัพพอร์ตเวกเตอร์รีเกรสชัน

predictedY2 <- predict(model,newdata=newdata2)

error2 = test$temp-predictedY2      %หาค่าความคลาดเคลื่อน

svrrmse2 = rmse(error2)      %แปลงค่าความคลาดเคลื่อนเป็น RMSE

svrrmae2 = mae(error2)      %แปลงค่าความคลาดเคลื่อนเป็น MAE

svrrmape2 <- mape(error2,test$temp) %แปลงค่าความคลาดเคลื่อนเป็น MAPE

svrrmse2 <- mse(error2)      %แปลงค่าความคลาดเคลื่อนเป็น MSE

svrr2 <- r(test$temp,predictedY2)      %แปลงค่าความคลาดเคลื่อนเป็น R

%สร้างชุดข้อมูลใหม่เพื่อใช้ในการพยากรณ์โมเดลค่าความคลาดเคลื่อน(เป็นข้อมูลทดสอบ)

newdata4 = data.frame(Date=test$Date,month=test$month,year=test$year)

```

```

predictedY3 <- predict(model2,newdata=newdata4)    %พยากรณ์ด้วยโมเดลค่าความ
คลาดเคลื่อน

error3 <- test$temp - (predictedY2+predictedY3)    %หาค่าความคลาดเคลื่อน

svrmse3 <- rmse(error3)                          %แปลงค่าความคลาดเคลื่อนเป็น RMSE

svrmae3 <- mae(error3)                          %แปลงค่าความคลาดเคลื่อนเป็น MAE

svrmape3 <- mape(error3,test$temp) %แปลงค่าความคลาดเคลื่อนเป็น MAPE

svrmse3 <- mse(error3)                          %แปลงค่าความคลาดเคลื่อนเป็น MSE

svrr3 <- r(test$temp , (predictedY2+predictedY3)) %แปลงค่าความคลาดเคลื่อนเป็น R

%สร้างโมเดลอาร์มาจากข้อมูลฝึกสอนและทดสอบโมเดล

library(forecast)                                %ใช้ library ชื่อ forecast ในการสร้างโมเดลอาร์มา

myts <- ts(train$temp, start=1, end=1022, frequency=1) %สร้างข้อมูลเพื่อใช้ในการสร้างโมเดลอาร์มา

auto.arima(myts)                                %หาค่าพารามิเตอร์ที่เหมาะสม

fit <- arima(myts, order=c(2,1,1))               %สร้างโมเดลอาร์มา

a <- predict(fit,n.ahead = 439)                 %ทำนายด้วยตัวแบบอาร์มา

predicta <- a$pred[1:439]                       %นำค่าที่ทำนายเก็บไว้ในตัวแปร predicta

arimaerror <- test$temp - predicta              %หาค่าความคลาดเคลื่อน

rmse(arimaerror)                               %แปลงค่าความคลาดเคลื่อนเป็น RMSE

mae(arimaerror)                               %แปลงค่าความคลาดเคลื่อนเป็น MAE

mse(error)                                     %แปลงค่าความคลาดเคลื่อนเป็น MSE

mape(error,test$temp) %แปลงค่าความคลาดเคลื่อนเป็น MAPE

r(test$temp , predicta) %แปลงค่าความคลาดเคลื่อนเป็น R

```

%สร้างโมเดลจากชุดข้อมูล milk

%สร้างโมเดลซัพพอร์ตเวกเตอร์เรกเรชันจากข้อมูลฝึกสอน

```

library(e1071)

train<-read.csv(file="milktrain.csv")

myformular<-milk ~ month+year

model <- svm(myformular,data=train)

newdata = data.frame(month=train$month,year=train$year)

predictedY <- predict(model,newdata=newdata)

error <- train$milk - predictedY

svrmse <- rmse(error)

svrmae <- mae(error)

svrmape <- mape(error,train$milk)

svrmse <- mse(error)

svrr <- r(train$milk , predictedY)

newdata3 = data.frame(error=error,month=train$month,year=train$year)

myformular2<-error ~ month+year

model2 <- svm(myformular2,data=newdata3)

newda = data.frame(month=train$month,year=train$year)

predicted<- predict(model2,newdata=newda)

trainerror <- train$milk - (predictedY+predicted)

trainrmse <- rmse(trainerror)

trainmae <- mae(trainerror)

trainmape <- mape(trainerror,train$milk)

trainmse <- mse(trainerror)

trainr <- r(train$milk , (predictedY+predicted))

```

%ทดสอบโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันด้วยข้อมูลทดสอบ

```
test<-read.csv(file="milktest.csv")

newdata2 = data.frame(month=test$month,year=test$year)

predictedY2 <- predict(model,newdata=newdata2)

error2 = test$milk-predictedY2

svrrmse2 = rmse(error2)

svrmae2 = mae(error2)

svrmape2 <- mape(error2,test$milk)

svrmse2 <- mse(error2)

svrr2 <- r(test$milk,predictedY2)

newdata4 = data.frame(month=test$month,year=test$year)

predictedY3 <- predict(model2,newdata=newdata4)

error3 <- test$milk - (predictedY2+predictedY3)

svrrmse3 <- rmse(error3)

svrmae3 <- mae(error3)

svrmape3 <- mape(error3,test$milk)

svrmse3 <- mse(error3)

svrr3 <- r(test$milk , (predictedY2+predictedY3))
```

%สร้างโมเดลอาร์มาจากข้อมูลฝึกสอนและทดสอบโมเดล

```
library(forecast)

myts <- ts(train$milk, start=1, end=118, frequency=1)

auto.arima(myts)

fit <- arima(myts, order=c(2,1,2))

a <- predict(fit,n.ahead = 50)

predicta <- a$pred[1:50]
```

```

error <- test$milk - predicta
rmse(error)
mae(error)
mse(error)
mape(error,test$milk)
r(test$milk , predicta)

%สร้างโมเดลจากชุดข้อมูล SLP

%สร้างโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันจากข้อมูลฝึกสอน

library(e1071)
train<-read.csv(file="slptrain.csv")
myformular<-SLP ~ month+year
model <- svm(myformular,data=train)
newdata = data.frame(month=train$month,year=train$year)
predictedY <- predict(model,newdata=newdata)
error <- train$SLP - predictedY
svrrmse <- rmse(error)
svrmae <- mae(error)
svrmape <- mape(error,train$SLP)
svrmse <- mse(error)
svrr <- r(train$SLP , predictedY)
newdata3 = data.frame(error=error,month=train$month,year=train$year)
myformular2<-error ~ month+year
model2 <- svm(myformular2,data=newdata3)
newda = data.frame(month=train$month,year=train$year)
predicted<- predict(model2,newdata=newda)

```



```
trainerror <- train$SLP - (predictedY+predicted)
```

```
trainrmse <- rmse(trainerror)
```

```
trainmae <- mae(trainerror)
```

```
trainmape <- mape(trainerror,train$SLP)
```

```
trainmse <- mse(trainerror)
```

```
trainr <- r(train$SLP , (predictedY+predicted))
```

%ทดสอบโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันด้วยข้อมูลทดสอบ

```
test<-read.csv(file="slptest.csv")
```

```
newdata2 = data.frame(month=test$month,year=test$year)
```

```
predictedY2 <- predict(model,newdata=newdata2)
```

```
error2 = test$SLP-predictedY2
```

```
svrrmse2 = rmse(error2)
```

```
svrmae2 = mae(error2)
```

```
svrmape2 <- mape(error2,test$SLP)
```

```
svrmse2 <- mse(error2)
```

```
svrr2 <- r(test$SLP,predictedY2)
```

```
newdata4 = data.frame(month=test$month,year=test$year)
```

```
predictedY3 <- predict(model2,newdata=newdata4)
```

```
error3 <- test$SLP - (predictedY2+predictedY3)
```

```
svrrmse3 <- rmse(error3)
```

```
svrmae3 <- mae(error3)
```

```
svrmape3 <- mape(error3,test$SLP)
```

```
svrmse3 <- mse(error3)
```

```
svrr3 <- r(test$SLP , (predictedY2+predictedY3))
```

%สร้างโมเดลอาร์มาจากข้อมูลฝึกสอนและทดสอบโมเดล

```

library(forecast)

myts <- ts(train$SLP, start=1, end=979, frequency=1)

auto.arima(myts)

fit <- arima(myts, order=c(5,0,0))

a <- predict(fit,n.ahead = 421)

predicta <- a$pred[1:421]

error <- test$SLP - predicta

rmse(error)

mae(error)

mse(error)

mape(error,test$SLP)

r(test$SLP , predicta)

%สร้างโมเดลจากชุดข้อมูล CO2

%สร้างโมเดลซัพพอร์ตเวกเตอร์เกรดชันจากข้อมูลฝึกสอน

library(e1071)

train<-read.csv(file="co2train.csv")

myformular<-CO2 ~ month+year

model <- svm(myformular,data=train)

newdata = data.frame(month=train$month,year=train$year)

predictedY <- predict(model,newdata=newdata)

error <- train$CO2 - predictedY

svrrmse <- rmse(error)

svrmae <- mae(error)

svrmape <- mape(error,train$CO2)

svrmse <- mse(error)

```

```

svrr <- r(train$CO2 , predictedY)

newdata3 = data.frame(error=error,month=train$month,year=train$year)

myformular2<-error ~ month+year

model2 <- svm(myformular2,data=newdata3)

newda = data.frame(month=train$month,year=train$year)

predicted<- predict(model2,newdata=newda)

trainerror <- train$CO2 - (predictedY+predicted)

trainrmse <- rmse(trainerror)

trainmae <- mae(trainerror)

trainmape <- mape(trainerror,train$CO2)

trainmse <- mse(trainerror)

trainr <- r(train$CO2 , (predictedY+predicted))

%ทดสอบโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันด้วยข้อมูลทดสอบ

test<-read.csv(file="co2test.csv")

newdata2 = data.frame(month=test$month,year=test$year)

predictedY2 <- predict(model,newdata=newdata2)

error2 = test$CO2-predictedY2

svrrmse2 = rmse(error2)

svrrmae2 = mae(error2)

svrrmape2 <- mape(error2,test$CO2)

svrrmse2 <- mse(error2)

svrr2 <- r(test$CO2,predictedY2)

newdata4 = data.frame(month=test$month,year=test$year)

predictedY3 <- predict(model2,newdata=newdata4)

error3 <- test$CO2 - (predictedY2+predictedY3)

svrrmse3 <- rmse(error3)

```

```
svrmae3 = mae(error3)
svrmape3 <- mape(error3,test$CO)
svrmse3 <- mse(error3)
svrr3 <- r(test$CO2 , (predictedY2+predictedY3))
```

%สร้างโมเดลอาร์มาจากข้อมูลฝึกสอนและทดสอบโมเดล

```
library(forecast)
myts <- ts(train$CO2, start=1, end=136, frequency=1)
auto.arima(myts)
fit <- arima(myts, order=c(2,1,2))
a <- predict(fit,n.ahead = 56)
predicta <- a$pred[1:56]
error <- test$CO2 - predicta
rmse(error)
mae(error)
mse(error)
mape(error,test$CO2)
r(test$CO2 , predicta)
```

%สร้างโมเดลจากชุดข้อมูล SOI

%สร้างโมเดลซัพพอร์ตเวกเตอร์เรกเรชันจากข้อมูลฝึกสอน

```
library(e1071)
train<-read.csv(file="SOItrain.csv")
myformular<-SOI ~ month+year
model <- svm(myformular,data=train)
newdata = data.frame(month=train$month,year=train$year)
predictedY <- predict(model,newdata=newdata)
```

```

error <- train$SOI - predictedY

svrrmse <- rmse(error)

svrmae <- mae(error)

svrmape <- mape(error,train$SOI)

svrmse <- mse(error)

svrr <- r(train$SOI , predictedY)

newdata3 = data.frame(error=error,month=train$month,year=train$year)

myformular2<-error ~ month+year

model2 <- svm(myformular2,data=newdata3)

newda = data.frame(month=train$month,year=train$year)

predicted<- predict(model2,newdata=newda)

trainerror <- train$SOI - (predictedY+predicted)

trainrmse <- rmse(trainerror)

trainmae <- mae(trainerror)

trainmape <- mape(trainerror,train$SOI)

trainmse <- mse(trainerror)

trainr <- r(train$SOI , (predictedY+predicted))

%ทดสอบโมเดลซัพพอร์ตเวกเตอร์รีเกรสชันด้วยข้อมูลทดสอบ

test<-read.csv(file="SOItest.csv")

newdata2 = data.frame(month=test$month,year=test$year)

predictedY2 <- predict(model,newdata=newdata2)

error2 = test$SOI-predictedY2

svrrmse2 = rmse(error2)

svrmae2 = mae(error2)

svrmape2 <- mape(error2,test$SOI)

svrmse2 <- mse(error2)

```

```

svrr2 <- r(test$SOI,predictedY2)

newdata4 = data.frame(month=test$month,year=test$year)

predictedY3 <- predict(model2,newdata=newdata4)

error3 <- test$SOI - (predictedY2+predictedY3)

svrrmse3 <- rmse(error3)

svrmae3 <- mae(error3)

svrmape3 <- mape(error3,test$SOI)

svrmse3 <- mse(error3)

svrr3 <- r(test$SOI , (predictedY2+predictedY3))

%สร้างโมเดลอาร์มาจากข้อมูลฝึกสอนและทดสอบโมเดล

library(forecast)

myts <- ts(train$SOI, start=1, end=378, frequency=1)

auto.arima(myts)

fit <- arima(myts, order=c(2,0,2))

a <- predict(fit,n.ahead = 162)

predicta <- a$pred[1:162]

error <- test$SOI - predicta

rmse(error)

mae(error)

mse(error)

mape(error,test$SOI)

r(test$SOI , predicta)

```



ภาคผนวก ข

บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่

มหาวิทยาลัยเทคโนโลยีสุรนารี

## รายชื่อบทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

Teetawat Kaewwijit, Nuntawut Kaoungku, Kittisak Kerdprasop, and Nittaya Kerdprasop

(2015)THE STUDY OF THE BEST PROBABILITY DISTRIBUTION TO PREDICT A GOLD PRICE WITH LINEAR REGRESSION. The 9th South East Asia Technical University Consortium (SEATUC) Symposium. Suranaree University of Technology. July 27-30, 2015.

ธีร์ธวัช แก้ววิจิตร, นิตยา เกิดประสพ, กิตติศักดิ์ เกิดประสพ. 2560.การเพิ่มประสิทธิภาพซอฟต์แวร์  
เวกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา.วารสารวิทยาศาสตร์และเทคโนโลยี  
มหาวิทยาลัยมหาสารคาม. (ปีที่ 36 ฉบับที่ 4 (กรกฎาคม-สิงหาคม 2560))





## THE STUDY OF THE BEST PROBABILITY DISTRIBUTION TO PREDICT A GOLD PRICE WITH LINEAR REGRESSION

Teetawat Kaewwijit, Nuntawut Kaoungku, Kittisak Kerdprasop, Nittaya Kerdprasop  
School of Computer Engineering, Suranaree University of Technology, Thailand

### ABSTRACT

This paper presents a best probability distribution for predict a gold price by linear regression. The dataset is a gold price, refer by oil price for compare and parameter to predict a gold price by used AIC to find the best model. The probability distribution used 3 models gaussian, gamma, inverse gaussian and link function used 2 models Identity, log total 6 models. We will implement these imputation techniques with R language. From experiment a best probability distribution is gamma and a link function is log by AIC is 55287.

Keyword: linear regression, gold price, probability distribution, R language

### 1. INTRODUCTION

From the present state of society in the era of globalization, the environment is changing rapidly. The perception is extremely important. The technology, trade, communication and information system have no borders. The competition is rising in all events. Including economic investment factors were also associated with an investment in a variety of formats for the exchange of goods and products directly, such as financial market shares or bonds, Investment in the oil market and gold market.

Gold is another important investment. There are currently investing in the gold market as a universal and global network. Gold is a metal that has a luster and beauty that has been popular in jewelry making. Gold have four basic features that make them stand out as the gold market demand, including beauty and durability is a rare mineral. And can be reused by recycling purified by annealing. It also has to be utilized such as gold.

1. Jewelry Industry From past to present, gold is a metal used in jewelry making has been the most popular.

2. Economic and fiscal stability since gold is a rare metal with a high demand for gold as a medium of exchange and is a metal that is used as an international reserve. As a result, gold is used as a tool for market speculation.

3. Gold in the Electronics Industry The gold was used as components in the electronics and communications telecommunication connections such as semiconductor materials and transistors.

4. Benefits of Transport and Communications. Gold has the ability to reflect radiation, such as infrared radiation. It has been used on the space shuttle and space suits.

5. Medical and dental benefits. Has been used in radiation and reflected radiation to be used in auscultation.

The demand for gold has made a number of high gold prices and price volatility. The composition and related factors are including the price of gold.

1. The price of crude oil on the world market. (Chang, et al., 2013), (Zhang, et al., 2010)

2. The US dollar is the main currency in the foreign currency exchange.

3. The need to buy (demand) and needs to sell (supply) at different times and circumstances.

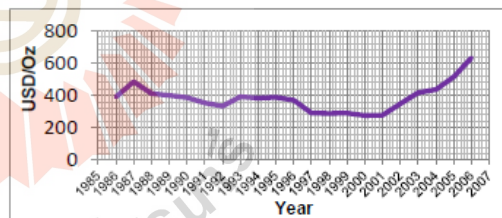


Fig. 1 shows the relationship between gold prices each year.

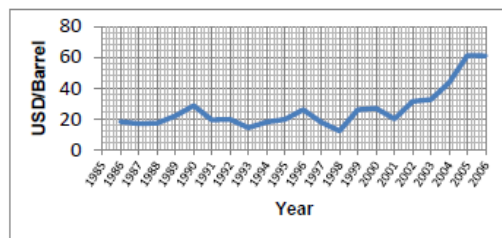


Fig. 2 shows the relationship between oil prices each year.

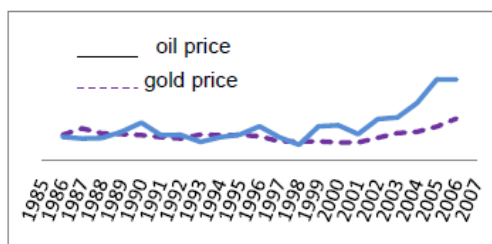


Fig. 3 shows the relationship between gold prices and the price trend of crude oil each year.

Figure 1 shows that the price of gold (World Gold Council, 2015) is increasing each year, which is significantly higher during the years 2001-2006.

Figure 2 shows that crude oil prices (Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (DCOILWTICO), 2015) is increasing each year, which is significantly higher during the years 2001-2006, as well as gold.

Figure 3 is a comparison chart between gold prices and the trend of crude oil prices. The price of crude oil and gold prices have different scales. It used for comparative only. The graph showed a trend in the same direction.

From the context of trade and investment in the gold market. The analysis Research and find out how to become a tool to predict gold prices as predicting gold prices by way of artificial neural networks (Parisi, et al., 2008) and to predict the price of gold by using an inflation factor (Ismail, et al., 2015), the researcher is interested to study. To find the best probability for predict the price of gold, based on correlation with the price of crude oil using a linear regression data as an alternative to predict the price of gold.

## 2. METHODOLOGY

This research aims to determine the best approach is probably best used for predict the price of gold, based on correlation with the price of crude oil using linear regression. And test the models using the AIC is a measure that models do best, which is how the research stage.

1. Studied Akaike's Information Criterion (AIC) (Akaike, 1981). AIC is the criteria for selecting the most appropriate model for comparing multiple criteria. It has been thought up by Mr. Akaike and Called Akaike's Information Criterion (AIC). The best model have minimum of AIC. This works best on large data samples. However, when the data is small compared to the number of parameters estimate. The models will be selected over a number of independent variables (Over-fit) by the AIC has the following equation.

$$AIC = 2k - 2\ln(L) \quad (1)$$

AIC = Akaike's Information Criterion

L = maximized value of the likelihood function  
k = the number of estimated parameters in the model

2. Studied Generalized Linear Model (GLM) General linear model equation model was first proposed by Nelder and Wedderburn in 1972 (Nelder, et al., 1972). The model consists of three parts. Random component, Systematic component and Link function by has the following equation.

$$y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_i X_{ii} \quad (2)$$

Random component are variables associated with the probability distribution of the dependent variable (response variable) or type of exponential family.

Systematic component is assigned the function of independent variables used to predict which variables are two types. 1.  $\beta$  = correlation 2. X = the values of the independent variables.

Link function is an indication of the relationship between the variable component and system components. The variable in the equation is  $y_i$

3. R language. R Language is a computer language used in statistics and display data in graphical form. R. language can be widely because of the function blocks (libraries packages) and many are free, so it is widely used in education. And business, in which the research is an important statement as follows (Kerdprasop, 2012).

```
-price<-read.csv(file="goldpriceoil.csv")
command is used to import data from
"goldpriceoil.csv" stored in a variable name "price".
-myformular<-price$gold ~ price$oil +
price$month + price$year command is used to
stored function to variable name "myformular".
-gaussian.iden<-glm(myformular,
family=gaussian(identity)) commands used in the
linear regression equation used is "myformular"
with a probability value is gaussian and link function
is identity. Model stored in the variable name
"gaussian.iden".
-pred.gauiden<-predict(gaussian.iden,
type="response") command is used to predict
"gaussian.iden" model and stored predict variable
in "pred.gauiden".
-summary(gaussian.iden) command is used
to showing the value of the variable in this example
is to show models "gaussian.iden".
```

```
- plot(price$num, price$dollar) command is
used to create graph. The first variable is plotted on
the x axis and the second variable is the axis y.
```

4. The concept of research. The main idea of this research is design the algorithms to find the best probability. The conceptual framework of the research shown in figure 4 and for the process used to find the best probability is divided into 3 steps.

Step 1: get gold and crude oil prices dataset to create a linear model. The probability is that the gaussian and link function is identity.

Step 2: use regression analysis to find the AIC and change the probability and link function.

Then use regression analysis to find the AIC with every model. (The probability has 3 types gaussian, gamma and inverse gaussian. The link function uses 2 types is identity and log. Total has 6 models).

Step 3: observe the AIC from each model and compared to find the best model.

### 3. EXPERIMENTAL RESULT

This research was test to compare the AIC of each model. The probability has 3 types, gaussian gamma and Inverse gaussian. The link function uses 2 types identity and locked. A total of 6 models with models that minimum of AIC is a best model, which means that the probability and the link function of the model is the best value to predicting the gold price.

From table 1 model was used probability gaussian and link function identity the AIC was 55826. Model was used probability gaussian and link function log the AIC was 55639. Model was used probability gamma and link function identity the AIC was 55420. Model was used probability gamma and link function log the AIC was 55287. Model was used probability inverse gaussian and link function identity the AIC was 55415. Model was used probability inverse gaussian and link function log the AIC was 55313. From 6 models with the minimum AIC is the model that the probability of gamma and the link function is log, the AIC is equal to 55287, it is the best model.

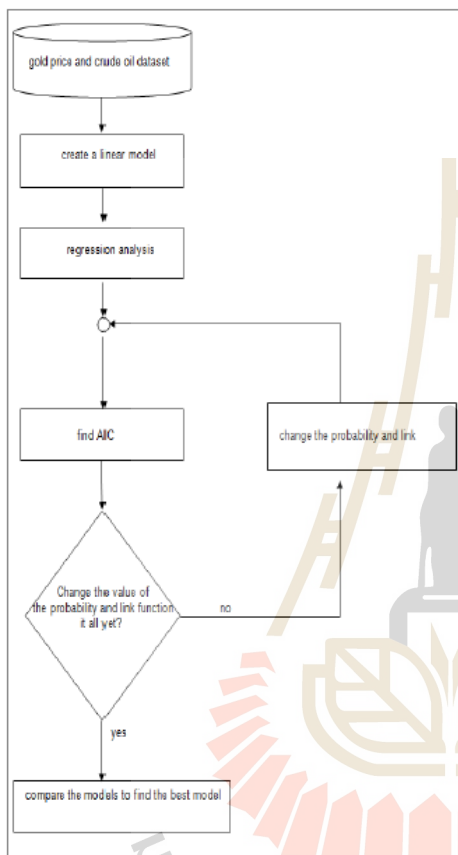


Figure 4 framework of research in the algorithm form.

Table 1 displays the results of the AIC from 6 models.

Probability Distribution	Link Function	
	Identity	Log
Gaussian	55826	55639
Gamma	55420	55287*
Inverse gaussian	55415	55313

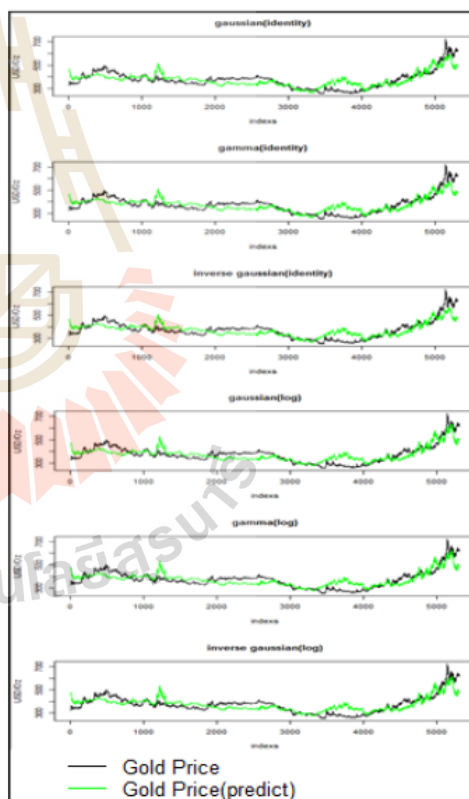


Figure 5 shows a comparison chart between the gold price and gold price prediction.



Figure 5 shows that the prediction is likely to be in line with the gold price. The graph to predict the individual values are very similar.

## CONCLUSION

Linear regression was used to predict from the linear equation to determine the relationship of the variables involved in functions that have been established. This can be used to predict the price of gold.

The results showed that the models used to value the probability of gamma and use the link function is locked, the AIC has minimal value in the six models, so this model is a best model for prediction the gold prices by using linear regression.

## REFERENCES

Akaike, H., Likelihood of a model and information criteria, *Journal of econometrics*, vol. 16(1), pp. 3-14, 1981.

Chang, H. F., Huang, L. C., Chin, M. C., Interactive relationships between crude oil prices, gold prices, and the NT-US dollar exchange rate—A Taiwan study, *Energy Policy*, vol. 63, pp. 441-448, 2013.

Crude Oil Prices: West Texas Intermediate (WTI) -Cushing, Oklahoma (DCOILWTICO), <https://research.stlouisfed.org/fred2/series/DCOILWTICO/downloaddata>, 2015.

Ismail, Z., Yahya, A., Shabri, A., Forecasting gold prices using multiple linear regression method, *American Journal of Applied Sciences*, vol. 6(8), pp. 1509, 2009.

Nelder, J. A., Baker, R. J., Generalized linear models, *Encyclopedia of Statistical Sciences*, 1972.

Parisi, A., Parisi, F., Diaz, D., Forecasting gold price changes: Rolling and recursive neural network models, *Journal of Multinational financial management*, vol. 18(5), pp. 477-487, 2008.

World Gold Council, <http://www.gold.org/research/download-the-gold-price-since-1978>, 2015.

Zhang, Y. J., Wei, Y. M., The crude oil market and the gold market: Evidence for counteraction, causality and price discovery, *Resources Policy*, vol. 35(3), pp. 168-177, 2010.

Kerdprasop, K., *Data Mining Methodology and Development*, Retrieved October 20, 2012, from <https://sites.google.com/site/kittisakthailand55/home/datamining2-55>



### Teetawat Kaewwijit

He is currently a master student with the School of Computer Engineering, Suranaree University of Technology, Thailand. His current research of interest includes Classification.



### Nuntawut Kaoungku

He is currently a doctoral student with the School of Computer Engineering, Suranaree University of Technology, Thailand. His current research includes semantic web and association.



### Kittisak Kerdprasop

He is an associate professor and chair of the School of Computer Engineering, Suranaree University of Technology, Thailand. His current research of interest includes Data mining, Artificial Intelligence, Functional and Logic Programming Languages, Computational Statistics.



### Nittaya Kerdprasop

She is an associate professor at the School of Computer Engineering, Suranaree University of Technology, Thailand. She is a member of ACM and IEEE Computer Society. Her research of interest includes Knowledge Discovery in Databases, Artificial Intelligence, Logic Programming, and Biomedical Informatics.

## การเพิ่มประสิทธิภาพซัพพอร์ตเวกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา The improvement of Support Vector Regression to Forecast Time Series

ธีรวัช แก้ววิจิตร<sup>1</sup>, นิตนา เกิดประสพ, กิตติศักดิ์ เกิดประสพ

Teetawat Kaewwijit, Nittaya Kerdprasop, Kittisak Kerdprasop

### บทคัดย่อ

ในปัจจุบันได้มีความพยายามในการหาเทคนิคใหม่ในการพยากรณ์ เพื่อให้การพยากรณ์มีความแม่นยำและความเร็วเพิ่มขึ้น โดยการคิดค้นเทคนิคใหม่ หรือการนำหลาย ๆ เทคนิคมาผสมกัน งานวิจัยนี้มีจุดประสงค์เพื่อเพิ่มประสิทธิภาพให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันในการพยากรณ์อนุกรมเวลา โดยใช้ค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำให้กับตัวแบบ ซึ่งข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลอนุกรมเวลาทั้งหมด 5 ชุดข้อมูลประกอบไปด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำฟิชเชอร์ ข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือน ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองดาร์วิน ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮิติและดาร์วิน โดยการเปรียบเทียบความแม่นยำของเทคนิคใหม่ จะทำการเปรียบเทียบกับเทคนิค 2 แบบคือเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม และเทคนิค ARIMA และวัดค่าโดยใช้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งจากผลการเปรียบเทียบพบว่าเทคนิคใหม่สามารถเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันได้

คำสำคัญ : ซัพพอร์ตเวกเตอร์รีเกรสชัน ข้อมูลอนุกรมเวลา ค่าความคลาดเคลื่อน

### Abstract

Currently, there are efforts to find new techniques in forecasting in order to improve precision and speed. The improvement is achieved by using new technique or a combination of techniques. This research aims to optimize support vector regression in forecasting time series by using the error to increase the accuracy of the model. The five datasets used in time series analysis are the daily temperature of the Fisher River, monthly milk production, the sea level pressure data at Darwin, carbon dioxide concentration at Mauna Loa mountain, and the atmospheric pressure difference between Tahiti and Darwin. The precision of the proposed model is compared against the traditional support vector regression and the ARIMA models using the root mean squared error and the mean absolute error metrics. From the experimental results, the proposed method can improve precision of the support vector regression technique.

Keyword : Support vector regression, Time Series, Error

## บทนำ

การทำนายตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง "การบอกเหตุการณ์หรือความเป็นไปที่จะเกิดในเบื้องหน้า" ซึ่งการทำนายนั้นได้มีมาตั้งแต่สมัยอดีต โดยการทำนายจะถูกใช้โดยผู้ทำพิธี ซึ่งอาจจะเป็นหมอดู พระ นักบวช คนทรง เป็นต้น โดยสิ่งที่ทำนายนั้นมักจะเป็นเรื่องเกี่ยวกับ ชีวิต ความรัก ความเป็นอยู่ ผลลัพธ์การพนันต่าง ๆ เป็นต้น โดยผลลัพธ์ในการทำนายนั้นอาจจะถูกต้องหรือไม่ถูกต้องก็ได้ เมื่อเวลาผ่านไปการทำนายได้มีการพัฒนาขึ้นโดยอาศัยหลักการต่าง ๆ เข้ามาช่วยเพื่อเพิ่มความแม่นยำในการทำนายเรียกว่าการพยากรณ์ โดยตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2554 หมายถึง "ทำนายหรือคาดการณ์โดยอาศัยหลักวิชา" ซึ่งหลักวิชาที่ว่านี้คือเทคนิคต่าง ๆ ที่มาช่วยในการพยากรณ์ โดยทางสถิติและคอมพิวเตอร์คือการนำข้อมูลต่าง ๆ ที่มีความเกี่ยวข้องกับสิ่งที่ต้องการพยากรณ์มาช่วยในการหาค่าพยากรณ์ ซึ่งค่าที่นิยมใช้ในการพยากรณ์จะเป็นค่าข้อมูลที่ต้องการพยากรณ์ย้อนหลัง โดยลักษณะของข้อมูลที่ใช้จะอยู่ในรูปแบบของอนุกรมเวลา (Time series) และมีการใช้เทคนิคต่าง ๆ ในการพยากรณ์ โดยเทคนิคที่นิยมใช้คือ เทคนิคสมการเชิงเส้นทั่วไป (Generalized Linear) เทคนิค ARIMA เทคนิคการถดถอย (Regression) เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) ซึ่งเทคนิคที่ได้กล่าวมานั้นเป็นเทคนิคที่ให้ผลลัพธ์ที่มีความแม่นยำสูง แต่ก็ยังมีความต้องการที่จะให้การพยากรณ์ที่ได้มีความแม่นยำที่สูงขึ้น ซึ่งในปัจจุบันได้มีนักวิจัยพัฒนาเทคนิคใหม่ ๆ เพื่อให้ได้ผลลัพธ์ที่ดีกว่าวิธีเดิม โดยมีการนำเทคนิคหลาย ๆ วิธีมาใช้ร่วมกัน เช่นการนำเทคนิควิชีเชิงพันธุกรรม (Genetic Algorithm) มาช่วยในการหาค่าพารามิเตอร์ในตัวแบบ (Model) เพื่อให้ได้ตัวแบบที่ดีที่สุด เป็นต้น เทคนิคหนึ่งที่น่าสนใจ และนำมาใช้ในการเพิ่มความแม่นยำคือการนำค่าความคลาดเคลื่อนที่ได้จากตัวแบบมาพิจารณาในการสร้างตัวแบบใหม่ โดยได้มีงานวิจัยที่ใช้เทคนิค ARIMA ในการสร้างตัวแบบ และนำเทคนิคเครือข่ายประสาทเทียมมาสร้างตัวแบบจากค่าความคลาดเคลื่อนที่ได้จากเทคนิค ARIMA จากนั้นนำตัว

แบบทั้ง 2 มาสร้างเป็นตัวแบบใหม่ ซึ่งแนวคิดของงานวิจัยนี้คือตัวแบบนี้ประกอบด้วยส่วนที่เป็นเชิงเส้น และส่วนที่ไม่เป็นเชิงเส้น โดยใช้เทคนิค ARIMA ในการพยากรณ์ส่วนที่เป็นเชิงเส้น และเทคนิคโครงข่ายประสาทเทียมใช้ในการพยากรณ์ส่วนที่ไม่เป็นเชิงเส้น โดยตั้งชื่อเทคนิคว่าไฮบริด (Hybrid) [1] ซึ่งเทคนิคนี้ได้มีการพัฒนาต่อโดยการเปลี่ยนแปลงเทคนิคที่ใช้ เช่น Chen และ Wang [2] เสนอเทคนิคที่ดัดแปลงมาจากเทคนิคไฮบริด ซึ่งเป็นการรวมเทคนิคของ ARIMA และซัพพอร์ตเวกเตอร์รีเกรสชัน โดยกำหนดให้ตัวแบบแรกใช้เทคนิค ARIMA จากนั้นนำค่าความคลาดเคลื่อนที่ได้นำมาใช้ในตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ และได้มีการใช้เทคนิควิชีเชิงพันธุกรรมในการช่วยหาค่าพารามิเตอร์ของซัพพอร์ตเวกเตอร์รีเกรสชัน Wang และคณะ [3] เสนอเทคนิคที่ดัดแปลงมาจากเทคนิคไฮบริดของ Zhang โดยการนำเทคนิค ARIMA โดยเริ่มต้นจากการพยากรณ์โดยใช้เทคนิค ARIMA ก่อนแล้วนำค่าความคลาดเคลื่อนที่ได้มาทำการสร้างตัวแบบ ARIMA อีกตัว จากนั้นนำตัวแบบทั้งสองตัวไปใช้ในการพยากรณ์ Khandelwal และคณะ [4] ได้มีการนำวิธีไฮบริดมาใช้ในการพยากรณ์ร่วมกับวิธี Discrete Wavelet Transform (DWT) โดยการนำ DWT ในการแบ่งข้อมูลออกเป็นข้อมูลความถี่สูง และความถี่ต่ำ จากนั้นนำวิธีการไฮบริดมาใช้ในการพยากรณ์ Oliveira และ Ludermir [5] ได้เสนอวิธีการโดยการนำขั้นตอนวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization : PSO) มาช่วยในการหาผลลัพธ์ที่ดีที่สุด และนำเทคนิคไฮบริดมาใช้ในการพยากรณ์ จากที่กล่าวมาข้างต้นพบว่าปัจจุบันได้มีความพยายามในการหาเทคนิคใหม่ ๆ เพื่อนำมาใช้ในการเพิ่มความแม่นยำในการพยากรณ์ โดยในงานวิจัยนี้ผู้วิจัยได้นำเสนอเทคนิคใหม่ โดยนำเทคนิคไฮบริดมาดัดแปลงใหม่ โดยมีสมมติฐานที่ว่าค่าความคลาดเคลื่อนที่ได้นั้นเป็นค่าความสัมพันธ์ที่ขาดหายไปในการพยากรณ์ครั้งแรก ดังนั้นเพื่อที่จะทำให้การพยากรณ์มีความแม่นยำมากขึ้น จึงได้ทำการนำค่าความคลาดเคลื่อนที่ได้มาทำการพยากรณ์อีกครั้งเพื่อเติมความสัมพันธ์ที่ขาดหายไป ซึ่งเทคนิคที่ใช้ในการทดลองคือเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน และได้มีการเปรียบเทียบความแม่นยำกับ



เทคนิคซัพพอร์ตเวกเตอร์เรกเรชันดั้งเดิม และเทคนิค ARIMA โดยเกณฑ์ที่ใช้ในการทดสอบความแม่นยำ คือ ค่ารากที่สองของค่าความคาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE)

### วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาวิธีการเพิ่มความแม่นยำให้กับเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันดั้งเดิม โดยการนำค่าความคลาดเคลื่อนมาช่วยในการสร้างตัวแบบ
2. เพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ของเทคนิคที่นำเสนอกับเทคนิคซัพพอร์ตเวกเตอร์เรกเรชันดั้งเดิม และเทคนิค ARIMA ด้วยค่ารากที่สองของค่าความคาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

### วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการทดลองเพื่อเพิ่มค่าความแม่นยำในการพยากรณ์ค่าให้กับตัวแบบโดยใช้ค่าความคาดเคลื่อนมาช่วยในการสร้างตัวแบบ ซึ่งตัวแบบที่ใช้ในการพยากรณ์สร้างโดยวิธีการซัพพอร์ตเวกเตอร์เรกเรชัน ซึ่งมีวิธีขั้นตอนการวิจัยดังนี้

#### 1. ซัพพอร์ตเวกเตอร์เรกเรชัน

ซัพพอร์ตเวกเตอร์เรกเรชัน (Support Vector Regression : SVR) [6] เป็นวิธีการหนึ่งที่ใช้พยากรณ์ค่าที่ได้รับความนิยมอย่างมาก และมีการใช้อย่างแพร่หลาย เนื่องจากค่าที่พยากรณ์นั้นมีความแม่นยำสูง เป็นการตัดแปลงมาจากวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยมีสมการดังนี้

$$f(x) = \langle w, x \rangle + b \quad (1)$$

โดยที่  $w \in X$ ,  $b \in \mathbb{R}$ ,  $\langle . . . \rangle$  คือการหาจุดทศนิยม,  $w$  คือน้ำหนักซัพพอร์ตเวกเตอร์,  $b$  คือค่าคงที่

โดยที่ข้อมูลที่ใช้ในการพยากรณ์จะอยู่ในรูป  $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset X \times \mathbb{R}$  โดยที่  $X$  คือขนาด

ชนิดของข้อมูลนำเข้า และ  $\mathbb{R}$  คือจำนวนจริง โดยเป้าหมายที่ต้องการคือหาค่าฟังก์ชัน  $f(x)$  โดยสามารถหาค่าได้โดยการใช้สมการของลากรานจ์ (Lagrange function) โดยจะมีการเพิ่มตัวแปรในสมการ ซึ่งเรียกว่า ตัวคูณลากรานจ์ (Lagrange multipliers) ซึ่งผลลัพธ์ที่ได้มีสมการดังสมการที่ 2

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (2)$$

โดยที่  $\alpha_i$  และ  $\alpha_i^*$  คือตัวคูณลากรานจ์

แต่ในกรณีที่ไม่สามารถทำการพยากรณ์ข้อมูลได้ใน 2 มิติจะมีการนำเคอร์เนลเข้ามาช่วย โดยมีสมการดังสมการที่ 3

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (3)$$

โดยที่  $k(x_i, x)$  คือค่าเคอร์เนล

#### 2. ARIMA

ARIMA (autoregressive integrated moving average) [7] เป็นเทคนิคที่เสนอโดยบ็อกซ์และเจนกินส์ในปี 1976 (Box and Jenkins, 1976) โดยเทคนิคแรกที่ถูกเสนอเรียกว่า ARMA โดยจะแบ่งออกเป็น 2 ส่วนคือ Autoregressive (AR) และ Moving average (MA) โดยข้อมูลที่จะนำมาสร้างตัวแบบ ARMA จำเป็นต้องมีลักษณะนิ่ง (stationary) ต่อมาจึงได้มีการพัฒนาเทคนิค ARIMA ขึ้นมาโดยพัฒนาให้เทคนิค ARMA สามารถทำงานได้กับข้อมูลที่มีลักษณะไม่นิ่งได้ ซึ่งเทคนิค ARIMA เป็นเทคนิคที่ได้รับความนิยมสูงเนื่องจากให้ผลการพยากรณ์ที่มีความแม่นยำสูง

**2.1 Autoregressive (AR)** เป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้า โดยกำหนดว่าค่าจริง  $Y_t$  ที่เวลา  $t$  ใด ๆ จะขึ้นกับค่าจริงที่เวลา  $t-1$ ,  $t-2$ ,  $t-3$ , ...,  $t-p$  โดยที่ค่า  $p$  คือค่าคาบเวลาที่ล่าช้า ซึ่งสามารถแสดงเป็นสมการได้ดังสมการที่ 4

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (4)$$

เมื่อ  $Y_t$  คือค่าจริงที่เวลา  $t$  ใด ๆ

$\theta_0$  คือค่าคงที่

$\theta_1, \theta_2, \dots, \theta_p$  คือพารามิเตอร์ถ่วงน้ำหนัก

$\varepsilon_t$  คือค่าคลาดเคลื่อนที่เวลา  $t$

## 2.2 Moving average (MA)

เป็นรูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าคลาดเคลื่อนก่อนหน้า โดยกำหนดว่าที่ค่าสังเกต  $Y_t$  ที่เวลา  $t$  ใด ๆ จะขึ้นกับค่าคลาดเคลื่อนที่เวลา  $t-1, t-2, t-3, \dots, t-q$  โดยที่ค่า  $q$  คือค่าความล่าช้า ซึ่งสามารถแสดงเป็นสมการได้ดังสมการที่ 5

$$Y_t = \theta_0 + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \quad (5)$$

เมื่อ  $Y_t$  คือค่าจริงที่เวลา  $t$  ใด ๆ

$\theta_0$  คือค่าคงที่

$\theta_1, \theta_2, \dots, \theta_q$  คือพารามิเตอร์ถ่วงน้ำหนัก

$\varepsilon_t$  คือค่าคลาดเคลื่อนที่เวลา  $t$

โดยเมื่อทำการรวม AR(p) และ MA(q) จะได้เทคนิคที่เรียกว่า ARMA(p,q) โดยมีสมการดังสมการที่ 6

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (6)$$

## 2.2.3 ตัวแบบ ARIMA

ตัวแบบ ARIMA หรือ ARIMA(p,d,q) เป็นเทคนิคที่พัฒนามาจาก ARMA โดยทำให้สามารถใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะนิ่งได้โดยทำการแปลงข้อมูลอนุกรมเวลาเดิม  $Y_t$  ซึ่งเป็นข้อมูลที่ไม่นิ่งให้เป็นอนุกรมเวลาใหม่  $Z_t$  ซึ่งเป็นข้อมูลที่นิ่ง โดยใช้การหาค่าผลต่างระหว่างค่าสังเกตในอนุกรมเวลาเดิม ดังสมการที่ 7

$$Z_t = \nabla^d Y_t \quad (7)$$

ถ้าให้  $d = 1$  จะได้  $Z_t = \nabla^1 Y_t = Y_t - Y_{t-1}$

$$\begin{aligned} \text{ถ้าให้ } d = 2 \text{ จะได้ } Z_t &= \nabla^2 Y_t = \nabla^1 Y_t - \nabla^1 Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \end{aligned}$$

## 3. เกณฑ์ที่ใช้ในการวัดความแม่นยำของตัวแบบ

### 3.1 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) เป็นวิธีที่การวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการดังสมการที่ 8 ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (8)$$

โดยที่ RMSE คือค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$n$  คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา  $t$  ใดๆ

$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใดๆ

### 3.2 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Deviation : MAD) หรือ (Mean Absolute Error : MAE) เป็นวิธีที่การวัดค่าความคลาดเคลื่อนที่นิยมอีกวิธีหนึ่ง ซึ่งวิธีนี้จะช่วยบอกถึงขนาดของความคลาดเคลื่อนรวมได้ โดยมีสมการดังสมการที่ 9 ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (9)$$

โดยที่ MAE คือค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$n$  คือจำนวนข้อมูลที่ใช้

$Y_t$  คือค่าจริงที่เวลา  $t$  ใดๆ



$\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์ที่เวลา  $t$  ใดๆ

#### 4. กรอบแนวคิดของงานวิจัย

แนวคิดของงานวิจัยนี้เกิดจากความต้องการที่จะเพิ่มความแม่นยำให้กับการพยากรณ์ โดยการใช้เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน และเพิ่มความแม่นยำโดยการใช้ค่าความคลาดเคลื่อน ซึ่งกรอบแนวคิดในการวิจัยมีขั้นตอนดังนี้

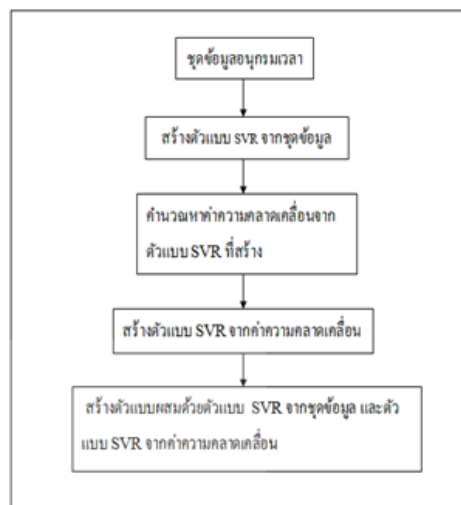
1. นำชุดข้อมูลอนุกรมเวลามาใช้ในการสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน
2. นำตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันที่ได้มาหาค่าความคลาดเคลื่อนโดยเทียบกับชุดข้อมูลอนุกรมเวลาที่ใช้ในการสร้างตัวแบบ
3. นำค่าความคลาดเคลื่อนที่ได้มาสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน
4. สร้างตัวแบบผสมโดยใช้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันจากชุดข้อมูล และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันจากค่าความคลาดเคลื่อน

ซึ่งสามารถแสดงกรอบแนวคิดได้ดังรูปที่ 1

จากรูปที่ 1 ได้แสดงรายละเอียดของกรอบแนวคิดและขั้นตอนในการวิจัย ซึ่งสามารถอธิบายรายละเอียดต่าง ๆ ได้ดังต่อไปนี้

##### 4.1 การแบ่งข้อมูลฝึกสอนและทดสอบ

ผู้วิจัยได้ทำการแบ่งชุดข้อมูลอนุกรมออกเป็น 2 ส่วน โดยในส่วนแรกเป็นข้อมูลฝึกสอน โดยใช้อัตราส่วน 70% ของข้อมูลทั้งหมด และส่วนที่สองเป็นข้อมูลทดสอบ โดยใช้อัตราส่วน 30% ของข้อมูลทั้งหมด เนื่องจากเป็นอัตราส่วนที่นิยมใช้ในงานสถิติ



รูปที่ 1 กรอบแนวคิดและขั้นตอนการวิจัย

##### 4.2 การสร้างตัวแบบด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชัน

ในการสร้างตัวแบบด้วยเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันนั้นจำเป็นต้องกำหนดสมการที่ใช้ เนื่องจากข้อมูลมีลักษณะเป็นอนุกรมเวลาที่ใช้ช่วยในการพยากรณ์คือค่า วัน เดือน ปี โดยกำหนดให้มีสมการดังสมการที่ 10

$$\hat{Y}_t = f(x) = \langle w, x \rangle + b \quad (10)$$

โดยที่ค่า  $\hat{Y}_t$  คือค่าที่ได้จากการพยากรณ์

$x$  คือข้อมูลนำเข้า ซึ่งแต่ละชุดข้อมูลประกอบด้วยข้อมูล วัน เดือน ปี และผลลัพธ์ ( $Y_t$ )

$b$  คือค่าคงที่

โดยในงานวิจัยนี้ผู้วิจัยได้กำหนดให้ใช้ค่าเคอร์เนลเป็นเรเดียลเบซิสฟังก์ชัน ค่าแกมมาเป็น  $1/\text{จำนวนมิติของข้อมูล}$  ค่าคอสเป็น 1 ค่าเอปซิลอนเป็น 0.1 เพื่อใช้ในการสร้างตัวแบบพยากรณ์ โดยใช้ข้อมูลฝึกสอนในการสร้างตัวแบบ และเมื่อนำค่าที่ได้จากการพยากรณ์และค่าจริงจากข้อมูลฝึกสอนมาทำการหาผลต่าง สิ่งที่ได้คือค่าความคลาดเคลื่อน ซึ่งหากค่าจริงมากกว่าค่าพยากรณ์ค่าความคลาดเคลื่อนจะมีค่าเป็นบวก แต่หากค่าจริงน้อยกว่าค่าพยากรณ์ค่าความคลาดเคลื่อนจะมีค่าเป็นลบดังสมการที่

$$\text{error} = |Y_t - \hat{Y}_t| \quad (11)$$

#### 4.3 การนำค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำ

การนำค่าความคลาดเคลื่อนมาช่วยในการเพิ่มความแม่นยำนั้นทำได้โดยการนำค่าความคลาดเคลื่อนมาสร้างตัวแบบพยากรณ์โดยค่าความคลาดเคลื่อนที่ได้ยังคงมีลักษณะเป็นอนุกรมเวลาอยู่ โดยมีสมการดังสมการที่ 12

$$\hat{Y}_{\text{error}} = f(x) = \langle w, x \rangle + b \quad (12)$$

โดยที่ค่า  $\hat{Y}_{\text{error}}$  คือค่าที่ได้จากการพยากรณ์

$x$  คือ ข้อมูลนำเข้า ซึ่งแต่ละชุดข้อมูลประกอบด้วยข้อมูล วัน เดือน ปี และค่าความคลาดเคลื่อน

$b$  คือค่าคงที่

โดยค่าอื่น ๆ ที่ใช้ในการสร้างตัวแบบเป็นค่าเดียวกับที่ใช้สร้างตัวแบบด้วยเทคนิคซัพพอร์ตเวกเตอร์เกรสชันโดยผลลัพธ์ที่ได้นั้นจะเป็นตัวแบบที่ได้จากค่าความคลาดเคลื่อน ซึ่งจากสมมติฐานที่ว่าค่าความคลาดเคลื่อนคือค่าความสัมพันธ์ที่ขาดหายไป ดังนั้นเพื่อเพิ่มความแม่นยำให้ตัวแบบที่สร้างด้วยเทคนิคซัพพอร์ตเวกเตอร์เกรสชันดั้งเดิมจึงได้นำตัวแบบที่ได้จากค่าความคลาดเคลื่อนมาช่วยในการพยากรณ์ ซึ่งตัวแบบใหม่ที่ได้แสดงดังสมการที่ 13 โดยเรียกตัวแบบนี้ว่าตัวแบบผสม (combine)

$$\hat{Y}_p = \hat{Y}_t + \hat{Y}_{\text{error}} \quad (13)$$

โดยที่ค่า  $\hat{Y}_p$  คือค่าที่ได้จากการพยากรณ์ด้วยตัวแบบผสม

#### 4.4 การทดสอบความแม่นยำของตัวแบบ

ในการทดสอบความแม่นยำของตัวแบบนั้นข้อมูลที่ใช้คือข้อมูลทดสอบ โดยใช้ตัวแบบผสมในการพยากรณ์ค่า ซึ่งวัดค่าโดยใช้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย โดยทำการเปรียบเทียบเทคนิคทั้งหมด 3 ชนิดคือ เทคนิค

ซัพพอร์ตเวกเตอร์เกรสชันดั้งเดิม เทคนิคซัพพอร์ตเวกเตอร์เกรสชันที่เพิ่มความแม่นยำด้วยค่าความคลาดเคลื่อน และเทคนิค ARIMA

#### 5. ข้อมูลที่ใช้

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่อยู่ในรูปแบบของอนุกรมเวลา โดยเป็นข้อมูลจาก Data Market (<https://datamarket.com/data/>) และ Duke University ([http://www2.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www2.stat.duke.edu/~mw/ts_data_sets.html)) ซึ่งข้อมูลที่น่ามาใช้มีทั้งหมด 5 ชุด ประกอบไปด้วยชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพีชเชอร์ (TEMP) ข้อมูลปริมาณการผลิตน้ำนมของวัวในแต่ละเดือน (MILK) ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองดาร์วิน (SLP) ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอา (CO2) และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและดาร์วิน (SOI)

#### ผลการทดลอง

จากตารางที่ 1 และ 2 พบว่าเมื่อนำค่าความคลาดเคลื่อนมาช่วยในการทำนาย ทำให้โมเดลเข้าใกล้ค่าของข้อมูลฝึกสอนมากขึ้น โดยทุกข้อมูลให้ผลลัพธ์ที่ดีขึ้นทั้งหมด จากตารางที่ 3 เป็นการเปรียบเทียบโดยใช้ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งผลลัพธ์ที่ได้คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์เกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล และเทคนิค ARIMA ให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์เกรสชัน

ตารางที่ 1 การเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลฝึกสอน

data	SVR	Proposed method	% Improvement
temp	5.15720	5.11310	0.855115179
milk	22.4174	21.4962	4.109307948
SLP	1.0115	1.0099	0.158180919
co2	0.4751	0.434	8.650810356
SOI	1.5845	1.5772	0.460713159

ตารางที่ 2 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลฝึกสอน

data	SVR	Proposed method	% Improvement
temp	3.96800	3.90040	1.703629032
milk	16.9748	15.1543	10.72472135
SLP	0.7916	0.7864	0.656897423
co2	0.3925	0.3262	16.89171975
SOI	1.2274	1.2123	1.23024279

พบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 4 ชุดข้อมูล และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพีซเซอร์ให้ผลดีขึ้น 1.34% ข้อมูลปริมาณการผลิตน้ำมันของวัวในแต่ละเดือนให้ผลดีขึ้น 17.15% ข้อมูล ค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 3.86% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 4.31% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและคาร์วินให้ผลแย่งลง 1.99% จากตารางที่ 4 เป็นการเปรียบเทียบโดยใช้ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งผลลัพธ์ที่ได้ คือ เทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 3 ชุดข้อมูล เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล และเทคนิค ARIMA ให้ผลลัพธ์ที่ดีที่สุด 1 ชุด ซึ่งเมื่อทำการเทียบเฉพาะเทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าเทคนิคใหม่ให้ผลลัพธ์ที่ดีที่สุด 4 ชุดข้อมูล และ

ตารางที่ 3 แสดงการเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากข้อมูลทดสอบ

data	SVR	Proposed method	ARIMA	% Improvement
temp	10.56466	10.42317	15.31096	1.33928
milk	60.28028	49.94115	85.42048	17.15176
SLP	1.921235	1.846996	2.548502	3.86413
co2	6.510226	6.229361	2.930608	4.31421
SOI	1.918531	1.956628	1.984341	-1.98574

ตารางที่ 4 แสดงการเปรียบเทียบค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากข้อมูลทดสอบ

data	SVR	Proposed method	ARIMA	% Improvement
temp	9.07392	8.96033	12.73791	1.25183
milk	48.35438	39.37634	66.36006	18.56717
SLP	1.57495	1.48809	2.155819	5.51510
co2	5.58294	5.27019	2.42601	5.60189
SOI	1.45858	1.49647	1.530834	-2.59773

เทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันให้ผลลัพธ์ที่ดีที่สุด 1 ชุดข้อมูล เมื่อทำการเทียบผลลัพธ์เทคนิคใหม่และเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันพบว่าชุดข้อมูลอุณหภูมิรายวันของแม่น้ำพีซเซอร์ให้ผลดีขึ้น 1.25% ข้อมูลปริมาณการผลิตน้ำมันของวัวในแต่ละเดือนให้ผลดีขึ้น 18.57% ข้อมูลค่าความดันที่ระดับน้ำทะเลที่เมืองคาร์วินให้ผลดีขึ้น 5.52% ข้อมูลปริมาณคาร์บอนไดออกไซด์ที่ภูเขาไฟเมานาโลอาให้ผลดีขึ้น 5.60% และข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและคาร์วินให้ผลแย่งลง 2.60%

#### สรุปผลการทดลอง

จากการทดลองพบว่าเทคนิคการเพิ่มประสิทธิภาพให้กับเทคนิคซัพพอร์ตเวกเตอร์รีเกรสชันดั้งเดิม โดยการนำค่าความคลาดเคลื่อนมาช่วยในการสร้างตัวแบบสามารถทำได้ เนื่องจากการนำค่าความคลาดเคลื่อนมาใช้ในการสร้างโมเดลจะทำให้โมเดลที่ได้นั้นเข้าใจตัวส่วนของข้อมูลฝึกสอนมากขึ้น ซึ่งทำให้ผลลัพธ์ที่ได้เมื่อนำไปใช้กับข้อมูลทดสอบได้ผลลัพธ์ที่ดีขึ้น แต่ในกรณีของข้อมูลค่าดัชนีที่คำนวณจากค่าความกดอากาศที่แตกต่างกันระหว่างจุด 2 จุดในตาฮีตีและคาร์วินให้ผลลัพธ์แย่งลง อาจเนื่องมาจากเกิดการโอเวอร์ฟิตติ้ง (Overfitting) ซึ่งเป็นเหตุการณ์ที่โมเดลสามารถให้ผลลัพธ์ได้ดีมากในข้อมูลฝึกสอนแต่ให้ผลลัพธ์ที่แย่งกับข้อมูลทดสอบ

#### กิตติกรรมประกาศ

ขอขอบพระคุณรองศาสตราจารย์ ดร. นิตยา เกิดประสพ อาจารย์ที่ปรึกษา และรองศาสตราจารย์ ดร. กิตติศักดิ์

- เกิดประสพ ที่คอยช่วยให้คำแนะนำและคำปรึกษาในการ  
ทำวิจัย
- เอกสารอ้างอิง**
- [1] Zhang, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;50:159-175.
- [2] Chen, Kuan-Yu, and Cheng-Hua Wang. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*. 2007;32.1:254-264.
- [3] Wang, Yuanyuan, et al. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. *Energy Policy*. 2012;48:284-294.
- [4] Khandelwal, Ina, Rathadip Adhikari, and Ghanshyam Verma. Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition. *Procedia Computer Science*. 2015;48:173-179.
- [5] de Oliveira, João FL, and Teresa B. Ludermir. A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*. 2016;180:27-34.
- [6] Smola, Alex J., and Bernhard Schölkopf. A tutorial on support vector regression. *Statistic and computing*. 2004;14.3:199-222
- [7] BOX, George EP; JENKINS, Gwilym M. Time series analysis, control, and forecasting. San Francisco, CA: Holden Day. 1976;3226.3228:10.
- [8] DataMarket [Internet]. Iceland:DataMarket; 2008 [cited 2016 april 17]. Available from: <https://datamarket.com/data/>
- [9] SOME TIME SERIES DATASETS [Internet]. America: Duke University [cited 2016 april 17]. Available from: [http://www2.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www2.stat.duke.edu/~mw/ts_data_sets.html)

## ประวัติผู้เขียน

นายธีรวัช แก้ววิจิตร เกิดเมื่อวันที่ 3 เมษายน พ.ศ. 2536 อำเภอเมืองร้อยเอ็ด จังหวัดร้อยเอ็ด เริ่มศึกษาชั้นประถมศึกษาที่โรงเรียนอนุบาลนครราชสีมา อำเภอเมืองนครราชสีมา จังหวัดนครราชสีมา จากนั้นได้ทำการศึกษาต่อชั้นมัธยมศึกษาที่โรงเรียนราชสีมาวิทยาลัย อำเภอเมืองนครราชสีมา จังหวัดนครราชสีมา และสำเร็จปริญญาตรี สาขาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี จังหวัดนครราชสีมา เมื่อปี พ.ศ. 2558

ในปี 2558 ได้ศึกษาต่อในระดับปริญญาโท สาขาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี

ในระหว่างการศึกษาได้รับความอนุเคราะห์เป็นอย่างดีจากอาจารย์ที่ปรึกษา และได้รับการตีพิมพ์เผยแพร่บทความวิชาการในงาน South East Asian Technical University Consortium (SEATUC) Symposium ที่มหาวิทยาลัยเทคโนโลยีสุรนารี ในหัวข้อ THE STUDY OF THE BEST PROBABILITY DISTRIBUTION TO PREDICT A GOLD PRICE WITH LINEAR REGRESSION ในปี พ.ศ. 2558

มหาวิทยาลัยเทคโนโลยีสุรนารี