

รายงานโครงการวิศวกรรมไฟฟ้า วิชา 2102499 ปีการศึกษา 2562

การเปรียบเทียบวิธีการพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ในระยะสั้นมาก

A comparison of intraday solar power forecasting methods

นายชฎานนท์ โพราวนานนท์ เลขประจำตัว 5930084921

นายสรารุต พรานนท์สกิตย์ เลขประจำตัว 5930515021

อาจารย์ที่ปรึกษา ผศ.ดร. จิตโกมุท ส่งคิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

การพยากรณ์กำลังผลิตไฟฟ้าระยะวาระใน 1 วันมีความสำคัญในการการบริหารจัดการระบบก้าวเก็บพลังงานสำรองพร้อมจ่ายให้มีความคุ้มค่าที่สุดในเชิงเศรษฐศาสตร์ และยังเพิ่มความมั่นคงของระบบโครงข่ายไฟฟ้า โครงงานนี้มีจุดประสงค์จะพัฒนาแบบจำลองการพยากรณ์ระยะ 4 ชั่วโมงล่วงหน้าโดยคำนวนทุกๆ 30 นาที ในกรอบอุปแบบนั้นจะพิจารณา 2 แนวทาง คือการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์ และการพยากรณ์กำลังผลิตไฟฟ้าโดยตรง โดยมีการแบ่งแบบจำลองออกตามช่วงเวลาใน 1 วัน กล่าวคือมีแบบจำลองสำหรับพยากรณ์ในช่วงเช้า กลางวัน และเย็น แบบจำลองที่พัฒนาขึ้นมาในนี้ใช้วิธีชั้พพร์ตเวกเตอร์ด้วยตัวแปรแบบจำลองป่าสุ่ม จากนั้นเบรย์บเทียนสมรรถนะกับแบบจำลองอื่นที่ใช้วิธีการลดถอยเชิงเส้น การลดถอยลดตัวแปรแบบปรับสันโค้ง (MARS) และโครงข่ายประสาทเทียม การทดลองจัดทำขึ้นโดยใช้ข้อมูลจาก 2 แหล่งคือ ข้อมูลที่ได้จากเครื่องมือวัด ณ ศึกษาภารมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และโรงไฟฟ้าพลังงานแสงอาทิตย์ในภาคกลางจำนวนหนึ่งโรงในช่วงระหว่างปี พ.ศ. 2560–2561 ผลการทดลองพบว่าการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์ด้วยแบบจำลองชั้พพร์ตเวกเตอร์ด้วยตัวแปรแบบจำลองป่าสุ่มมีค่าความผิดพลาด NRMSE ในกรอบ 30 นาที ล่วงหน้าเท่ากับ 7.14% และ 6.93% ตามลำดับ ส่วนกรณีพยากรณ์กำลังผลิตไฟฟ้าโดยตรง พบว่าค่าความผิดพลาด NRMSE มีค่าเท่ากับ 6.40% และ 6.02% ตามลำดับ ผลลัพธ์ที่เห็นว่าการพยากรณ์กำลังผลิตไฟฟ้าโดยตรงให้สมรรถนะที่ดีกว่าการพยากรณ์ผ่านความเข้มแสงอาทิตย์ และเทคนิคการพยากรณ์ที่มีสมรรถนะดีที่สุดคือการพยากรณ์กำลังผลิตไฟฟ้าโดยตรงด้วยแบบจำลองป่าสุ่ม

คำสำคัญ: การพยากรณ์ความเข้มแสงอาทิตย์, การพยากรณ์ในระยะระหว่างวัน, กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์, แบบจำลองป่าสุ่ม, แบบจำลองชัพพร์ตเวกเตอร์

Abstract

Intraday solar power forecasting is crucial to ensuring power continuity and economical dispatch in PV systems. This study is focused on an 4-hour ahead solar power forecasting in a resolution of 30 min. We present 2 approaches, namely, indirect approach and direct approach which are forecasting solar power via irradiance prediction and PV conversion and forecasting solar power directly respectively. The proposed models for the two approaches are split and responsible for providing predictions at three different times of the days: morning, midday, and evening. In this work, we develop SVR and RF models and compare the performance to baseline models which are linear regression, MARs and ANN models. Every model are designed to produce intraday solar power forecasts using ground data which were collected from two measurement stations in central region of Thailand from 2017-2018. The result shows that the direct approach yielded better performance, achieving NRMSE of 7.14% and 6.93%, compared to the indirect approach which achieved NRMSE of 6.40% and 6.02% on SVR and RF model respectively. The best model in terms of forecast accuracy is achieved by the random forest model that directly predicts solar power.

Keywords: solar power forecasting, intraday forecast, Photovoltaic system, random forest, support vector regression

สารบัญ

1	บทนำ	7
2	ภาพรวมของโครงการ	8
2.1	วัตถุประสงค์	8
2.2	ขอบเขต	9
2.3	ผลลัพธ์ที่คาดหวัง	9
3	หลักการและทฤษฎีที่เกี่ยวข้อง	9
3.1	การคัดเลือกคุณลักษณะ	10
3.1.1	สถิติพื้นฐาน	10
3.1.2	วิธีการลดถอยเชิงเส้นแบบขั้นตอน	11
3.2	แบบจำลองทัองฟ้าใส	11
3.2.1	การตรวจจับวันท้องฟ้าใสจากข้อมูลวัด	12
3.3	เทคนิคการประมาณ	13
3.3.1	Linear regression	13
3.3.2	Multivariate adaptive regression splines (MARS)	14
3.3.3	Support Vector Regression	14
3.3.4	Random Forest	16
3.4	ตัวชี้วัดประสิทธิภาพของการพยากรณ์	17
4	การจัดเตรียมข้อมูล	18
4.1	ที่มาของข้อมูล	18
4.2	การประมวลข้อมูลเบื้องต้น	18
4.2.1	การจัดการกับข้อมูลสูญหาย	18
4.2.2	การลดอัตราสุ่มข้อมูล	19
4.2.3	การจัดการกับข้อมูลที่ผิดพลาด	19
4.3	การวิเคราะห์ลักษณะของข้อมูลเบื้องต้น	20
5	แบบจำลองการพยากรณ์	21
5.1	แบบจำลองพยากรณ์ความเข้มแสงอาทิตย์	22
5.2	แบบจำลองสำหรับแปลงความเข้มแสงอาทิตย์เป็นกำลังไฟฟ้า	24
5.3	แบบจำลองพยากรณ์กำลังผลิตไฟฟ้าโดยตรง	25
6	ผลลัพธ์ของโครงการ	26
6.1	การคัดเลือกคุณลักษณะ	27
6.2	การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์	27
6.3	การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง	30
6.4	ตัวอย่างผลลัพธ์การพยากรณ์	31
6.5	การเปรียบเทียบความซับซ้อนในการคำนวณ	31
7	การวิเคราะห์และวิจารณ์ผลลัพธ์ของโครงการ	33
8	บทสรุป	35
9	กิตติกรรมประกาศ	36
	เอกสารอ้างอิง	39

10 ภาคผนวก	40
10.1 ผลการปรับค่าพารามิเตอร์ของแบบจำลอง Support Vector Regression	40
10.2 ผลการปรับค่าพารามิเตอร์ของแบบจำลอง Random Forest	42
10.3 ชุดโปรแกรมคำสั่ง	43

สารบัญรูป

1	สัดส่วนของงานวิจัยจำแนกตามเทคนิคที่ใช้ในการพยากรณ์ [AOE ^{+16]}	8
2	ค่าความเข้มแสงอาทิตย์ท้องฟ้าใส่อ้างอิงในแต่ละการวนซ้ำโดยใช้ข้อมูลจากโรงไฟฟ้าภาคกลางและตีกวิศวกรรมไฟฟ้า	13
3	ตัวอย่างค่าความเข้มแสงของวันที่ถูกเลือกเป็นวันท้องฟ้าใส่ด้วยขั้นตอนการตรวจสอบนี้	13
4	รูปแสดงฟังก์ชันสูญเสียแบบ ε -incentive ของ linear SVR [SS04]	15
5	ตัวอย่างการแบบจำลองต้นไม้สำหรับปริมาณตัวแปรต้น 2 มิติ	16
6	ตัวอย่างข้อมูลกำลังผลิตไฟฟ้าและค่าความเข้มแสงในหนึ่งวันจากโรงไฟฟ้าภาคกลางและตีกวิศวกรรมไฟฟ้า	19
7	ตัวอย่างข้อมูลที่ผิดพลาดของกำลังผลิตไฟฟ้าและความเข้มแสงอาทิตย์จากตีกวิศวกรรมไฟฟ้าและโรงไฟฟ้าภาคกลาง	19
8	ความสัมพันธ์ระหว่างกำลังผลิตไฟฟ้าและความเข้มแสงอาทิตย์ของข้อมูลจากโรงไฟฟ้าภาคกลางและตีกวิศวกรรมไฟฟ้า	20
9	การกระจายตัวของความเข้มแสงอาทิตย์ในช่วงเวลาต่างๆ	20
10	ตัวอย่างรูปแบบการพยากรณ์ทางตรงของแบบจำลอง Random forest ณ วันท้องฟ้าใส	21
11	ตัวอย่างรูปแบบการพยากรณ์ทางตรงของแบบจำลอง Random forest ณ วันสภาพอากาศทั่วไป	22
12	ขั้นตอนการพยากรณ์กำลังผลิตไฟฟ้า	22
13	แบบจำลองพยากรณ์ k-step ของความเข้มแสงอาทิตย์แยกตามช่วงเวลา	24
14	แบบจำลองพยากรณ์ k-step ของกำลังผลิตไฟฟ้าแยกตามช่วงเวลา	26
15	RMSE ของความเข้มแสงอาทิตย์ในแต่ละระยะการพยากรณ์	28
16	MBE ของความเข้มแสงอาทิตย์ในแต่ละระยะการพยากรณ์	28
17	RMSE ของความเข้มแสงอาทิตย์ในการพยากรณ์ที่แต่ละจุดเวลา	28
18	MBE ของความเข้มแสงอาทิตย์ในการพยากรณ์ที่แต่ละจุดเวลา	29
19	NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์ผ่านความเข้มแสงอาทิตย์	29
20	NMBE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์ผ่านความเข้มแสงอาทิตย์	29
21	NRMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์ผ่านความเข้มแสงอาทิตย์ที่แต่ละจุดเวลา	30
22	NMBE ของกำลังผลิตไฟฟ้าในการพยากรณ์ผ่านความเข้มแสงอาทิตย์ที่แต่ละจุดเวลา	30
23	NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์โดยตรง	30
24	NMBE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์โดยตรง	31
25	NRMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์โดยตรงที่แต่ละจุดเวลา	31
26	NMBE ของกำลังผลิตไฟฟ้าในการพยากรณ์โดยตรงที่แต่ละจุดเวลา	31
27	ตัวอย่างผลลัพธ์การพยากรณ์โดยตรงในระยะ 30 นาทีด้วยแบบจำลอง Random Forest, SVR, MARs	32

สารบัญตาราง

1	สัญลักษณ์และตัวแปร	9
2	พารามิเตอร์การกระจายของความเข้มแสงอาทิตย์แต่ละช่วงเวลา	21
3	การคัดเลือกคุณลักษณะสำหรับพยากรณ์ $I(t+1)$	27
4	ความซับซ้อนในการคำนวณของวิธีพยากรณ์ด้วยแบบจำลองต่างๆ	33
5	NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามระยะพยากรณ์ ณ ตีกวิศวกรรมไฟฟ้า	35
6	NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามระยะพยากรณ์ ณ โรงไฟฟ้าภาคกลาง	35
7	NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามเวลา ณ ตีกวิศวกรรมไฟฟ้า	36
8	NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามเวลา ณ โรงไฟฟ้าภาคกลาง	37
9	สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า C	41
10	สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า γ	41
11	สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า ε	41
12	สมรรถนะของแบบจำลอง RF เมื่อปรับค่า $n_{min_samples_split}$ และ $n_{min_samples_leaf}$	42
13	สมรรถนะของแบบจำลอง RF เมื่อปรับค่า d	43
14	สมรรถนะของแบบจำลอง RF เมื่อปรับค่า m	43

1 บทนำ

ในปัจจุบันประเทศไทยมีนโยบายที่จะลดปริมาณการใช้พลังงานจากก๊าซธรรมชาติ และส่งเสริมการผลิตไฟฟ้าจากพลังงานทางเลือกอาทิ พลังงานจากเซลล์แสงอาทิตย์ ตามแผนพัฒนาพลังงานทดแทนและพลังงานทางเลือก ในปี พ.ศ.2558 (AEDP 2015) โดยกรมพัฒนาพลังงานทดแทนและอนุรักษ์พลังงานมีนโยบายที่จะเพิ่มสัดส่วนการใช้พลังงานทดแทนภายในประเทศ และมาตรการประเมินภายใต้สิ้นปี พ.ศ.2579 สัดส่วนของพลังงานไฟฟ้าที่ผลิตได้จากพลังงานแสงอาทิตย์จะคิดเป็นสัดส่วนถึง 30.5 % ของพลังงานในกลุ่มพลังงานทดแทนทั้งหมด ซึ่งสอดคล้องกับการที่ต้นทุนในการลงทุนติดตั้งระบบผลิตกำลังไฟฟ้าจากเซลล์แสงอาทิตย์ที่มีแนวโน้มที่ลดลงอย่างต่อเนื่อง จึงทำให้การผลิตไฟฟ้าจากพลังงานแสงอาทิตย์ เข้ามายึด主导 สำหรับผู้ประกอบการ อย่างไรก็ตามค่าความเข้มแสงอาทิตย์ที่เป็นตัวแปรสำคัญในการผลิต กำลังไฟฟ้าจากพลังงานแสงอาทิตย์ มีความแปรปรวนซึ่งขึ้นอยู่กับปัจจัยสำคัญ คือ สภาพภูมิอากาศ ทำให้ต้องคำนึงถึงการใช้พลังงานทดแทนที่มีประสิทธิภาพ ในการบริหารจัดการกำลังผลิตไฟฟ้าให้สอดคล้องกับความต้องการของผู้ใช้ในแต่ละช่วงเวลา จำกปัญหาข้างต้น การพัฒนาประสิทธิภาพของการพยากรณ์กำลังไฟฟ้าที่ผลิตได้จากเซลล์แสงอาทิตย์จึงมีความสำคัญ ทั้งในด้านการรักษาความมั่นคงของระบบ ตลอดจนลดต้นทุนอันเนื่องมาจากการสำรองกำลังผลิตไฟฟ้าโดยทั่วไป การพยากรณ์กำลังไฟฟ้าที่ผลิตได้จากเซลล์แสงอาทิตย์ สามารถแบ่งได้เป็น 4 ประเภท

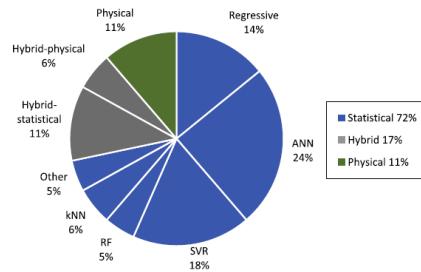
1. การพยากรณ์ในระยะสั้นมาก (very short-term forecast หรือ intra-day) เป็นการพยากรณ์ในระยะ 1-6 ชั่วโมงล่วงหน้า มีประโยชน์ในการการรักษาความมั่นคงของระบบโครงข่ายไฟฟ้า รวมไปถึงถึงการใช้งานระบบกักเก็บพลังงานสำรองพร้อมจ่ายทันทีและเพื่อบริหารจัด การพลังงานไฟฟ้า (จากพลังงานหมุนเวียน) ส่วนเกินในบางช่วงเวลา
2. การพยากรณ์ในระยะสั้น (short-term forecast หรือ day-ahead) เป็นการพยากรณ์ในระยะ 1-3 วันล่วงหน้า มีประโยชน์ในการบริหารจัดการ ความต้องการใช้ไฟฟ้า เพื่อเตรียมการสั่งเดินเครื่องในโรงงานที่สามารถควบคุมกำลังผลิตไฟฟ้าได้ เพื่อให้กำลังผลิตไฟฟ้า ในแต่ละช่วงเวลาเหมาะสม และ เป็นไปตามกลไกตลาดซื้อขายไฟฟ้าไว้ล่วงหน้า ทันทีเพื่อให้ต้นทุนการจัดทำไฟฟ้าโดยรวมของพื้นที่มีความคุ้มค่าที่สุดในเชิงเศรษฐศาสตร์ และการใช้งานเชือเพลิงแต่ละชนิดเป็นไปอย่างเพียงพอและมีประสิทธิภาพ
3. การพยากรณ์ในระยะกลาง (medium-term forecast) เป็นการพยากรณ์ในระยะ 1 สัปดาห์-1 เดือนล่วงหน้า มีประโยชน์ในการวางแผนกำหนดบำรุงรักษาโดยการทำนายความพร้อมใช้งานของกำลังผลิตไฟฟ้าในอนาคต
4. การพยากรณ์ในระยะยาว (long-term forecast) เป็นการพยากรณ์ในระยะ 1 เดือน-1 ปีล่วงหน้า มีประโยชน์ในการบริหารจัดการระบบผลิตกำลังไฟฟ้าในระยะยาว เช่น การสร้างโรงงานผลิตไฟฟ้าแห่งใหม่ หรือการจัดทำแผนประมาณการกำลังไฟฟ้าที่จะผลิตได้ในอนาคต

การศึกษาและพัฒนาความแม่นยำของการพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์เป็นที่น่าสนใจในวงกว้าง โดยแบ่งออกได้เป็นหลักๆ 2 วิธี คือ วิธีการพยากรณ์ทางตรงและวิธีการพยากรณ์ทางอ้อม วิธีการพยากรณ์ทางอ้อมจะเริ่มจากการพยากรณ์ค่าความเข้มแสงอาทิตย์ก่อน จากนั้นใช้แบบจำลองของระบบผลิตไฟฟ้าในการแปลงค่าความเข้มแสงอาทิตย์จากการพยากรณ์ไปเป็นค่ากำลังไฟฟ้าที่คาดว่าจะผลิตได้ในขณะที่ วิธีการพยากรณ์ทางตรงเป็นการพยากรณ์ค่ากำลังผลิตไฟฟ้าที่ได้จากระบบผลิตไฟฟ้าโดยตรง ทั้งนี้หากหลักฐานวิจัยในอดีตจะให้ความสนใจ เฉพาะการพยากรณ์ค่าความเข้มแสงอาทิตย์ เนื่องจากเป็นส่วนที่ยากในการพยากรณ์ และมีการประยุกต์ใช้ที่หลากหลายบนอุปกรณ์ ภายนอกและการพยากรณ์ กำลังผลิตไฟฟ้า อย่างไรก็ตามทั้งการพยากรณ์ทางตรงและทางอ้อมต่างมีข้อตอนวิธีการและเทคนิคที่คล้ายคลึงกัน [AOE⁺16]

หลักฐานวิจัยในอดีต ได้นำเสนอวิธีที่หลักหลายในการพยากรณ์แสงอาทิตย์และกำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ [AOE⁺16, IPC13] โดยสามารถแบ่งวิธีการได้ออกเป็นหลักๆ 3 ประเภท คือ 1) วิธีการทางสถิติ (statistical methods) 2) วิธีการทางกายภาพ (physical methods) 3) วิธีการแบบผสมผสาน (hybrid methods) [AOE⁺16] วิธีการทางสถิติเป็นการใช้ข้อมูลในอดีตที่วัดได้ เช่น ข้อมูลสภาพอากาศ ค่า กำลังผลิตไฟฟ้าในอดีต ใน การพยากรณ์ โดยไม่จำเป็นต้องใช้ข้อมูลความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม ตัวอย่างที่นิยม เช่น วิธีการในกลุ่มการเรียนรู้ด้วยเครื่อง วิธีการทางกายภาพ (physical methods) เป็นวิธีที่อาศัยการคำนวนโดยใช้สมการความสัมพันธ์ทางพิสิกส์ระหว่าง ตัวแปรต้นและตัวแปรตาม โดยวิธีที่เป็นที่นิยมได้แก่ การพยากรณ์โดยการคำนวน ค่าพยากรณ์สภาพอากาศเชิงเลข (Numerical Weather Prediction) และ การพยากรณ์โดยใช้วิธีข้างต้นร่วมกันเรียกว่าวิธีการแบบผสมผสาน

แผนภูมิดังรูปที่ 1 แสดงให้เห็นว่าในอดีตมีการนำเสนอวิธีที่หลักหลายในการพยากรณ์ค่ากำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์อาทิ

- 1) Regressive methods 2) Artificial neural network (ANN) 3) Support vector regression (SVR) 4) k-Nearest neighbors (k-NN)
- 5) Random forest (RF) ทั้งนี้ในโครงงานฉบับนี้จะเลือกพิจารณาการพยากรณ์ในระยะสั้นมาก (very short-term forecast หรือ intra-day) เพื่อประโยชน์ในการบริหารและรักษาความมั่นคงในระบบโครงข่ายไฟฟ้า โดยวิธีการพยากรณ์ที่เป็นที่นิยมแพร่หลายในระยะนี้ คือ วิธีการทางสถิติ (statistical methods) ซึ่งมีหลักหลายวิธี ตั้งแต่ การใช้แบบจำลองเชิงเส้น ไปจนถึงวิธีที่มีความซับซ้อนสูง เช่น โครงข่ายประสาทเทียม (neural network) ทั้งนี้การใช้แบบจำลองเชิงเส้นซึ่งที่มีความซับซ้อนต่ำในการพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์มีประสิทธิภาพ ต่ำ [SSY] ซึ่งอาจเกิดมาจากการปัจจัยที่มีความสัมพันธ์แบบไม่เป็นเชิงเส้นกับค่ากำลังผลิตไฟฟ้า ด้วยปัญหาข้างต้นในปัจจุบันวิธีการพยากรณ์ในกลุ่มการเรียนรู้ด้วยเครื่องปัจจุบัน แนะนำให้ใช้ในการพยากรณ์กำลังผลิตไฟฟ้าจากพลังงานแสงอาทิตย์อาทิ ANN, SVR, RF, KNN ซึ่งมีงานวิจัยในอดีตที่เกี่ยวข้องดังนี้



รูป 1: สัดส่วนของงานวิจัยจำแนกตามเทคนิคที่ใช้ในการพยากรณ์ [AOE⁺ 16]

M.Rana [MIG16] เปรียบเทียบการใช้วิธี SVR และวิธี NN-ensemble ในการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 5 ถึง 60 นาที โดยใช้มูลกำลังผลิตไฟฟ้าในอดีตเพียงอย่างเดียว และการใช้ข้อมูลกำลังผลิตไฟฟ้าในอดีตร่วมกับข้อมูลสภาพอากาศ นอกจากนี้ยังมีการประยุกต์ใช้วิธี Correlation-based Feature selection (CFS) ในการคัดเลือกคุณลักษณะของข้อมูลที่ใช้ในการทำนาย จากผลลัพธ์พบว่าในการพยากรณ์ระยะใกล้ NN-ensemble และ SVR ให้ความแม่นยำในการพยากรณ์ใกล้เคียงกัน ส่วนในระยะไกลออกไป NN-ensemble จะพยากรณ์แม่นยิ่งกว่า SVR

S.Vagropoulos [VKS16] ใช้วิธี SARIMA ในการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 1 ชั่วโมงโดยใช้ข้อมูล ความเข้มรังสีแสงอาทิตย์ในอดีต ร่วมกับข้อมูลสภาพอากาศ ผลลัพธ์การพยากรณ์มีค่า NRMSE เท่ากับ 8.12% โดยเป็นค่าที่ถูกปรับเทียบด้วยค่ากำลังติดตั้งขนาด 0.15-MW

M.Bouzerdoum [BMP13] เปรียบเทียบการใช้วิธี seasonal auto-regressive integrated moving average (SARIMA) , SVR และ การ ผสมผสานของ SARIMA และ SVR เพื่อพยากรณ์กำลังผลิตไฟฟ้าในระยะ 1 ชั่วโมง โดยใช้ข้อมูลจากกำลังผลิตไฟฟ้าในอดีตและค่าอุณหภูมิ พบว่า SARIMA-SVR ให้ผลลัพธ์ที่แม่นยำที่สุดมีค่า NRMSE เท่ากับ 9.40 % โดยเป็นค่าที่ถูกปรับเทียบด้วยค่ากำลังติดตั้งขนาด 20-kW

R. Xu และคณะ [XCS12] ประยุกต์ใช้วิธี SVR ร่วมกับการวิเคราะห์ความคล้ายกันของแต่ละวัน ในการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 2 ชั่วโมง โดยใช้ข้อมูลจากกำลังผลิตไฟฟ้าและค่าแสงอาทิตย์ในอดีตและค่าอุณหภูมิ พบว่าผลลัพธ์การพยากรณ์ที่ได้มีค่า NRMSE เท่ากับ 9.34% ซึ่งมีความแม่นยำสูงกว่าวิธี NN ที่ได้ค่า NRMSE เท่ากับ 13.19% โดยเป็นค่าที่ถูกปรับเทียบด้วยค่ากำลังติดตั้งขนาด 500-kW

W. Björk และคณะ [BEO16] เปรียบเทียบการใช้วิธี SVR , KNN และ combined weight SVM-kNN ในการพยากรณ์ในระยะ 1 ชั่วโมงและ 6 ชั่วโมงโดยใช้ข้อมูลกำลังผลิตไฟฟ้าในอดีต ข้อมูลสภาพอากาศ เวลาที่พยากรณ์ และค่าดัชนีไฟฟ้า พบว่าผลลัพธ์ที่ดีที่สุดของทั้งสองระยะการพยากรณ์มาจากวิธี combined weight SVR-KNN ในระยะการพยากรณ์ 1 ชั่วโมงให้ค่า NRMSE เท่ากับ 6.08% และ ในระยะการพยากรณ์ 6 ชั่วโมง ได้ค่า NRMSE เท่ากับ 10.16% โดยผลลัพธ์ที่ได้ถูกประเมินจากข้อมูลจากโรงไฟฟ้า 87 โรงในประเทศเยอรมนี

W.A. Muhammad [AMR18] เปรียบเทียบการใช้วิธี SVR และ RF เพื่อพยากรณ์กำลังผลิตไฟฟ้าในระยะ 1 ชั่วโมง โดยใช้ข้อมูลกำลังผลิตไฟฟ้าในอดีต ความเข้มรังสีแสงอาทิตย์ในอดีต ข้อมูลสภาพอากาศ วันและเดือนที่พยากรณ์ ได้ผลลัพธ์การพยากรณ์จากวิธี RF และ SVR มีค่า RMSE เท่ากับ 2.2470 kWh และ 2.3973 kWh ตามลำดับ โดยใช้ข้อมูลที่ใช้ในการทดลองวัดจากระบบไฟฟ้าซึ่งมีกำลังติดตั้งสูงสุดประมาณ 40-kW

จากการวิจัยที่เกี่ยวข้องข้างต้น แสดงให้เห็นว่าวิธี SVR เป็นวิธีที่นิยมแพร่หลายและมีสมรรถนะที่ดีในการพยากรณ์ในระยะสั้นมาก และ วิธี RF เป็นวิธีที่ [AMR18] นำเสนอว่าให้ผลลัพธ์ที่ดีกว่า SVR โครงงานนี้จึงสนใจที่จะทดลองเปรียบเทียบกลุ่มวิธี แบบจำลองในกลุ่มการเรียนรู้ด้วยเครื่องเพิ่มเติม อันได้แก่ 1) linear regression model , 2) Multivariate adaptive regression spline (MARS) , 3) SVR, 4) RF โดยที่ 2 วิธีในกลุ่มแรกจะทำขึ้นเพื่อเป็นแบบจำลองฐาน (baseline model)

2 ภาพรวมของโครงงาน

2.1 วัตถุประสงค์

1. เพื่อศึกษาและสรุปปัจจัยที่ส่งผลต่อกำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ในระยะสั้นมาก
2. เพื่อเปรียบผลลัพธ์การพยากรณ์ความเข้มแสงจากแบบจำลองในกลุ่ม Linear Regression, Multivariate adaptive regression spline, Support vector regression และ Random forest โดยใช้ตัวชี้วัดสมรรถนะของการพยากรณ์ทางสถิติ

3. เพื่อเปรียบเทียบความซับซ้อนในการคำนวณ (computational complexity) ที่เกิดขึ้นในขั้นตอนการเรียนรู้แบบจำลองและการคำนวณค่าพยากรณ์ของแบบจำลองในข้อ 2.

2.2 ขอบเขต

- การทดลองหลักจะทดลองบนข้อมูลที่วัดได้ ณ ชั้นดادฟ้าตีภาควิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย ในช่วงเวลาตั้งแต่ เดือนมกราคม พ.ศ. 2560 จนถึงเดือนธันวาคม พ.ศ. 2561 ซึ่งประกอบด้วย ค่าแสงอาทิตย์ต่อพื้นที่, ค่ากำลังไฟฟ้า, ค่าความชื้นสัมพัทธ์, อุณหภูมิ, ความเร็วลม, ดัชนีรังสีอัลตราไวโอเลต (UV Index) โดยข้อมูลทั้งหมดถูกลดอัตราสุ่มลงเป็น 30 นาที ส่วนข้อมูลสำรองเป็นข้อมูลที่วัดได้จากโรงไฟฟ้าในภาคกลางจำนวน 1 โรง ในช่วงเวลาตั้งแต่ เดือนมกราคม พ.ศ. 2560 จนถึง เดือนธันวาคม พ.ศ. 2561 ซึ่งประกอบด้วย ค่าแสงอาทิตย์ต่อพื้นที่, ค่ากำลังไฟฟ้า, อุณหภูมิ โดยข้อมูลทั้งหมดถูกลดอัตราสุ่มลงเป็น 30 นาที (จะมีผลการทดลอง เมื่อมีข้อมูลเพียงพอ)
- วิเคราะห์หาตัวแปรต่างๆ ที่ส่งผลต่อกำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ ด้วยวิธีการคัดเลือกคุณลักษณะ ได้แก่ การวิเคราะห์สหสัมพันธ์, การวิเคราะห์สหสัมพันธ์แบบแยกส่วน และการทดสอบเชิงเส้นแบบขั้นตอน
- เปรียบเทียบผลลัพธ์จากพยากรณ์ความเข้มแสงอาทิตย์ล่วงหน้า 4 ชั่วโมง (ค่าผลลัพธ์การพยากรณ์มีความล落เอี้ยด 30 นาที ก่อนคือ 30, 60, 90, ..., 240 นาทีล่วงหน้า) โดยพยากรณ์ในช่วงเวลา 5:30 น. ถึง 17.00 น. (พยากรณ์ทุกๆ 30 นาที) เพื่อให้ได้ค่าพยากรณ์ในช่วงเวลาตั้งแต่ 6:00 น. ถึง 17.30 น.
- การเปรียบเทียบแบบจำลองจะพิจารณาคุณลักษณะแบบจำลองอันได้แก่ 1) Linear Regression 2) Multivariate Adaptive Regression Splines (MARS) 3) Support Vector Regression 4) Random Forest โดย 2 วิธีแรก จัดทำขึ้นเพื่อเป็นแบบจำลองฐาน (baseline model)
- เปรียบเทียบผลลัพธ์การพยากรณ์กับผลลัพธ์จากแบบจำลอง ANN ซึ่งทีมวิจัยสามารถทกริด จุฬาลงกรณ์มหาวิทยาลัยได้จัดทำขึ้น
- ใช้แบบจำลองการแปลงความเข้มแสงอาทิตย์ไปเป็นกำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์

2.3 ผลลัพธ์ที่คาดหวัง

- ผลลัพธ์การเปรียบเทียบการพยากรณ์ความเข้มแสงอาทิตย์ด้วยแบบจำลอง Linear Regression, MARS, SVR, RF ทั้งในแง่ของสมรรถนะการพยากรณ์ และความซับซ้อนในการคำนวณของแต่ละแบบจำลอง ทั้งในส่วนของการเรียนรู้ทางสถิติและขั้นตอนการดำเนินการพยากรณ์
- แบบจำลองการพยากรณ์กำลังผลิตไฟฟ้าในระยะสั้นมากในกลุ่มการเรียนรู้ด้วยเครื่องอันได้แก่ Linear Regression, MARS, SVR, RF และชุดคำสั่งโปรแกรมสำหรับพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ในระยะสั้นมาก

3 หลักการและทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงสัญลักษณ์และตัวแปรที่ใช้ในโครงงานฉบับนี้ หลักการในการคัดเลือกคุณลักษณะและ แบบจำลองท่องฟ้าไซต์ลดจันทร์เทคนิคการประมาณทั้ง 4 วิธีและสุดท้ายจะนำเสนอแบบจำลองการพยากรณ์ค่ากำลังผลิตไฟฟ้า

กำหนดให้ X_1, X_2, \dots, X_p แทนตัวแปรต้น, Y แทนตัวแปรตามคือค่าความเข้มแสงอาทิตย์หรือกำลังผลิตไฟฟ้าตามบริบทกำหนดสัญลักษณ์และตัวแปรดังนี้

ตาราง 1: สัญลักษณ์และตัวแปร

ตัวแปร	ความหมาย	หน่วย
I	ความเข้มแสงอาทิตย์	วัตต์ต่อตารางเมตร
P	กำลังผลิตไฟฟ้า	วัตต์
RH	ความชื้นสัมพัทธ์	เปอร์เซ็นต์
WS	ความเร็วลม	เมตรต่อวินาที
UV	ดัชนีรังสีอัลตราไวโอเลต	-
T	อุณหภูมิภายนอก	องศาเซลเซียส
$\cos(\theta)$	โคไซน์ของมุมของดวงอาทิตย์เทียบกับแนวตั้งจากพื้นโลก	-

- ตัวแปรที่เขียนในรูป $x(t)$ หมายถึงค่า x ณ เวลา t
- ตัวแปรที่เขียนในรูป $\hat{x}(t)$ หมายถึงค่าพยากรณ์ หากเขียนในรูป $x(t)$ หมายถึงค่าที่วัดได้จริง
- ตัวแปรที่เขียนในรูป $\hat{x}_A(t)$ หมายถึงค่าพยากรณ์ของ x จากวิธี A
- การใช้คำดับเวลาจะเขียนอยู่ในรูป $x(t)$ หมายถึงตัวแปร x ที่เวลา t ในวันนี้ๆ หากอยู่เขียนในรูป $x^{(d)}(t)$ หมายถึงตัวแปร x ที่วันที่ d ในเวลา t

ในการทดลองจะกำหนด index ของเวลาดังนี้

- t แทน index ของเวลาปัจจุบัน
- $t - 1, t - 2, \dots, t - k$ หมายถึงเวลา $30, 60, \dots, 30k$ นาทีก่อนหน้า
- $t + 1, t + 2, \dots, t + k$ หมายถึงเวลา $30, 60, \dots, 30k$ นาทีข้างหน้า

ยกตัวอย่างเช่น $I^{d-1}(t)$ หมายถึงความเข้มแสลงอาทิตย์ในวันก่อนหน้าที่เวลาเดียวกันกับเวลา t ปัจจุบัน

3.1 การคัดเลือกคุณลักษณะ

3.1.1 สหสัมพันธ์

สหสัมพันธ์ (Correlation) เป็นค่าที่บ่งบอกความสัมพันธ์เชิงเส้นระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป โดยในการพิจารณาความสัมพันธ์เชิงเส้นระหว่างตัวแปรว่ามีมากน้อยเพียงใด สามารถอุดได้จากค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ซึ่งสามารถคำนวณจากวิธีการทางสถิติได้หลายวิธีซึ่งขึ้นอยู่กับลักษณะของตัวแปรนั้นๆ ในการวัดความสัมพันธ์แต่ละแบบจะต้องมีการทดสอบนัยสำคัญทางสถิติของตัวแปรคุณนั้นๆ ก่อนจะสามารถสรุปความสัมพันธ์ระหว่างตัวแปรได้ การวิเคราะห์ความสัมพันธ์ในรูปแบบนี้จะสามารถตีความถึงความสอดคล้องไปด้วยกันของตัวแปร แต่ไม่ได้หมายความถึงการเป็นเหตุและผลกันระหว่างตัวแปรนั้นๆ

1) สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson correlation Coefficient) เป็นวิธีที่ใช้วัดความสัมพันธ์เชิงเส้นระหว่างตัวแปร 2 ชุดในเขตของตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยสามารถคำนวณได้จากสูตรดังนี้

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (1)$$

โดยที่ ρ แทนสัมประสิทธิ์สหสัมพันธ์เพียร์สัน, $\text{cov}(X, Y)$ แทนความแปรปรวนร่วมของตัวแปร X และ Y , σ_x, σ_y แทนส่วนเบี่ยงเบนมาตรฐานของตัวแปร X และ Y ตามลำดับ

2) สัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน

สัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน (Partial correlation coefficient) เป็นวิธีที่ใช้วัดความสัมพันธ์เชิงเส้นระหว่างตัวแปร 2 ชุด โดยคำนวณจากความคลาดเคลื่อนคงค้างของตัวแปร 2 ชุดนั้นหลังจากกำจัดอิทธิพลเชิงเส้นจากตัวแปรอื่นๆ ออกด้วยการต่อไปนี้

$$\text{cov}(Y_i, Y_j | X) = \text{cov}(Y_i - \hat{Y}_i(X), Y_j - \hat{Y}_j(X)) \quad (2)$$

โดยที่ \hat{Y}_i คือค่าประมาณของ Y_i จาก การวิเคราะห์การลดถอยแบบเชิงเส้นบนข้อมูล X และ \hat{Y}_j คือค่าประมาณของ Y_j จาก การวิเคราะห์การลดถอยแบบเชิงเส้นบนข้อมูล X

$$\rho_{Y_i, Y_j | X} = \frac{\text{cov}(Y_i, Y_j | X)}{\sqrt{\text{var}(Y_i - \hat{Y}_i(X)) \text{var}(Y_j - \hat{Y}_j(X))}} \quad (3)$$

ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรสุ่มแบบเก้าส์เชียน 2 ตัวใดๆ สามารถคำนวณจากเมทริกซ์ความแปรปรวนร่วมผกผันได้ดังนี้

$$\rho_{X_i X_j | V \setminus \{X_i, X_j\}} = -\frac{\Sigma_{ij}^{-1}}{\sqrt{\Sigma_{ii}^{-1} \Sigma_{jj}^{-1}}} \quad (4)$$

โดยที่ ρ แทนสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วน V แทนเขตของตัวแปรสุ่ม X_1, X_2, \dots, X_K , Σ แทนเมทริกซ์ความแปรปรวนร่วมของตัวแปรสุ่ม ในเขต V

3.1.2 วิธีการถดถอยเชิงเส้นแบบขั้นตอน

วิธีการถดถอยเชิงเส้นแบบขั้นตอน (Stepwise linear regression) เป็นวิธีหนึ่งในการหาสมการถดถอยเชิงเส้นแสดงความสัมพันธ์ระหว่างตัวแปรตัวแปรที่พิจารณา ซึ่งแตกต่างจากวิธีการถดถอยเชิงเส้น (linear regression) ตรงที่จะมีการเพิ่ม/ลดตัวแปรตัวแปรที่ใช้ในการสร้างสมการโดยการใช้ค่าสถิติเป็นเกณฑ์ในการเลือกตัวแปรตัวแปรที่จะเพิ่ม/ลด แต่ละขั้นตอน ซึ่งวิธีการถดถอยเชิงเส้นแบบขั้นตอน เป็นวิธีที่เกิดจากการประยุกต์ระหว่างวิธีการเลือกแบบก้าวหน้า (forward selection) และวิธีการตัดทิ้งแบบถอยหลัง (backward deletion)

- วิธีการเลือกแบบก้าวหน้า (forward selection) จะเริ่มต้นจากการสร้างสมการค่าคงที่สำหรับประมาณค่าตัวแปรตามที่พิจารณา จากนั้นในแต่ละขั้นตอนจะทดลองเพิ่มตัวแปรตัวแปรที่ใช้ในการสร้างสมการถดถอย จากนั้นตรวจสอบว่าการเพิ่มตัวแปรตัวแปรตัวแปรนั้นส่งผลให้ค่า RMSE ในประมาณค่าตัวแปรตามลดลงอย่างมีนัยสำคัญหรือไม่ โดยการทดสอบนัยสำคัญทางสถิติ จากนั้นจึงตัดสินใจเพิ่มตัวแปรที่มี $p\text{-value}$ ต่ำสุดเข้าไปในกลุ่มตัวแปรที่จะใช้ในการสร้างสมการถดถอย และดำเนินกระบวนการต่อจนกระทั่งกระบวนการจะสิ้นสุดเมื่อ $p\text{-value}$ จากการทดสอบตัวแปรตัวแปรทุกตัวมีค่ามากกว่าค่าที่กำหนด
- วิธีการตัดทิ้งแบบถอยหลัง (backward deletion) จะเริ่มต้นจากการสร้างสมการถดถอยเชิงเส้นที่ประกอบด้วยตัวแปรตัวแปรทุกตัวในสมการก่อน จากนั้นในแต่ละขั้นตอนจะทดลองตัดตัวแปรตัวแปรออกจากกลุ่มตัวแปรที่ใช้ในการสร้างสมการที่เหลือ จากนั้นตรวจสอบว่าการที่มีแปรตัวแปรตัวแปรที่อยู่ในกลุ่มนั้น ส่งผลให้ค่า RMSE ในประมาณค่าตัวแปรตามลดลงอย่างมีนัยสำคัญหรือไม่ (เมื่อเทียบกับหลังตัดตัวแปรออก) โดยการทดสอบนัยสำคัญทางสถิติ จากนั้นจึงตัดสินใจตัดตัวแปรที่มี $p\text{-value}$ สูงสุดออกจากกลุ่มตัวแปรที่จะใช้ในการสร้างสมการถดถอย และดำเนินกระบวนการต่อจนกระทั่งกระบวนการจะสิ้นสุดเมื่อ $p\text{-value}$ จากการทดสอบตัวแปรตัวแปรทุกตัวมีค่าต่ำกว่าค่าที่กำหนด

สำหรับวิธีการถดถอยเชิงเส้นแบบขั้นตอนในแต่ละขั้นตอนจะเพิ่มตัวแปรตัวแปรที่ใช้ในการสร้างสมการถดถอยโดยวิธีการเลือกแบบก้าวหน้า และเมื่อสิ้นสุดขั้นตอนการเพิ่มตัวแปรแล้วจะทำการทดสอบนัยสำคัญทางสถิติ จึงตัดตัวแปรตัวแปรที่ไม่มีความสัมพันธ์เชิงเส้นกับตัวแปรตาม โดยการพิจารณาตัวแปรที่อยู่ในกลุ่มตัวแปรที่ใช้ในการสร้างสมการถดถอย และดำเนินกระบวนการต่อจนกระทั่งกระบวนการจะสิ้นสุดเมื่อ $p\text{-value}$ จากการทดสอบตัวแปรตัวแปรที่ใช้ในการสร้างสมการถดถอยหลังจากการดำเนินกระบวนการเลือกแบบก้าวหน้า

3.2 แบบจำลองห้องฟ้าใส

แบบจำลองห้องฟ้าใส เป็นแบบจำลองที่ใช้ในการคำนวณความเข้มแสงอาทิตย์ที่ตกกระทบบนพื้นผิวโลกในสภาวะที่ห้องฟ้าปราศจากเมฆ [จ57] ซึ่งในโครงงานฉบับจะใช้ค่าความเข้มแสงอาทิตย์ในสภาวะที่ห้องฟ้าปราศจากเมฆเป็นคุณลักษณะหนึ่งในการพยากรณ์ ซึ่งมีการนำเสนอแบบจำลองสำหรับประมาณค่าความเข้มแสงอาทิตย์ไว้หลากหลายแบบจำลอง ดังนี้ [จ57]

- แบบจำลองของ Haunwitz ถูกพัฒนาโดยใช้ข้อมูลค่ารังสีอาทิตย์ที่วัดได้ ณ ตอนใต้ของเมืองบอสตันประเทศสหรัฐอเมริกา [Ber45]

$$I(t) = 1098 \cos(\theta(t)) e^{-0.057/\cos(\theta(t))} \quad (5)$$

- แบบจำลองของ Berger-Duffie [Vio97] (I_0 เป็นค่าคงที่มีค่าเท่ากับ 1366.1W/m^2)

$$I(t) = I_0(0.7 \cos(\theta(t))) \quad (6)$$

- แบบจำลองของ Adnote-Bourges-Campana-Gicquel [Vio97]

$$I(t) = 951.39 \cos(\theta(t))^{1.15} \quad (7)$$

- แบบจำลองของ Robledo-Soler [Vio97] ($\theta(t)$ มีหน่วยเป็นองศา)

$$I(t) = 1159.24 \cos(\theta(t))^{1.179} e^{-0.0019(90^\circ - \theta(t))} \quad (8)$$

- แบบจำลอง ASHRAE ที่พัฒนาจากข้อมูลสภาพภูมิอากาศในประเทศไทยโดย Pansak และ Chumnong [PC07] โดยที่ค่ารังสีอาทิตย์รวมคำนวณได้จากการรวมของรังสีตรงและรังสีกระจายดังสมการ

$$\begin{aligned} I(t) &= I_{\text{direct}}(t) + I_{\text{diffuse}}(t) \\ &= Ae^{-B \sec(\theta(t))} + CAe^{-B \sec(\theta(t))} \\ &= Ke^{-B \sec(\theta(t))} \end{aligned} \quad (9)$$

โดยที่ค่าคงที่ K, B เป็นค่าคงที่ ที่ได้จากการประมาณโดยใช้ข้อมูลที่วัดได้ในอดีต

6. แบบจำลองของ Kasten [PR02, MEPV12]

$$I(t) = 0.84I_0 \cos(\theta(t))e^{-0.027AM(t)(f_{h1}+f_{h2}(T_L-1))} \quad (10)$$

โดยที่ $AM(t)$ คือ มวลอากาศ ณ เวลา t คำนวนจาก $AM(t) = 1/(\cos(\theta(t))+0.50572(96.07995^\circ-\theta(t))-1.6364$
 $f_{h1} = e^{-h/8000}$, $f_{h2} = e^{-h/1250}$, h คือ ระดับความสูงจากน้ำทะเล, T_L คือ ค่าความชุ่มของบรรยากาศ

7. แบบจำลองของ Ineichen [PR02]

$$I(t) = a_1 I_0 \cos(\theta(t))e^{-a_2 AM(t)(f_{h1}+f_{h2}(T_L-1))} \quad (11)$$

โดยที่ $a_1 = 5.09 \times 10^{-5}h + 0.868$, $a_2 = 3.92 \times 10^{-5}h + 0.0387$

ซึ่งในบทต่อๆ ไปเราจะใช้สัญลักษณ์ I_{clr} แทนค่าความเข้มแสงอาทิตย์ในสภาวะที่ห้องฟ้าใส
 สำหรับแบบจำลองของ ASHRAE, Kasten และ Ineichen มีค่าพารามิเตอร์ที่ยังไม่ทราบค่า โดยเราสามารถประมาณได้โดยใช้ข้อมูลวัด

3.2.1 การตรวจจับวันท่องฟ้าใสจากข้อมูลวัด

จากแบบจำลองท่องฟ้าใสข้างต้นจะเห็นได้ว่าแบบจำลอง ASHRAE, Kasten และ Ineichen ต้องการข้อมูลแสงอาทิตย์ในอดีตของวันท่องฟ้าใสในการประมาณค่าพารามิเตอร์ต่างๆ ที่ใช้ในแบบจำลอง จึงมีการนำเสนอขั้นตอนที่ใช้ในการตรวจจับวันท่องฟ้าใสจากข้อมูลดิบดังนี้

1. คำนวนค่าเฉลี่ยความเข้มแสงอาทิตย์ในแต่ละจุดเวลา เพื่อให้เป็นค่าความเข้มแสงอาทิตย์อ้างอิง
2. ในแต่ละวันคำนวนค่า cosine distance ระหว่างค่าความเข้มแสงอาทิตย์อ้างอิงกับค่าความเข้มแสงอาทิตย์ของวันหนึ่งๆ จาก (12)

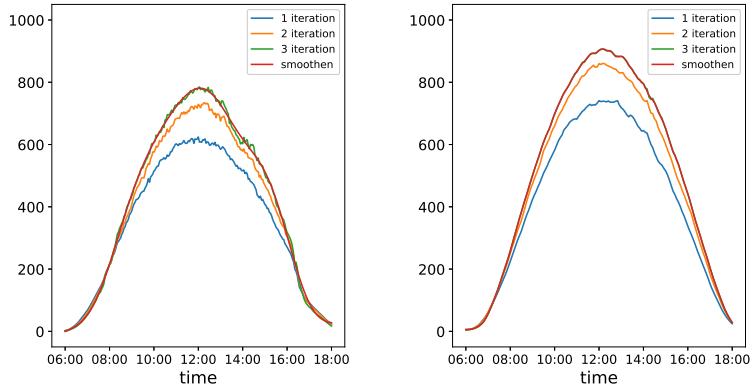
$$\text{cosine distance} = 1 - \frac{I_{ref} \cdot I_i}{\|I_{ref}\|_2 \|I_i\|_2} \quad (12)$$

โดยที่ I_{ref} คือเวกเตอร์ค่าความเข้มแสงอาทิตย์อ้างอิง, I_i คือเวกเตอร์ค่าความเข้มแสงอาทิตย์ของวันที่ i

ค่า cosine distance ที่คำนวนได้จะมีค่าอยู่ระหว่าง 0-1 ถ้าในวันใด มีค่าเข้าใกล้ 0, หมายความว่าความเข้มแสงอาทิตย์ของวันนั้นๆ มีความใกล้เคียงกับความเข้มแสงอาทิตย์อ้างอิงมาก (เป็นวันท่องฟ้าใส)

3. เลือกวันที่มีค่า cosine distance น้อยกว่าเบอร์เซ็นไทล์ที่ 25 จากวันทั้งหมดในข้อมูลมาเพื่อคำนวนหาค่าเฉลี่ยความเข้มแสงอาทิตย์ในแต่ละจุดเวลา และกำหนดให้เป็นค่าความเข้มแสงอาทิตย์อ้างอิงใหม่
4. คำนวนค่า cosine distance ระหว่างค่าความเข้มแสงอาทิตย์อ้างอิงใหม่กับค่าความเข้มแสงอาทิตย์ของวันที่ถูกเลือก
5. ทำขั้นตอนที่ 3 และ 4 จำนวนวันที่เลือกมา มีน้อยกว่าจำนวนวันที่กำหนด และกำหนดให้ค่าความเข้มแสงอาทิตย์อ้างอิงที่คำนวนได้เป็นความเข้มแสงอาทิตย์ท่องฟ้าใสอ้างอิง
6. ตั้งเกณฑ์ว่าวันท่องฟ้าใสจะมีค่า cosine distance ระหว่างความเข้มแสงอาทิตย์กับความเข้มแสงอาทิตย์ท่องฟ้าใสอ้างอิงน้อยกว่าค่าค่านึง เพื่อคัดเลือกหารันท่องฟ้าใส โดยในงานนี้เลือกใช้ค่าเท่ากับ 0.006

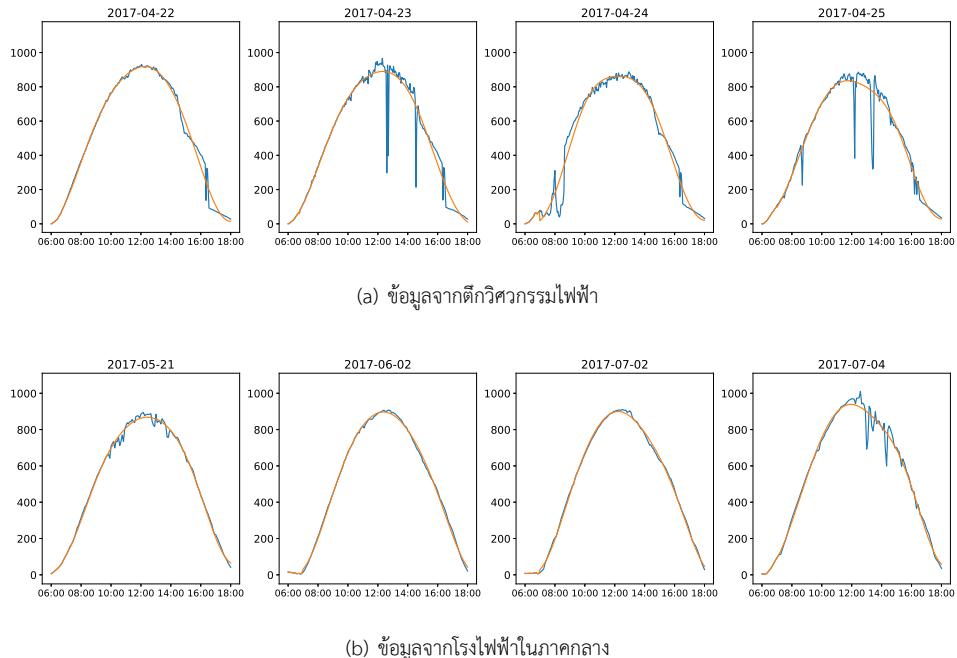
เนื่องจากจุดประสงค์ของการตรวจจับวันท่องฟ้าใสคือเพื่อหาข้อมูลอ้างอิงในการคำนวนค่าพารามิเตอร์ของแบบจำลองท่องฟ้าใส ดังนั้นจึงเลือกใช้ข้อมูลวัดที่มีความละเอียดสูง และลดความแปรปรวนของข้อมูลวัดของวันที่ถูกเลือกเป็นวันท่องฟ้าใสโดยนำข้อมูลดังกล่าวไปผ่าน butter-worth low-pass filter แบบสองทิศทาง



(a) ข้อมูลจากตึกวิศวกรรมไฟฟ้า

(b) ข้อมูลจากโรงไฟฟ้าในภาคกลาง

รูป 2: ค่าความเข้มแสงอาทิตย์ท้องฟ้าใส่อ้างอิงในแต่ละการวนซ้ำโดยใช้ข้อมูลจากโรงไฟฟ้าภาคกลางและตึกวิศวกรรมไฟฟ้า
เส้นสีแดงคือเส้นแสดงข้อมูลความเข้มแสงอาทิตย์ท้องฟ้าใส่อ้างอิงจากการวนซ้ำครั้งสุดท้ายหลังนำไปผ่านตัวกรอง



รูป 3: ตัวอย่างค่าความเข้มแสงของวันที่ถูกเลือกเป็นวันท้องฟ้าใส่ด้วยขั้นตอนการตรวจจับนี้

เส้นสีฟ้าแสดงถึงข้อมูลดิบของความเข้มแสงอาทิตย์และเส้นสีส้มแสดงถึงค่าความเข้มแสงที่ได้หลังนำข้อมูลดิบไปผ่าน butterworth low-pass filter

3.3 เทคนิคการประมาณ

ในโครงการนี้ เราชาระพยากรณ์ค่าความเข้มแสงอาทิตย์ในช่วงเวลาล่วงหน้า 4 ชั่วโมง (ค่าผลลัพธ์การพยากรณ์มีความละเอียด 30 นาที กล่าวคือ จะพยากรณ์ 30, 60, 90, ..., 240 นาทีล่วงหน้า) โดยวิธีที่จะนำมาเปรียบเทียบมี 4 วิธีคือ 1) Linear Regression 2) Multivariate Adaptive Regression Splines (MARS) 3) Support Vector Regression (SVR) 4) Random Forest (RF) โดย 2 วิธีแรกจะทำขึ้นเพื่อเป็นแบบจำลองฐาน (baseline model)

3.3.1 Linear regression

เราจะพยากรณ์ค่าความเข้มแสงอาทิตย์ $I(t+1), I(t+2), \dots, I(t+8)$ โดยใช้วิธีการลดถอยเชิงเส้นซึ่งมีตัวแปรตันดังนี้

- ค่าความเข้มแสงอาทิตย์ในอดีตประกอบด้วย

- $I(t), I(t-1), \dots, I(t-7)$

- $I^{(d-1)}(t+1), I^{(d-1)}(t+2), \dots, I^{(d-1)}(t+8)$

2. ค่าความเข้มแสงอาทิตย์ในสภาวะท้องฟ้าสีประกอบด้วย

- $I_{\text{clr}}(t+1), I_{\text{clr}}(t+2), \dots, I_{\text{clr}}(t+8)$

3.3.2 Multivariate adaptive regression splines (MARS)

เราจะพิจารณาค่าความเข้มแสงอาทิตย์ $I(t+1), I(t+2), \dots, I(t+8)$ โดยใช้ปรัตนและตัวแปรตามเช่นเดียวกับวิธี Linear regression โดย Multivariate Adaptive Regression Splines (MARS) เป็นวิธีหนึ่งในการเรืองลดโดยในสร้างสมการความสัมพันธ์แบบไม่เป็นเชิงเส้น (เชิงเส้นแบบเป็นช่วง) ระหว่างตัวแปรตามและตัวแปรตาม ดังนี้ [FHT01]

$$\hat{Y}(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X) \quad (13)$$

โดยที่ β แทนสัมประสิทธิ์, M แทนจำนวนช่วง และ $h(X)$ แทนฟังก์ชันเชิงเส้นแบบเป็นช่วงในตัวแปร X ซึ่งเรียกว่า basis function เกี่ยวน์ได้ในรูป $h_m(X) = \max(0, X - c_m)$ ในขั้นตอนทำแบบจำลองเราจะเลือกค่า β และ $h(X)$ ที่ทำให้ค่าผลรวมของค่าเศษเหลือกำลังสอง (residual sum of squares) มีค่าน้อยที่สุดในชุดข้อมูลฝึกโดยในโครงงานฉบับนี้ได้ทำการทดลองโดยใช้ Python library ที่มีชื่อว่า Py-Earth ซึ่งใช้ขั้นตอนของ Jerome Friedman [Jer91] ในการประมาณพารามิเตอร์ของแบบจำลอง เนื่องจากวิธี MARS เป็นตัวอย่างของวิธีการพยากรณ์แบบไม่เป็นเชิงเส้น (เชิงเส้นแบบเป็นช่วง) ซึ่งเข้าใจได้ง่ายและมีพื้นฐานมาจากวิธี Linear regression ดังนั้นเราจึงเลือกวิธีนี้เป็นหนึ่งในวิธีที่จะใช้เป็นแบบจำลองฐาน (baseline model) สำหรับเปรียบเทียบกับวิธีการอื่น

3.3.3 Support Vector Regression

Support vector regression เป็นเทคนิคการเรียนรู้ด้วยเครื่องที่ได้รับการพัฒนามาจาก Vapnik (1995) [CV95] และได้รับความนิยมเนื่องจากสมรรถนะที่ดีในการจัดการกับปัญหาการพยากรณ์อนุกรมเวลาโดยใช้หลักการวิเคราะห์การลดโดยที่สามารถอธิบายได้ทั้งรูปแบบความสัมพันธ์เชิงเส้นและไม่เชิงเส้นดังนี้ [FAGJ15, SS04, Vap99] ภายใต้ชุดข้อมูลฝึก $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ โดยที่ $x_i \in \mathbb{R}^p$ เป็นข้อมูลฝึกษาเข้า และ $y_i \in \mathbb{R}$ เป็นข้อมูลฝึกษาออก หลักการของ Support vector regression คือการเปลี่ยนปริภูมิของข้อมูลฝึกษาเข้า X ไปยังปริภูมิใหม่ (\mathcal{H}) ผ่านฟังก์ชัน $\varphi(x)$ หลังจากนั้นวิเคราะห์การลดโดยในปริภูมิใหม่เพื่อหาฟังก์ชันในการประมาณ y_i ดัง (14) โดยหาก $\varphi(x)$ เป็นฟังก์ชันไม่เชิงเส้นแล้วฟังก์ชันเชิงเส้นที่ได้ในปริภูมิใหม่นั้นจะสมนัยกับฟังก์ชันไม่เชิงเส้นในปริภูมิเดิม

$$f(x) = \langle w, \varphi(x) \rangle + b \quad (14)$$

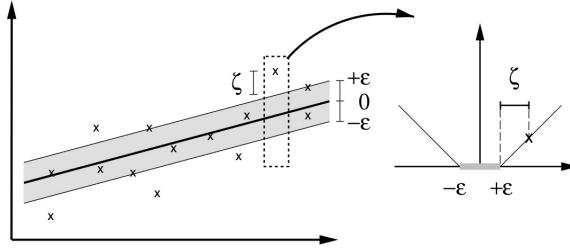
โดยที่ $w \in \mathcal{H}, b \in \mathbb{R}$ แทนเวกเตอร์ค่าน้ำหนัก และค่าคงที่ตามลำดับในการหาค่า w, b สามารถทำได้โดยแก้ไขปัญหาการหาค่าที่เหมาะสมโดยให้ข้อจำกัดดังนี้

$$\begin{aligned} & \underset{w, b, u_i, v_i}{\text{minimize}} \quad (1/2) \|w\|^2 + C \sum_{i=1}^n (u_i + v_i) \\ & \text{subject to} \quad y_i - \langle w, \varphi(x_i) \rangle - b \leq \varepsilon + u_i, \quad i = 1, 2, \dots, n, \\ & \quad \langle w, \varphi(x_i) \rangle + b - y_i \leq \varepsilon + v_i, \quad i = 1, 2, \dots, n, \\ & \quad u_i, v_i \geq 0 \quad , \quad i = 1, 2, \dots, n \end{aligned} \quad (15)$$

ε แทนพารามิเตอร์ที่กำหนดขนาดของบริเวณค่าความคลาดเคลื่อนที่ยอมรับได้ โดยค่าความคลาดเคลื่อนที่ตกลอยู่ภายในบริเวณนี้จะไม่ถูกนำไปคิดในฟังก์ชันสูญเสีย u, v แทนตัวแปรหย่อน (Slack Variable) ซึ่งเป็นค่าที่ยอมให้บางจุดข้อมูลมีค่าความคลาดเคลื่อนมากกว่าค่า ε ที่กำหนดได้ดังแสดงในสมการข้อจำกัด

ฟังก์ชันวัตถุประสงค์ต้องการที่จะหาค่าต่ำสุดของพจน์ $(1/2) \|w\|^2$ ซึ่งเป็นค่าที่ลงโทษความซับซ้อนของแบบจำลองและยังสอดคล้องกับการหาระยะห่างมากที่สุดของระบบของบริเวณค่าความคลาดเคลื่อนที่ยอมรับได้ และพจน์ $C \sum_{i=1}^n (u_i + v_i)$ ซึ่งแสดงถึงฟังก์ชันสูญเสียแบบ ε -incentive ดังแสดงใน (16) พจน์นี้สอดคล้องกับการพิจารณาฟังก์ชันลงโทษ (penalty function) ที่ลงโทษตัวแปรหย่อนที่ยอมให้เกิดความคลาดเคลื่อนมากกว่าค่า ε ในบางจุดข้อมูล ส่วนค่าคงที่ C เป็นค่าน้ำหนักที่ควบคุมความสมดุลในการหาค่าต่ำสุดระหว่าง 2 พจน์ดังกล่าว โดยสรุปการหาค่าต่ำสุดของ (14) สอดคล้องกับหลักการการเรียนรู้ทางสถิติที่ต้องการควบคุมทั้งค่าความคลาดเคลื่อนในชุดข้อมูลฝึกและความซับซ้อนของแบบจำลอง [JWHT13]

$$|y - f(x)|_{\varepsilon} = \begin{cases} |y - f(x)| - \varepsilon, & \text{if } |y - f(x)| > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (16)$$



รูป 4: รูปแสดงฟังก์ชันสูญเสียแบบ ε -incentive ของ linear SVR [SS04]

ในการแก้ปัญหา (15) (primal form) เราพบว่ามีการคำนวนหา w ซึ่งอยู่ในปริภูมิ \mathcal{H} ที่อาจมีมิติสูง จึงอาจมีความจำเป็นต้องใช้กำลังในการคำนวนสูง ดังนั้นเราประยุกต์ใช้หลักการ Lagrange duality เปลี่ยนมาพิจารณา dual form ของปัญหานี้ภายใต้เงื่อนไข Karush-Kuhn-Tucker (KKT) แทน ซึ่งเป็นการคำนวนหา λ, ν ซึ่งอยู่ในปริภูมิ \mathbb{R}^n ดังนี้

$$\begin{aligned} & \underset{\lambda, \nu}{\text{maximize}} \quad - (1/2)(\lambda - \nu)^T Q(\lambda - \nu) - \varepsilon \sum_{i=1}^n (\lambda_i + \nu_i) + \sum_{i=1}^n y_i(\lambda_i - \nu_i) \\ & \text{subject to} \quad \mathbf{1}^T(\lambda - \nu) = 0 \quad \text{และ} \quad \lambda_i, \nu_i \in [0, C], \quad i = 1, 2, \dots, n \end{aligned} \quad (17)$$

โดย $Q_{ij} = \varphi(x_i)^T \varphi(x_j)$ และค่าคงที่บวก $\lambda, \nu \in \mathbb{R}^n$ แทนตัวคุณลักษณะ ซึ่งจาก (17) ได้ผลลัพธ์ดังนี้

$$w = \sum_{i=1}^n (\lambda_i - \nu_i) \varphi(x_i), \quad \text{ดังนั้น} \quad f(x) = \sum_{i=1}^n (\lambda_i - \nu_i) \langle \varphi(x_i), \varphi(x) \rangle + b = \sum_{i=1}^n (\lambda_i - \nu_i) k(x_i, x) + b \quad (18)$$

โดยที่ x_i แทนจุดข้อมูลขาเข้าในชุดข้อมูลฝึก ส่วน x แทนจุดข้อมูลขาเข้าในชุดข้อมูลตรวจสอบ หรือชุดข้อมูลทดสอบ จาก (18) การคำนวนผลลัพธ์ที่ได้ขึ้นกับ support vectors โดยไม่ขึ้นกับมิติของปริภูมิ \mathcal{H} นอกจากนี้ยังสามารถประยุกต์ใช้ Kernel Trick โดยการคำนวนฟังก์ชันเครื่องเรียนแทนการคำนวนผลคุณแบบจุดของข้อมูลขาเข้าในปริภูมิ \mathcal{H} หนึ่งๆ จากข้อได้เปรียบทั้งต้นจึงเห็นว่าในการพิจารณา dual problem สามารถลดกำลังการคำนวนในการแก้ปัญหาลงได้มาก ส่วนเงื่อนไข Karush-Kuhn-Tucker (KKT) ที่ทำให้การแก้สมการในรูป dual form ได้ผลลัพธ์เดียวกับการแก้สมการในรูป primal form ได้แก่

$$\lambda_i(\varepsilon + u_i - y_i + \langle w, \varphi(x_i) \rangle + b) = 0 \quad (19)$$

$$\nu_i(\varepsilon + v_i + y_i - \langle w, \varphi(x_i) \rangle - b) = 0 \quad (20)$$

$$(C - \lambda_i)u_i = 0 \quad (21)$$

$$(C - \nu_i)v_i = 0 \quad (22)$$

จากสมการเงื่อนไขดังกล่าว สามารถสรุปได้ดังนี้

1. มีเฉพาะคู่ค่าดับ (x_i, y_i) ที่มีค่าตัวคุณลักษณะเท่ากับ C เท่านั้นที่ ตกลงอยู่นอกบริเวณความคลาดเคลื่อนที่ยอมรับได้
2. $\lambda_i \nu_i = 0$ หรือกล่าวได้ว่าคู่ค่าดับ (λ_i, ν_i) ใดๆ จะมีค่าใดค่าหนึ่งเท่ากับศูนย์เสมอ
3. ถ้า $\lambda_i \in (0, C)$ และ $u_i = 0$ และถ้า $\nu_i \in (0, C)$ และ $v_i = 0$ ดังนั้นจากเงื่อนไขสามารถนำไปใช้ในการคำนวนหาค่า b จาก (19), (20) ดังนี้

$$\begin{aligned} b &= y_i - \langle w, \varphi(x_i) \rangle - \varepsilon \quad \text{สำหรับ } \lambda_i \in (0, C) \\ b &= y_i - \langle w, \varphi(x_i) \rangle + \varepsilon \quad \text{สำหรับ } \nu_i \in (0, C) \end{aligned} \quad (23)$$

ฟังก์ชันเครื่องเรียนที่เป็นที่นิยมสำหรับ Support Vector Regression มีดังนี้

1. Linear kernel : $k(x, x') = \langle x, x' \rangle$
2. Polynomial kernel : $k(x, x') = (\gamma \langle x, x' \rangle + r)^d$

3. RBF kernel : $k(x, x') = \exp(-\gamma \|x - x'\|^2)$
4. Sigmoid kernel : $k(x, x') = \tanh(\gamma \langle x, x' \rangle + r)$

โดยที่ r, d, γ เป็นพารามิเตอร์ของฟังก์ชันเครื่องเรียน

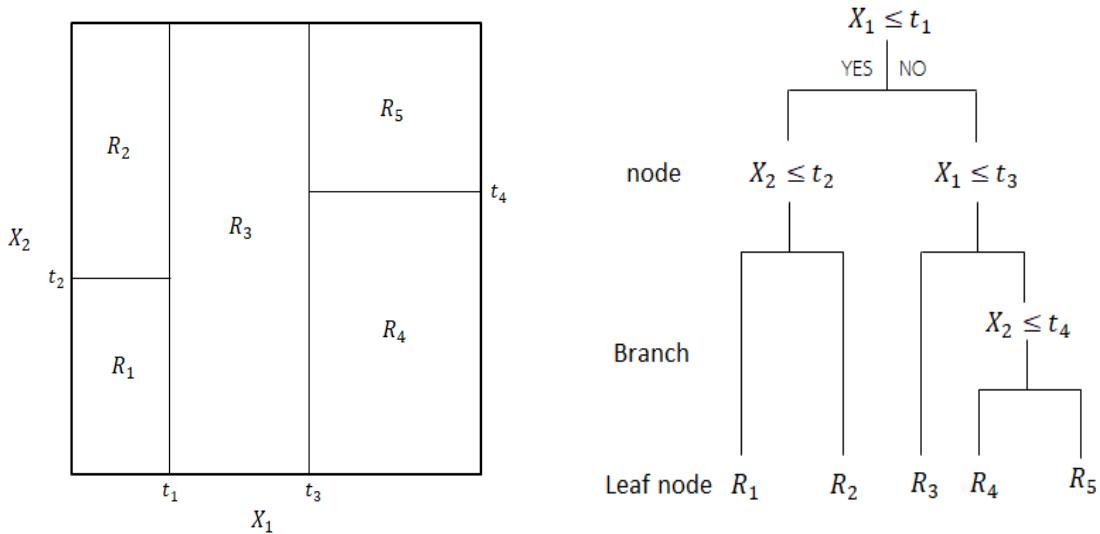
โครงงานฉบับนี้ได้ทดลองโดยใช้ the Python package scikit-learn ซึ่งมีพื้นฐานมาจาก LIBSVM library ในโครงงานนี้ค่าพารามิเตอร์ที่ใช้คือ

$$C = 128, \gamma = 0.125, \varepsilon = 4 \quad (24)$$

โดยใช้วิธีการเลือกพารามิเตอร์แสดงในหัวข้อที่ 10.1

3.3.4 Random Forest

แบบจำลอง Random forest ถูกนำเสนอครั้งแรกในปี ค.ศ. 1995 โดย Tin Kam Ho เป็นวิธีที่อิงจากแบบจำลองต้นไม้ตัดโดย (Regression tree model) ดังที่จะอธิบายต่อไป แบบจำลองต้นไม้ตัดโดยเป็นการประยุกต์หลักการของแบบจำลองต้นไม้ตัดสินใจ (Decision tree model) เพื่อใช้ในการพยากรณ์ค่าของตัวแปรที่พิจารณาโดยอาศัยวิธีการแบ่งกลุ่มของตัวแปรต้น หลักการของแบบจำลองต้นไม้ตัดโดย สามารถอธิบายได้เป็น 2 ขั้นตอนดังนี้ [JWHT13]



รูป 5: ตัวอย่างการแบบจำลองต้นไม้สำหรับปริภูมิตัวแปรต้น 2 มิติ

1. แบ่งปริภูมิของตัวแปรต้น X_1, X_2, \dots, X_p ออกเป็น J ส่วนที่ไม่มีการซ้อนทับซึ่งกันและกัน, ให้ปริภูมิย่อยนั้นเรียกว่า R_1, R_2, \dots, R_J
2. สำหรับทุกๆ ข้อมูลของตัวแปรต้นที่อยู่ใน R_j เราจะพยากรณ์ค่าของตัวแปรตามให้มีค่าเท่ากับค่าเฉลี่ยของค่าตัวแปรตามในชุดข้อมูลฝึกทั้งหมด ซึ่งค่าของตัวแปรต้นโดยใน R_j

โดยในขั้นตอนที่ 1 จะเลือกแบ่งปริภูมิของตัวแปรต้นออกเป็น ปริภูมิย่อย R_1, R_2, \dots, R_J ซึ่งมีลักษณะเป็น high-dimensional rectangles เพื่อให้ได้ ปริภูมิย่อย R_1, R_2, \dots, R_J ซึ่งให้ค่า residual squared error (RSS) ที่น้อยที่สุด กำหนดโดย

$$\text{RSS} = \sum_{j=1}^J \sum_{i \in R_j} \|y_i - \hat{y}_{R_j}\|_2^2 \quad (25)$$

ในทางปฏิบัติ ขั้นตอนวิธีที่ใช้หาค่าตอบของปัญหาข้างต้น เรียกว่าวิธี recursive binary splitting คือ การแบ่งปริภูมิออกเป็นปริภูมิย่อยทีละสองปริภูมิ โดยวิธีการวนซ้ำ ซึ่งมีเงื่อนไขคือในแต่ละรอบจะเลือกการแบ่งปริภูมิที่ทำให้ค่า RSS มีค่าลดลงมากที่สุด รูปที่ 5 แสดงตัวอย่าง

ในกรณีที่ปริภูมิของตัวแปรต้นเป็นปริภูมิ 2 มิติ และจาก รูปที่ 5(b) จะเห็นว่าด้วยลักษณะของขั้นตอนวิธีนี้ องทำให้ลักษณะของแบบจำลองนี้คล้ายการแตกกิ่งของต้นไม้และถูกเรียกว่าแบบจำลองต้นไม้ โดยแต่ละส่วนจะถูกเรียกว่า ปม (node) และแขนง (branch) แบบจำลองต้นไม้ ลดด้อยดังที่กล่าวไว้ข้างต้นมักประสบกับปัญหาเรื่องความแปรปรวนที่สูง ซึ่งหมายความว่าหากเราแบ่งชุดข้อมูลฝึกออกเป็นสองส่วน จำนวนน้ำหนาแบบจำลองโดยใช้ข้อมูลฝึกแต่ละส่วน ผลลัพธ์ในการพยากรณ์ที่ได้จะแตกต่างกันมาก ซึ่งในทางตรงกันข้ามแบบจำลองที่มีความแปรปรวนต่ำจะให้ผลลัพธ์ที่ใกล้เคียงกันแม้ว่าจะเปลี่ยนชุดข้อมูลฝึก แบบจำลอง Random forest เป็นการรวมผลการพยากรณ์จากแบบจำลองต้นไม้ ลดด้อยจำนวนมาก โดยผลการพยากรณ์ของแบบจำลอง Random forest จะกำหนดให้เป็นค่าเฉลี่ยของค่าพยากรณ์จากทุกๆแบบจำลองต้นไม้ ย่อย โดยในแต่ละปม (node) ของแบบจำลองย่อย จะสุ่มเลือกใช้ จำนวนคุณลักษณะของตัวแปรต้นเพียง m คุณลักษณะจากทั้งหมด p คุณลักษณะ ซึ่งพิจารณาได้ว่ากระบวนการหางต้นเป็นการลดทดสอบพื้นที่ของกลุ่มแบบจำลองต้นไม้ ซึ่งจะทำให้แบบจำลองรวมมีความแปรปรวนลดลง และมีความคงทนต่อการเปลี่ยนแปลงชุดข้อมูล นอกจากนี้แบบจำลอง Random forest ยังสามารถประยุกต์ใช้ร่วมกับวิธี bootstrap ซึ่งมีหลักการคือ ในขั้นตอนฝึกของแต่ละแบบจำลองต้นไม้ย่อย จะมีการสุ่มตัวอย่างชุดข้อมูลฝึกที่จะใช้ในการฝึกแต่ละแบบจำลองจากชุดข้อมูลฝึกทั้งหมด ซึ่งจะทำให้แบบจำลองต้นไม้ย่อยแต่ละแบบมีความแตกต่างกันมากขึ้น พารามิเตอร์สำคัญที่เป็นตัวกำหนดเงื่อนไขของแบบจำลอง และยังส่งต่อประสิทธิภาพ/ความซับซ้อนในการคำนวณของการพยากรณ์มีดังนี้

1. จำนวนแบบจำลองต้นไม้ทั้งหมดภายในป่า เขียนแทนด้วย n_{tree}
2. จำนวนระดับหรือความลึกมากสุดของต้นไม้ที่ยอมรับได้ เขียนแทนด้วย d
คือจำนวนปม (node) ทั้งหมดที่มากที่สุด เมื่อนับตั้งแต่ใบ (leaf node) ไปจนถึงปมนสุด ดังรูปที่ 5(b)
3. จำนวนตัวอย่างจากชุดข้อมูลฝึกน้อยสุดที่ยินยอมให้มีการเริ่มต้นแบ่งปริภูมิ เขียนแทนด้วย $n_{min_samples_split}$
คือจำนวนตัวอย่างในข้อมูลฝึกที่น้อยที่สุดในแต่ละปม (node) ก่อนเริ่มการแตกใบ (leaf node) ดังรูปที่ 5(b)
4. จำนวนตัวอย่างจากชุดข้อมูลฝึกน้อยสุดที่ยินยอมไว้แต่ละปริภูมิย่อย เขียนแทนด้วย $n_{min_samples_leaf}$
คือจำนวนตัวอย่างในข้อมูลฝึกที่น้อยที่สุดในแต่ละใบ (leaf node) ดังรูปที่ 5(b)
5. จำนวนคุณลักษณะของตัวแปรต้นใช้ในแต่ละปมของแบบจำลองต้นไม้ เขียนแทนด้วย m

ในโครงงานนี้ค่าพารามิเตอร์ที่ใช้คือ

$$n_{tree} = 1000, n_{min_samples_split} = 34, n_{min_samples_leaf} = 16, d = 10, m = 13 \quad (26)$$

โดยใช้วิธีการเลือกพารามิเตอร์แสดงในหัวข้อที่ 10.2

3.4 ดัชนีการวัดประสิทธิภาพของการพยากรณ์

การวัดประสิทธิภาพของแบบจำลองการพยากรณ์นั้นมีหลากหลายวิธี โดยดัชนีตัวชี้วัดสมรรถนะที่นิยมใช้ในงานประยุกต์การพยากรณ์พัฒนาซึ่งอยู่ในรูปของค่าความผิดพลาดในการพยากรณ์ มีดังนี้

หมายเหตุ : ในที่นี้จะใช้สัญลักษณ์ตัวแปร x และ \hat{x} แทนค่าทั้งจริงและท่าพยากรณ์ตามลำดับ โดย x อาจแทนค่าความเข้มแสง หรือค่ากำลังผลิตไฟฟ้า

1. Root Mean Square Error (RMSE): เป็นการหาค่าเฉลี่ยของกำลังสองสัมบูรณ์ซึ่งเทียบกับ 2-นอร์มของเวกเตอร์ ค่าความผิดพลาด

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t))^2} \quad (27)$$

2. Mean Bias Error (MBE): เป็นค่าเฉลี่ยของค่าความผิดพลาด ซึ่งอาจมีค่าเป็นบวกหรือลบ เราจะใช้ดัชนีนี้เป็นการบอกว่าแบบจำลองนั้น ประเมินค่าสูงกว่าความเป็นจริง (overestimate) หรือต่ำกว่าความเป็นจริง (underestimate) ได้

$$MBE = \frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t)) \quad (28)$$

3. Normalized Root Mean Square Error (NRMSE): การใช้ดัชนี RMSE นั้นไม่ได้คำนึงถึงขนาดของค่าตัวแปร เมื่อนำดัชนีนี้ไปเปรียบเทียบกับ ข้อมูลชุดอื่นที่มีขนาดต่างกัน จึงอาจจะเปรียบเทียบไม่ได้สมเหตุสมผล ดังนั้นการ normalization แบบต่างๆ จึงได้ถูกเสนอขึ้น เพื่อทำให้สามารถเทียบสมรรถนะกับงานอื่นๆ ที่ทดสอบบนข้อมูลชุดอื่นได้

a) Normalized by the mean

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t))^2}}{\bar{x}} \times 100\% \quad (29)$$

โดยที่ $\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t)$ คือค่าเฉลี่ยของ $x(t)$

b) Normalized by a constant

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}(t) - x(t))^2}}{\text{Constant}} \times 100\% \quad (30)$$

โดยที่ Constant ในที่นี้ขึ้นอยู่กับตัวแปร x ที่ใช้ยกตัวอย่าง เช่น ในกรณีที่ x เป็นกำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ เราจะทำให้เป็นปกติด้วยค่ากำลังผลิตที่ติดตั้ง (Capacity) ส่วนกรณีที่ x เป็นความเข้มแสงอาทิตย์เราจะทำให้เป็นปกติด้วยค่าคงที่ซึ่งในรายงานนี้ค่า 1000 W/m^2

4 การจัดเตรียมข้อมูล

ในบทนี้จะอธิบายถึงที่มาของข้อมูล การประมวลผลเบื้องต้นในการตรวจสอบข้อมูลก่อนนำไปใช้ในการพัฒนาแบบจำลองการพยากรณ์ในลำดับถัดไป

4.1 ที่มาของข้อมูล

1. ข้อมูลจากตึกภาควิชาวิศวกรรมไฟฟ้า : เป็นข้อมูลที่ถูกเก็บจากแผงเซลล์แสงอาทิตย์ที่มีค่า installed capacity 8kW ในงานวิจัยนี้จะใช้ข้อมูลที่เก็บในช่วงวันที่ 1 มกราคม 2017 จนถึงวันที่ 31 ธันวาคม 2018 โดยข้อมูลวัดที่เก็บจะเป็นราย 3 นาที มีตัวแปรคือ ความเข้มรังสีดวงอาทิตย์ กำลังผลิตไฟฟ้า ความชื้นสัมพัทธ์ ความเร็วลม ดัชนีรังสีอัลตราไวโอเลต อุณหภูมิภายนอก
2. ข้อมูลจากโรงไฟฟ้านภาคกลาง : เป็นข้อมูลที่ถูกเก็บจากแผงเซลล์แสงอาทิตย์ที่มีค่า installed capacity 126 MW ในงานวิจัยนี้จะเก็บข้อมูลในช่วงวันที่ 1 มกราคม 2017 จนถึงวันที่ 31 ธันวาคม 2018 โดยข้อมูลวัดที่เก็บจะเป็นราย 5 นาที มีตัวแปรคือ ความเข้มรังสีดวงอาทิตย์ กำลังผลิตไฟฟ้า อุณหภูมิภายนอก และอุณหภูมิของแผงเซลล์แสงอาทิตย์

ข้อมูลทั้ง 2 โรงไฟฟ้าได้มาจากการฐานข้อมูล PV ของหน่วยวิจัยสมาร์ทกริด จุฬาลงกรณ์มหาวิทยาลัย [SGR]

4.2 การประมวลผลข้อมูลเบื้องต้น

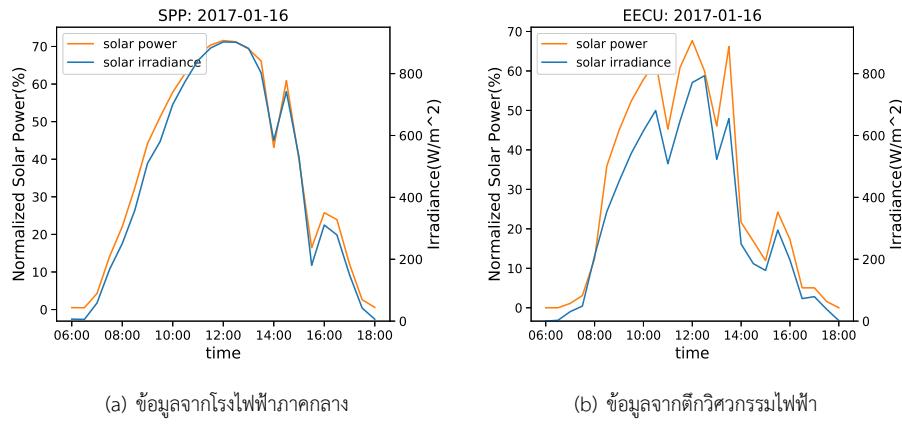
เนื่องจากชุดข้อมูลที่ใช้เป็นข้อมูลอนุกรมเวลา และมีรูปแบบเฉพาะตัวดังแสดงในรูปที่ 6 จึงมีการเสนอกระบวนการจัดเตรียมข้อมูลก่อนนำไปใช้ในการทดลอง ซึ่งประกอบไปด้วยกระบวนการดังนี้

4.2.1 การจัดการกับข้อมูลสูญหาย

วิธีการเติมข้อมูลสูญหาย เป็นวิธีการประมวลผลค่าของข้อมูลที่สูญหายจากข้อมูลที่มีอยู่ ในที่นี้กำหนดให้สัญลักษณ์ $x(t)$ แทนข้อมูล ณ เวลา t ที่สูญหาย และ $\hat{x}(t)$ เป็นค่าจากการประมวลผล ข้อมูลที่สูญหายสามารถแบ่งออกเป็น 3 ประเภทคือ

1. ข้อมูลสูญหายระยะสั้น: เป็นข้อมูลที่มีการสูญหายต่อเนื่องน้อยกว่า 30 นาที ข้อมูลที่สูญหายในกลุ่มนี้จะถูกประมวลผลด้วยการประมาณค่าในช่วงเชิงเส้น หากสมมุติว่าข้อมูลที่ขาดหายคือ $x(t+1), x(t+2), \dots, x(t+k)$ เราสามารถคำนวณค่าประมาณ $\hat{x}(t+n)$ โดยที่ $n = 1, 2, \dots, k$ ได้จาก

$$\hat{x}(t+n) = x(t) + \frac{n}{k+1} (x(t+k+1) - x(t)) \quad (31)$$



รูป 6: ตัวอย่างข้อมูลกำลังผลิตไฟฟ้าและค่าความเข้มแสงในหนึ่งวันจากโรงไฟฟ้าภาคกลางและตึกวิศวกรรมไฟฟ้า

2. ข้อมูลสูญหายระยะยาว: เป็นข้อมูลที่มีการสูญหายต่อเนื่องมากกว่า 30 นาที แต่ไม่สูญหายทั้งวัน ข้อมูลที่สูญหายในกลุ่มนี้จะถูกประมาณด้วยการใช้ค่าเฉลี่ยของวันใกล้เคียง กำหนดให้ $x^{(d)}(t)$ คือข้อมูล ณ วันที่ d เวลา t และค่าข้อมูลในวัน เวลาเดียวกัน จำนวน N วัน เราสามารถคำนวณค่าประมาณ $\hat{x}^{(d)}(t)$ ได้จาก

$$\hat{x}^{(d)}(t) = \frac{x^{(d-N)}(t) + \cdots + x^{(d-1)}(t) + x^{(d+1)}(t) + \cdots + x^{(d+N)}(t)}{2N} \quad (32)$$

3. ข้อมูลสูญหายทั้งวัน: ข้อมูลที่สูญหายในกลุ่มนี้จะถูกตัดทิ้งไม่นำมาใช้ในการทดลอง

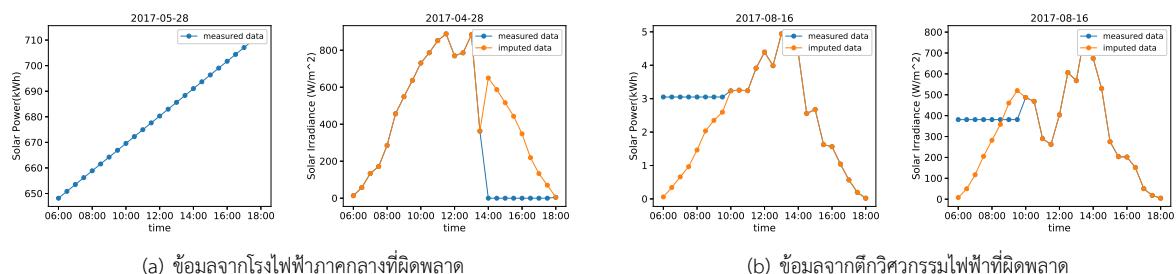
4.2.2 การลดอัตราสุ่มข้อมูล

การลดอัตราสุ่มข้อมูลเป็นการลดความถี่ในการสุ่มตัวอย่างข้อมูลเพื่อให้ได้คาบเวลาของข้อมูลตามต้องการซึ่งคือข้อมูลราย 30 นาที ในการทดลองนี้เลือกใช้วิธีการหาค่าเฉลี่ยของข้อมูลตั้งแต่ในช่วงที่ครอบคลุมครบที่ต้องการ downsample โดยสามารถคำนวณได้จาก

$$\hat{x}(t+n) = \frac{1}{n+1} \sum_{i=1}^n x(t+i) \quad (33)$$

4.2.3 การจัดการกับข้อมูลที่ผิดพลาด

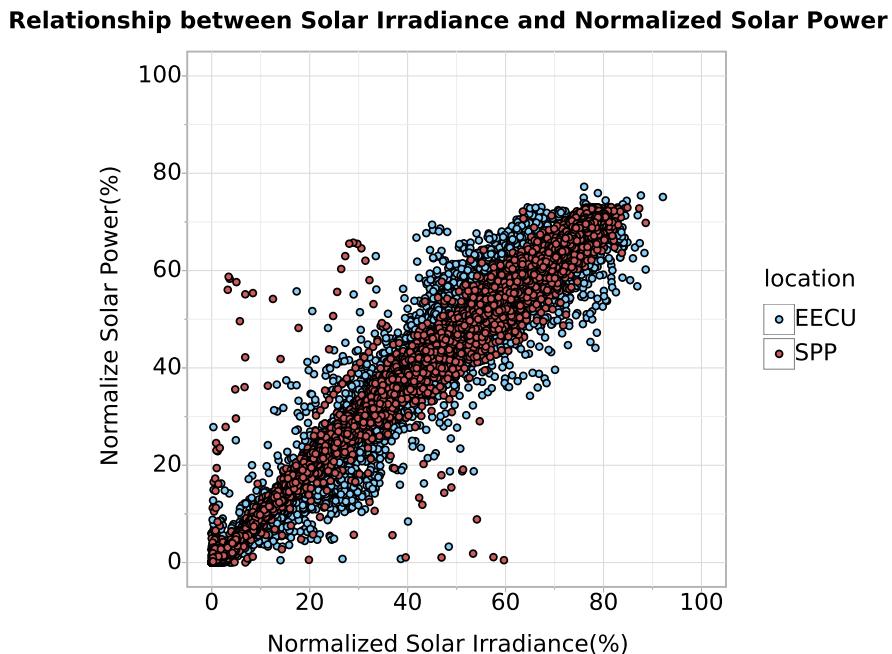
การจัดการกับข้อมูลที่ผิดพลาดเป็นการวิเคราะห์และคัดกรองข้อมูลเบื้องต้นก่อนนำไปใช้ในการทดลองโดยตรวจจับข้อมูลที่มีลักษณะไม่สอดคล้องกับลักษณะเฉพาะตัวของข้อมูล อย่างเช่นข้อมูลความเข้มแสงอาทิตย์และข้อมูลกำลังผลิตไฟฟ้ามีค่าคงที่ต่อเนื่องกันในช่วงเวลากลางวัน ข้อมูลความเข้มแสงอาทิตย์มีค่าเข้าใกล้ศูนย์ตลอดทั้งวัน โดยข้อมูลที่มีลักษณะดังกล่าวจะถูกจัดการในหลักเกณฑ์เดียวกับข้อมูลที่สูญหายดังแสดงในรูปที่ 7



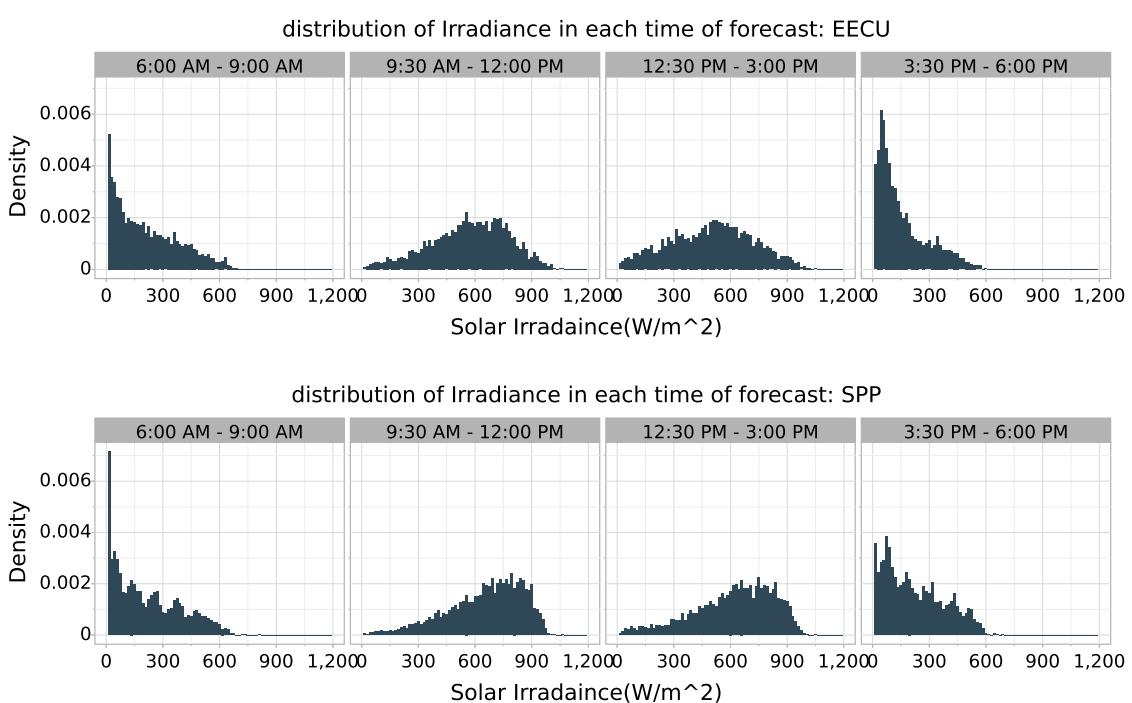
รูป 7: ตัวอย่างข้อมูลที่ผิดพลาดของกำลังผลิตไฟฟ้าและความเข้มแสงอาทิตย์จากตึกวิศวกรรมไฟฟ้าและโรงไฟฟ้าภาคกลาง ข้อมูลที่ผิดพลาดเหล่านี้จะถูกจัดการด้วยหลักเกณฑ์เดียวกับข้อมูลที่สูญหายกล่าวคืออ้างข้อมูลผิดพลาดในบางช่วงเวลาจะเติมข้อมูลด้วยค่าเฉลี่ยของวันใกล้เคียงที่เวลาันนๆ ส่วนข้อมูลที่ผิดพลาดทั้งวันจะตัดข้อมูลส่วนนั้นทิ้ง

4.3 การวิเคราะห์ลักษณะของข้อมูลเบื้องต้น

ในส่วนนี้จะวิเคราะห์ลักษณะพื้นฐานของข้อมูลจากทั้ง 2 สถานที่ โดยพิจารณาเฉพาะข้อมูลในช่วงกลางวันกล่าวคือช่วงเวลา 6:00-18:00 นาฬิกา เท่านั้น จากรูปที่ 9 และตารางที่ 2 จะเห็นว่าข้อมูลความเข้มแสงอาทิตย์ในแต่ละช่วงเวลา มีการกระจายตัวที่ต่างกัน สังเกตจากค่าส่วนเบี่ยงเบนมาตรฐานและความเบี้ยวของข้อมูล โดยสามารถแบ่งข้อมูลตามลักษณะการกระจายตัวได้เป็น 3 ช่วงเวลา ได้แก่ช่วงเช้า กลางวันและช่วงเย็น นอกจากนี้ข้อมูลจากสถานที่ต่างกันยังมีการกระจายตัวที่ต่างกัน และมีลักษณะการเฉพาะในการผลิตกำลังไฟฟ้าจากความเข้มแสงอาทิตย์ที่ต่างกันอีกด้วย โดยข้อมูลจากศึกษาวิชาวิศวกรรมไฟฟ้าที่ค่าความเข้มแสงหนึ่งๆ ค่ากำลังผลิตไฟฟ้ามีการกระจายที่มากกว่าข้อมูลจากโรงไฟฟ้าภาคกลาง ดังแสดงในรูปที่ 8



รูป 8: ความสัมพันธ์ระหว่างกำลังผลิตไฟฟ้าและความเข้มแสงอาทิตย์ของข้อมูลจากโรงไฟฟ้าภาคกลางและศึกษาวิชาวิศวกรรมไฟฟ้า



รูป 9: การกระจายตัวของความเข้มแสงอาทิตย์ในช่วงเวลาต่างๆ

ตาราง 2: พารามิเตอร์การกระจายของความเข้มแสงอาทิตย์แต่ละช่วงเวลา

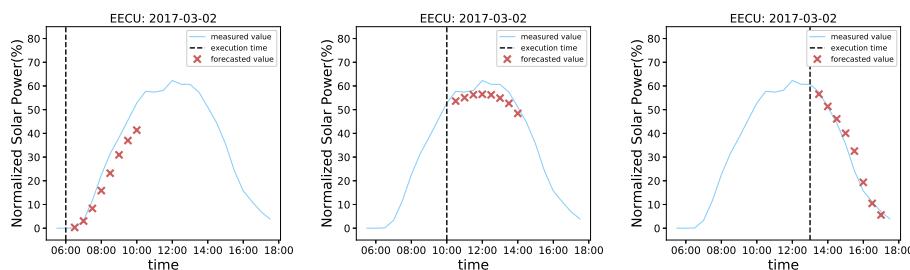
พารามิเตอร์การกระจายตัว	ข้อมูล ณ ตึกวิศวกรรมไฟฟ้า				ข้อมูล ณ โรงไฟฟ้าภาคกลาง			
	ช่วงเช้า	ช่วงสาย	ช่วงบ่าย	ช่วงเย็น	ช่วงเช้า	ช่วงสาย	ช่วงบ่าย	ช่วงเย็น
ส่วนเบี่ยงเบนมาตรฐาน (S.D.) [W/m ²]	172.81	206.88	221.46	131.37	179.78	199.12	211.42	156.42
ความเบี้ยงของข้อมูล (skewness)	0.773	-0.357	-0.077	1.116	0.723	-0.747	-0.639	0.547

5 แบบจำลองการพยากรณ์

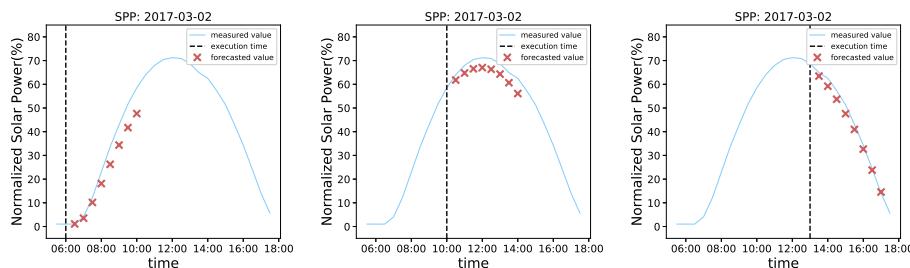
ในโครงการนี้มีการกำหนดปัญหาที่สนใจดังต่อไปนี้

การกำหนดปัญหา

ในโครงการนี้เราต้องการพยากรณ์ในช่วงเวลาตั้งแต่ 6:00 น. ถึง 17.30 น. ล่วงหน้า 4 ชั่วโมง (ค่าพยากรณ์มีความล่วงอีกด้วย 30 นาที) กล่าวคือจะพยากรณ์ 30, 60, 90, ..., 240 นาทีล่วงหน้า) ค่าพยากรณ์จะถูกคำนวณในช่วงเวลา 5:30 น. ถึง 17.00 น. ด้วยความถี่การพยากรณ์ 30 นาทีตั้งแต่แสดงในรูปที่ 10 และรูปที่ 11 แทน --- คือเวลาพยากรณ์ซึ่งจะเลื่อนไปทุกๆ 30 นาทีและ \times คือค่าพยากรณ์ ความยากของปัญหานี้คือความผันผวนของสภาพอากาศในระหว่างวันตั้งแต่แสดงในรูปที่ 11 ซึ่งเป็นตัวอย่างวันที่สภาพอากาศทั่วไปโดยความผันผวนมักเกิดขึ้นในช่วงเวลากลางวัน เนื่องจากส่วนบีบีงบนมาตรฐานของความเข้มแสงอาทิตย์ในช่วงเวลากลางวันมีค่าสูงกว่าในช่วงเวลาเช้าและเย็นตั้งแต่แสดงในตารางที่ 2 ทำให้วันสภาพอากาศทั่วไปการพยากรณ์ในช่วงเวลากลางวันจึงเป็นปัญหาที่ยากกว่าเทียบกับการพยากรณ์ในช่วงเช้าและเย็น ขณะที่รูปที่ 10 เป็นตัวอย่างวันที่สภาพอากาศปราศจากเมฆ การพยากรณ์ความเข้มแสงอาทิตย์จึงเป็นปัญหาที่ง่ายกว่าในกรณีแรก



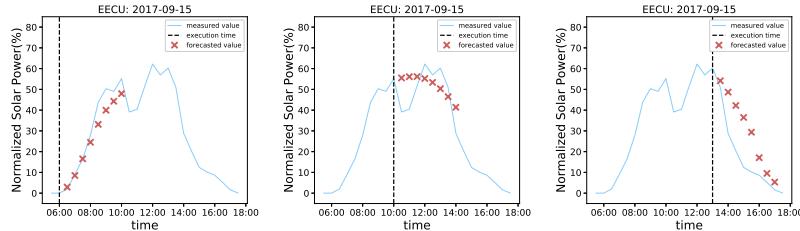
(a) การพยากรณ์ ณ ตึกวิศวกรรมไฟฟ้า



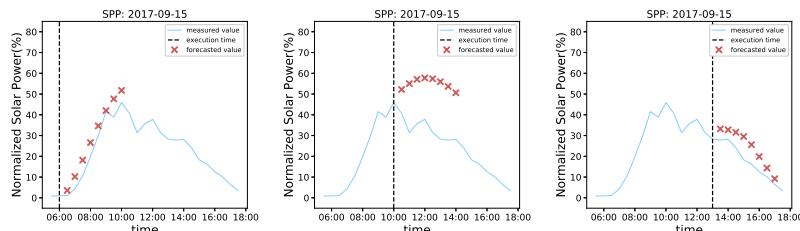
(b) การพยากรณ์ ณ โรงไฟฟ้าภาคกลาง

รูป 10: ตัวอย่างรูปแบบการพยากรณ์ทางตรงของแบบจำลอง Random forest ณ วันท่องฟ้าใส

ต่อไปเราจะนำเสนอรูปแบบการทดลองและขั้นตอนที่ใช้พยากรณ์ค่ากำลังผลิตไฟฟ้า โดยแบ่งวิธีการออกเป็น 2 วิธีคือ การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์ และการพยากรณ์กำลังผลิตไฟฟ้าโดยตรง ตั้งแต่แสดงในรูปที่ 12 กล่าวคือวิธีแรกเป็นการพยากรณ์ความเข้มแสงอาทิตย์ก่อนจากนั้นจึงใช้แบบจำลองอีกส่วนหนึ่งแปลงค่าความเข้มแสงอาทิตย์ที่พยากรณ์ได้ไปเป็นกำลังผลิตไฟฟ้า ส่วนอีกวิธีเป็นการพยากรณ์ค่ากำลังผลิตไฟฟ้าโดยตรง ซึ่งในแต่ละวิธีจะมีตัวแปรต้นที่ใช้แตกต่างกันดังที่จะกล่าวในส่วนถัดไป และเทคนิคการพยากรณ์ จะมีเทคนิคที่เป็นตัวเปรียบเทียบได้แก่ Linear regression, MARS และ ANN และมี SVR, RF เป็นแบบจำลองที่นำเสนอด้วย

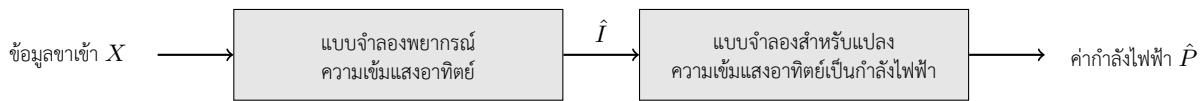


(a) การพยากรณ์ ณ ตึกภาควิชาวารมณ์ไฟฟ้า



(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 11: ตัวอย่างรูปแบบการพยากรณ์ทางตรงของแบบจำลอง Random forest ณ วันสภาพอากาศทั่วไป

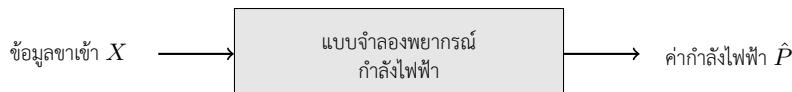


แบ่งออกเป็น 4 วิธีการพยากรณ์

- Linear regression
- MARS
- SVR
- Random forest

แบบจำลองถูกโดยแบบพหุนาม

(a) การพยากรณ์ค่ากำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์



แบ่งออกเป็น 4 วิธีการพยากรณ์

- Linear regression
- MARS
- SVR
- Random forest

(b) การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง

รูป 12: ขั้นตอนการพยากรณ์กำลังผลิตไฟฟ้า

5.1 แบบจำลองพยากรณ์ความเข้มแสงอาทิตย์

ในส่วนนี้จะนำเสนอแบบจำลองสำหรับพยากรณ์ความเข้มแสงอาทิตย์ ซึ่งเราระบุพยากรณ์ค่าความเข้มแสงอาทิตย์ 30, 60, 90, ..., 240 นาที ล่วงหน้าเพียงแทนด้วย $\hat{I}(t+1), \hat{I}(t+2), \dots, \hat{I}(t+8)$ ตัวแปรต้นทั้งหมดมีดังนี้

1. ความเข้มแสงอาทิตย์ในอดีตประกอบด้วย

- ค่าในระหว่างวันเดียวกัน $I(t), I(t-1), \dots, I(t-7)$
- ค่าในวันก่อนหน้า ณ เวลาที่ต้องการพยากรณ์ $I^{(d-1)}(t+1), I^{(d-1)}(t+2), \dots, I^{(d-1)}(t+8)$
- ค่าเฉลี่ยเคลื่อนที่แบบยกกำลัง $I_{ema}^{(d)}(t) = \alpha I^{(d)}(t) + (1-\alpha)I^{(d)}(t-1), \alpha = 0.8$
- ค่า ณ สภาวะท้องฟ้าใส $I_{clr}(t+1), I_{clr}(t+2), \dots, I_{clr}(t+k)$

2. โคง์เช่นของมุ่งระบห่วงดวงอาทิตย์กับแนวตั้งจากพื้นโลกประกอบด้วย

- ค่าณ เวลาที่ต้องการพยากรณ์ $\cos(\theta(t+1)), \cos(\theta(t+2)), \dots, \cos(\theta(t+8))$

โดยพยากรณ์ตั้งแต่เวลา 5:30 - 17:00 น ทุกๆ 30 นาที มีโครงสร้างแบบจำลองดังนี้

Linear regression และ Multivariate adaptive regression splines

แบบจำลองทั้ง 2 แบบแรกจัดทำขึ้นเพื่อเป็นแบบจำลองฐาน (baseline model) สำหรับเปรียบเทียบสมรรถนะกับแบบจำลอง SVR และ RF ใช้ตัวแปรตันดังนี้

1. ค่าความเข้มแสงอาทิตย์ในอดีตประกอบด้วย

- $I(t), I(t-1), \dots, I(t-7)$
- $I^{(d-1)}(t+1), I^{(d-1)}(t+2), \dots, I^{(d-1)}(t+8)$

2. ค่าความเข้มแสงอาทิตย์ในสภาวะท้องฟ้าใสประกอบด้วย

- $I_{clr}(t+1), I_{clr}(t+2), \dots, I_{clr}(t+8)$

Artificial neural network

แบบจำลองนี้เป็นแบบจำลองที่มีวิจัยสามารถทกริด จุฬาลงกรณ์มหาวิทยาลัยได้จัดทำขึ้นในรายงานนี้จะนำผลลัพธ์ของแบบจำลองนี้มาร่วมใช้ในการเปรียบเทียบสมรรถนะของแบบจำลองด้วย โดยแบบจำลอง ANN มีตัวแปรตันประกอบด้วย

1. ความเข้มแสงอาทิตย์ในอดีต

- ค่าในระหว่างวันเดียวกัน $I(t)$
- ค่าเฉลี่ยเคลื่อนที่ระหว่างวันแบบยกกำลัง $I_{ema*}^{(d-1)}(t+1), I_{ema*}^{(d-1)}(t+2), \dots, I_{ema*}^{(d-1)}(t+7)$
โดยที่ $I_{ema*}^{(d-1)}(t+k) = \alpha I^{(d-1)}(t+k) + (1-\alpha)I^{(d-2)}(t+k)$, $\alpha = 0.9$

2. ค่ากำลังผลิตไฟฟ้าในอดีต $P(t)$

3. ค่าอุณหภูมิในอดีต $T(t)$

Support Vector Regression และ Random forest

จากการวิเคราะห์ข้อมูลในหัวข้อที่ 4 ซึ่งพบการกระจายตัวที่แตกต่างกันของค่าความเข้มแสงอาทิตย์ในแต่ละช่วงเวลาซึ่งหมายความว่าในช่วงเวลาเช้า, กลางวัน, เย็น ความเข้มแสงอาทิตย์มีความผันผวนและการเปลี่ยนแปลงที่แตกต่างกัน ดังนั้นเราจึงเสนอการแบ่งแบบจำลองออกเป็น 3 ส่วนย่อย แยกตามเวลาของค่าพยากรณ์ด้วยสมมติฐานที่ว่าในช่วงเวลาเช้า, กลางวัน, เย็น ความเข้มแสงอาทิตย์มีลักษณะเฉพาะที่แตกต่างกัน โดยแต่ละส่วนย่อยของแบบจำลองจะใช้ตัวแปรตันที่แตกต่างกันดังนี้สำหรับ k หนึ่งๆ (โดย k แทนการพยากรณ์ $k-step$)

1. แบบจำลองสำหรับช่วงเวลา 6:00 - 9:00 น. คือ ช่วงเวลาเช้าซึ่งค่าความเข้มแสงอาทิตย์ในช่วงเวลาจะมีค่าต่ำ และมีความแปรปรวนที่น้อย ดังนั้นตัวแปรตันที่ใช้จะประกอบด้วย

$$I(t), I^{(d-1)}(t+k), \cos(\theta(t+k)), I_{clr}(t+k)$$

โดยมีสมมติฐานว่าค่า $I(t+1)$ (ค่าพยากรณ์) จะประไปตามแนวโน้มของมุ่งแสงอาทิตย์เป็นส่วนใหญ่

2. แบบจำลองสำหรับช่วงเวลา 9:30 - 15:00 น. คือ ช่วงเวลากลางวันซึ่งค่าความเข้มแสงอาทิตย์ในช่วงเวลาจะมีความแปรปรวนที่มากและมีความไม่แน่นอน ตัวแปรตันที่ใช้จะประกอบด้วย

$$I(t), I(t-1), \dots, I(t-7), I^{(d-1)}(t+k), I_{ema}(t), \cos(\theta(t+k)), I_{clr}(t+k)$$

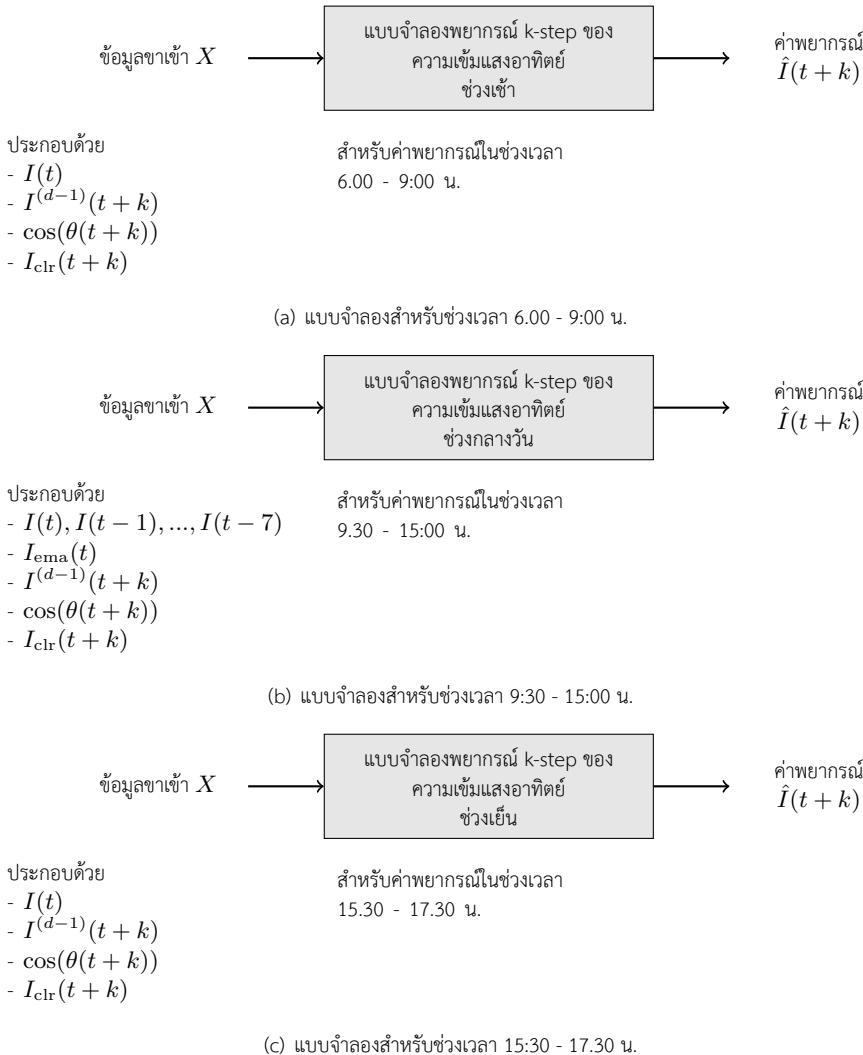
3. แบบจำลองสำหรับช่วงเวลา 15:30 - 17:30 น. คือ ช่วงเวลาเย็นซึ่งค่าความเข้มแสงอาทิตย์มีค่าต่ำใกล้ศูนย์ 0 และมีความแปรปรวนที่น้อย ดังนั้นจึงเลือกตัวแปรตันเช่นเดียวกับ แบบจำลองในเวลาเช้าประกอบด้วย

$$I(t), I^{(d-1)}(t+k), \cos(\theta(t+k)), I_{clr}(t+k)$$

จะเห็นว่าตัวแปรที่แบบจำลองช่วงกลางวันแตกต่างจากช่วงเช้าและเย็นคือ $I(t-1), \dots, I(t-7)$ และ $I_{ema}(t)$ เนื่องจากในช่วงเวลากลางวันความเข้มแสงอาทิตย์มีความผันผวนสูง ซึ่งตัวแปรข้างต้นจะช่วยอธิบายความเปลี่ยนแปลงในระหว่างวัน ส่วนตัวแปร

$I^{(d-1)}(t+k)$ เป็นตัวแปรที่อธิบายพฤติกรรมในอดีตย้อนหลัง 1 วันของค่าพยากรณ์ และ $I_{clr}(t+k)$ เป็นตัวแทนของค่าความเข้มแสงอาทิตย์ที่สภาวะท้องฟ้าใสซึ่งตัวแปรทั้งสองนี้จะมีความเข้มแสงอาทิตย์ที่เวลาต่อๆ กัน จึงใช้เป็นตัวแปรตันหลักในการพยากรณ์ทั้งสามช่วงเวลา

โครงสร้างของแบบจำลองข้างต้นแสดงดังรูปที่ 13 สำหรับแต่ละแบบจำลองช่วงเวลา (เช้า, กลางวัน, เย็น) แบบจำลองที่ k จะหมายถึงการพยากรณ์ $k-step$ ล่วงหน้า ดังนั้นวิธีการพยากรณ์หนึ่งๆ (เช่น SVR) จะใช้แบบจำลองทั้งหมด $3 \times 8 = 24$ แบบจำลอง



รูป 13: แบบจำลองพยากรณ์ k -step ของความเข้มแสงอาทิตย์แยกตามช่วงเวลา

5.2 แบบจำลองสำหรับแปลงความเข้มแสงอาทิตย์เป็นกำลังไฟฟ้า

ในการคำนวณค่ากำลังผลิตไฟฟ้าจากความเข้มแสงอาทิตย์ เราต้องคำนึงถึงตัวแปรที่มีผลต่อการคำนวณ เช่น จำนวนชั่วโมงที่แสงแดดส่อง direct บนพื้นผิว จำนวนชั่วโมงที่มีเมฆบดบัง direct แสง และอุณหภูมิที่มีผลต่อการเปลี่ยนแปลงของค่ากำลังผลิตไฟฟ้า ดังนั้นเราต้องคำนึงถึงตัวแปรที่มีผลต่อการคำนวณค่ากำลังผลิตไฟฟ้า เช่น อุณหภูมิ จำนวนชั่วโมงที่มีเมฆบดบัง และจำนวนชั่วโมงที่ไม่มีเมฆบดบัง

ในโครงงานฉบับนี้ใช้แบบจำลองความสัมพันธ์แบบไม่เป็นเชิงเส้นระหว่างค่ากำลังผลิตไฟฟ้าและค่ารังสีดิจิตอลที่ได้จากการสำรวจ ซึ่งเราใช้วิธีการลดด้อยเชิงเส้นแบบขั้นตอน (Stepwise linear regression) ในการเลือกโครงสร้างของแบบจำลอง โดยมีตัวแปรตัวเดียวที่มีผลต่อค่ากำลังผลิตไฟฟ้า คือ อุณหภูมิ และจำนวนชั่วโมงที่ไม่มีเมฆบดบัง ซึ่งเราสามารถเขียนแบบจำลองได้ดังนี้

- แบบจำลองที่ใช้ค่ารังสีดิจิตอลที่เพียงอย่างเดียว

$$P(I) = a_1 I + a_2 I^2 + a_3 I^3 \quad (34)$$

- แบบจำลองที่ใช้ค่ารังสีดิจิตอลที่เพียงอย่างเดียว

$$P(I, T) = b_1 I + b_2 I^2 + b_3 I^3 + b_4 IT \quad (35)$$

ในการคำนวณค่ากำลังผลิตไฟฟ้า เราต้องคำนึงถึงตัวแปรที่มีผลต่อค่ากำลังผลิตไฟฟ้า เช่น อุณหภูมิ จำนวนชั่วโมงที่ไม่มีเมฆบดบัง และจำนวนชั่วโมงที่มีเมฆบดบัง ซึ่งเราสามารถเขียนแบบจำลองได้ดังนี้

5.3 แบบจำลองพยากรณ์กำลังผลิตไฟฟ้าโดยตรง

ในส่วนนี้จะนำเสนอแบบจำลองสำหรับพยากรณ์กำลังผลิตไฟฟ้าโดยตรง ซึ่งเราระบุพยากรณ์ค่ากำลังผลิตไฟ 30, 60, 90, ..., 240 นาทีล่วงหน้าเขียนแทนด้วย $\hat{P}(t+1), \hat{P}(t+2), \dots, \hat{P}(t+8)$ ตัวแปรต้นที่ใช้ประกอบด้วยกลุ่มตัวแปร เช่นเดียวกับการพยากรณ์ความเข้มแสงอาทิตย์และมีตัวแปรที่เพิ่มขึ้นประกอบด้วย

- กำลังผลิตไฟฟ้าในอดีตระหว่างวันเดียวกัน $P(t), P(t-1), \dots, P(t-7)$
- กำลังผลิตไฟฟ้าในวันก่อนหน้า ณ เวลาที่ต้องการพยากรณ์ $P^{(d-1)}(t+1), P^{(d-1)}(t+2), \dots, P^{(d-1)}(t+8)$
- ค่าเฉลี่ยเคลื่อนที่แบบยกกำลังของกำลังผลิตไฟฟ้า $P_{\text{ema}}^{(d)}(t) = \alpha P^{(d)}(t) + (1 - \alpha)P^{(d)}(t-1)$, $\alpha = 0.8$
- อุณหภูมิของเซลล์แสงอาทิตย์ $T(t)$

โดยพยากรณ์ตั้งแต่เวลา 5:30 - 17:00 น. ทุกๆ 30 นาที มีโครงสร้างแบบจำลองดังนี้

Linear regression และ Multivariate adaptive regression splines

แบบจำลองทั้ง 2 แบบแรกจัดทำขึ้นเพื่อเป็นแบบจำลองฐาน (baseline model) สำหรับเปรียบเทียบสมรรถนะกับแบบจำลอง SVR และ RF ใช้ตัวแปรต้นคือดังนี้

1. ค่ากำลังผลิตไฟฟ้าและความเข้มแสงอาทิตย์ในอดีตประกอบด้วย
 - $I(t), I(t-1), \dots, I(t-7)$
 - $I^{(d-1)}(t+1), I^{(d-1)}(t+2), \dots, I^{(d-1)}(t+8)$
 - $P(t), P(t-1), \dots, P(t-7)$
 - $P^{(d-1)}(t+1), P^{(d-1)}(t+2), \dots, P^{(d-1)}(t+8)$
2. ค่าความเข้มแสงอาทิตย์ในสภาวะท้องฟ้าใสประกอบด้วย
 - $I_{\text{clr}}(t+1), I_{\text{clr}}(t+2), \dots, I_{\text{clr}}(t+8)$

Artificial neural network

แบบจำลองนี้เป็นแบบจำลองที่พิเศษ สามารถคำนวณหาวิทยาลัยได้จัดทำขึ้น ในรายงานนี้จะนำผลลัพธ์ของแบบจำลองนี้มาร่วมใช้ในการเปรียบเทียบสมรรถนะของแบบจำลองด้วย โดยแบบจำลอง ANN มีตัวแปรต้นประกอบด้วย

1. ความเข้มแสงอาทิตย์ในอดีต
 - ค่าในระหว่างวันเดียวกัน $I(t)$
 - ค่าเฉลี่ยเคลื่อนที่ระหว่างวันแบบยกกำลัง $I_{\text{ema}*}^{(d-1)}(t+1), I_{\text{ema}*}^{(d-1)}(t+2), \dots, I_{\text{ema}*}^{(d-1)}(t+7)$
โดยที่ $I_{\text{ema}*}^{(d-1)}(t+k) = \alpha I^{(d-1)}(t+k) + (1 - \alpha)I^{(d-2)}(t+k)$, $\alpha = 0.9$
2. ค่ากำลังผลิตไฟฟ้าในอดีต $P(t)$
3. ค่าอุณหภูมิในอดีต $T(t)$

Support Vector Regression และ Random forest

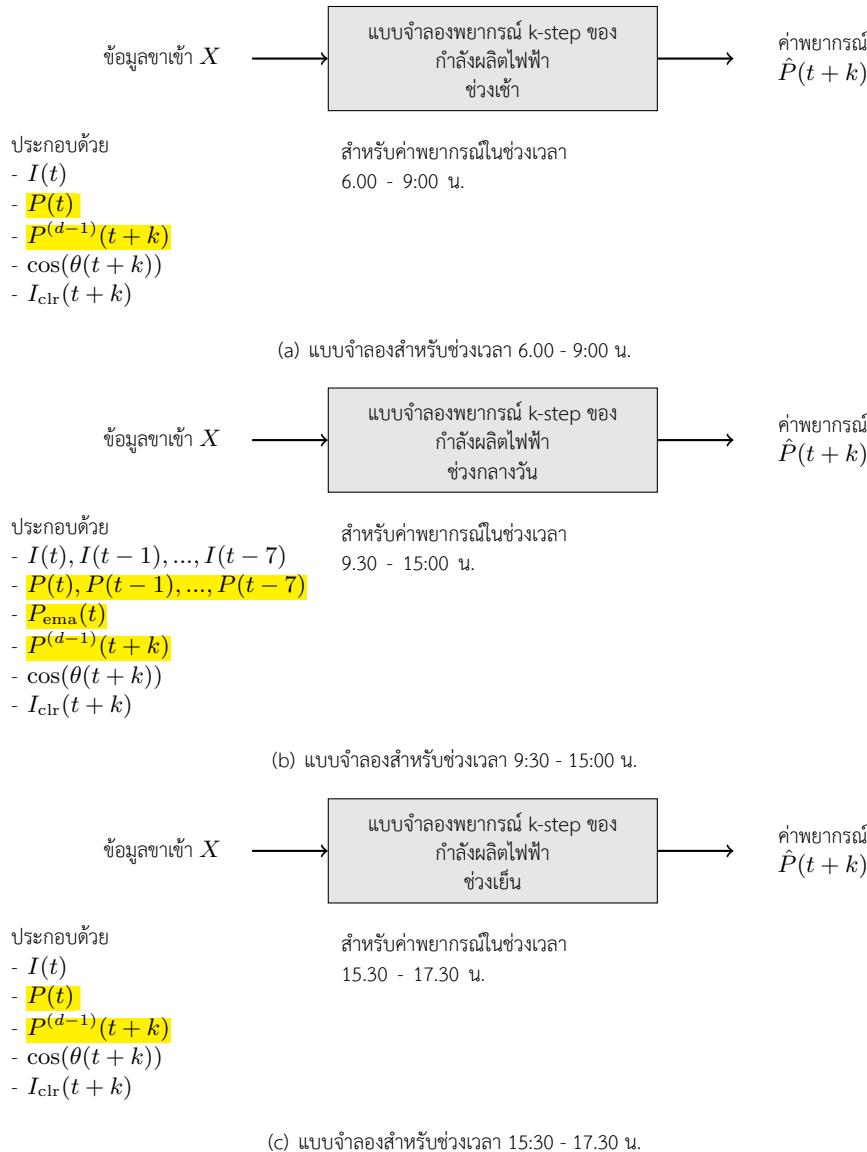
สำหรับแบบจำลองพยากรณ์กำลังผลิตไฟฟ้าโดยตรงจะใช้กลุ่มตัวแปรต้นเช่นเดียวกับแบบจำลองพยากรณ์ผ่านความเข้มแสงอาทิตย์ แต่จะเพิ่มตัวแปรที่ใช้อธิบายลักษณะของระบบผลิตไฟฟ้าคือ กำลังผลิตไฟฟ้าที่ผลิตได้ในอดีต ดังแสดงในรูปที่ 14 และด้วยเหตุผลเรื่องการกระจายตัวของข้อมูลที่แตกต่างกันในแต่ละช่วงเวลา เช่นเดียวกับการพยากรณ์ความเข้มแสงอาทิตย์ เรายังแบ่งแบบจำลองออกเป็น 3 ส่วนย่อย แยกตามเวลาของค่าพยากรณ์ โดยแต่ละส่วนย่อยของแบบจำลองจะใช้ตัวแปรต้นที่แตกต่างกันดังนี้

1. แบบจำลองสำหรับช่วงเวลา 6.00 - 9:00 น. คือ ช่วงเวลาเข้าซึ่งค่ากำลังผลิตไฟฟ้าในช่วงเวลาที่มีค่าต่ำ และมีความแปรปรวนที่น้อย ดังนั้นตัวแปรต้นที่ใช้จะประกอบด้วย $I(t), P(t), P^{(d-1)}(t+k), \cos(\theta(t+k)), I_{\text{clr}}(t+k)$
2. แบบจำลองสำหรับช่วงเวลา 9:30 - 15:00 น. คือ ช่วงเวลากลางวันซึ่งค่าความเข้มแสงอาทิตย์ในช่วงเวลาที่มีความแปรปรวนที่มากและมีความไม่แน่นอน ตัวแปรต้นที่ใช้จะประกอบด้วย $I(t), I(t-1), \dots, I(t-7), P(t), P(t-1), \dots, P(t-7), P^{(d-1)}(t+k), \cos(\theta(t+k)), I_{\text{clr}}(t+k), P_{\text{ema}}(t)$

3. แบบจำลองสำหรับช่วงเวลา 15:30 - 17:30 น. คือ ช่วงเวลาเย็นซึ่งค่าความเข้มแสงอาทิตย์มีค่าต่ำใกล้ศูนย์ 0 และมีความแปรปรวนที่น้อย ดังนั้นจึงเลือกตัวแปรต้นเช่นเดียวกับ แบบจำลองในเวลาเช้า ประกอบด้วย $I(t), P(t), P^{(d-1)}(t+k), \cos(\theta(t+k)), I_{clr}(t+k)$

จะเห็นว่าตัวแปรที่แบบจำลองช่วงกลางวันแตกต่างจากช่วงเช้าและเป็นคือ $P(t-1), \dots, P(t-7)$ และ $P_{ema}(t)$ เนื่องจากในช่วงเวลากลางวันความเข้มแสงอาทิตย์มีความผันผวนสูง ซึ่งตัวแปรข้างต้นจะช่วยอธิบายความเปลี่ยนแปลงในระหว่างวัน สรุนตัวแปร $P^{(d-1)}(t+k)$ เป็นตัวแปรที่อธิบายพุทธิกรรมในอดีตย้อนหลัง 1 วันของค่าพยากรณ์และ $I_{clr}(t+k)$ เป็นตัวแทนของค่าความเข้มแสงอาทิตย์ที่สภาวะห้องฟ้าใสซึ่งตัวแปรทั้งสองนี้สะท้อนถึงแนวโน้มค่าความเข้มแสงอาทิตย์และกำลังผลิตไฟฟ้าที่เวลานั้นๆ จึงใช้เป็นตัวแปรต้นหลักในการพยากรณ์ทั้งสามช่วงเวลา

โครงสร้างของแบบจำลองข้างต้นแสดงดังรูปที่ 14 สำหรับแต่ละแบบจำลองช่วงเวลา (เช้า, กลางวัน, เย็น) แบบจำลองที่ k จะหมายถึงการพยากรณ์ $k-step$ ล่วงหน้า ดังนั้นวิธีการพยากรณ์หนึ่งๆ (เช่น SVR) จะใช้แบบจำลองทั้งหมด $3 \times 8 = 24$ แบบจำลอง



รูป 14: แบบจำลองพยากรณ์ k -step ของกำลังผลิตไฟฟ้าแยกตามช่วงเวลา
ตัวแปรที่เน้นด้วยสีเหลือง คือตัวแปรที่แตกต่างจากแบบจำลองพยากรณ์ความเข้มแสง

6 ผลลัพธ์ของโครงการ

ในส่วนของการดำเนินงาน จะนำเสนอผลลัพธ์ในส่วนของการคัดเลือกลักษณะและผลลัพธ์การพยากรณ์จาก 4 วิธี คือ Linear regression, Multivariate adaptive regression splines, Support vector regression และ Random forest ณ ตึกภาควิชาวารมณ์ไฟฟ้า และโรงไฟฟ้าในภาคกลาง โดยแบ่งออกเป็นการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์ และการพยากรณ์ผลิตผลิตไฟฟ้าโดยตรง จำนวนจะนำ

ผลลัพธ์ที่ได้เปรียบเทียบกับผลลัพธ์จากการพยากรณ์ด้วยวิธี Artificial neural network ซึ่งทีมวิจัยระบบสมาร์ทกริด จุฬาลงกรณ์มหาวิทยาลัย ได้ทดลองโดยใช้ชุดข้อมูลชุดเดียวกัน นอกจากนี้ยังแสดงตัวอย่างผลลัพธ์การพยากรณ์ของแบบจำลอง

6.1 การคัดเลือกคุณลักษณะ

ตัวแปรต้นที่พิจารณาคือ $I(t-1), I(t-2), \dots, I(t-7), I^{(d-1)}(t+1), T(t), RH(t), UV(t), WS(t), \cos(\theta(t+1))$ เมื่อกำหนดให้ตัวแปรตามคือ $I(t+1)$

ตาราง 3: การคัดเลือกคุณลักษณะสำหรับพยากรณ์ $I(t+1)$

ตัวแปร	สหสัมพันธ์		สหสัมพันธ์แยกส่วน		การคัดโดยใช้เส้นแบบขั้นตอน	
	สัมประสิทธิ์	p-value	สัมประสิทธิ์	p-value	สัมประสิทธิ์ในสมการลดด้อย	p-value
$I(t)$	0.8956	0	0.4366	0	0.6574	10^{-34}
$I(t-1)$	0.7789	0	-0.0014	0.8675	-	-
$I(t-2)$	0.6478	0	0.0101	0.2397	-	-
$I(t-3)$	0.5018	0	-0.0172	0.0466	-0.0191	10^{-2}
$I(t-4)$	0.3610	0	-0.0099	0.2533	-	-
$I(t-5)$	0.2260	10^{-155}	-0.0344	0.0001	-0.0510	10^{-7}
$I(t-6)$	0.1039	10^{-33}	-0.0202	0.0192	-0.0288	10^{-2}
$I(t-7)$	-0.0059	0.4955	-0.0720	0	-0.0834	10^{-21}
$I^{(d-1)}(t+1)$	0.7369	0	0.1021	0	0.0876	10^{-50}
$T(t)$	0.4290	0	0.0035	0.6825	-	-
$RH(t)$	-0.1291	10^{-51}	-0.0638	0	-1.2015	10^{-17}
$UV(t)$	0.8540	0	0.1090	0	1.4957	10^{-39}
$WS(t)$	0.1388	10^{-59}	-0.0088	0.3090	-	-
$\cos(\theta(t+1))$	0.7810	0	0.0910	0	109.78	10^{-75}

หมายเหตุ : ตัวแปรที่ไม่ถูกเลือกในการสร้างสมการลดด้อยโดยใช้เส้นแบบขั้นตอนคือ

$$I(t-1), I(t-2), I(t-4), WS(t), T(t)$$

จากผลลัพธ์ทั้งตารางที่ 3 จะเห็นว่าผลลัพธ์จากการวิเคราะห์ด้วยสัมประสิทธิ์สหสัมพันธ์และสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนแตกต่างกัน เพราะสัมประสิทธิ์สหสัมพันธ์แยกส่วนเป็นการวิเคราะห์ความสัมพันธ์เชิงเส้นระหว่างตัวแปร ในขณะที่กำหนดให้ตัวแปรอื่นๆ เป็นค่าคงที่ ในการคัดเลือกคุณลักษณะที่มีนัยสำคัญเพื่อใช้ในการพยากรณ์ค่า $I(t+1)$ เรายังพิจารณา p-value จากการทดสอบนัยสำคัญทางสถิติเป็นเกณฑ์สำหรับสัมประสิทธิ์สหสัมพันธ์จะเห็นว่า p-value ของทุกๆตัวแปรมีค่าใกล้เคียงศูนย์ อย่างไรก็ด้วยเหตุผลที่กล่าวไปข้างต้น เรายังพิจารณา p-value จากสัมประสิทธิ์สหสัมพันธ์แบบแยกส่วนประกอบกัน จะพบว่ากลุ่มตัวแปรที่มี p-value ต่ำ ซึ่งหมายความว่าเป็นกลุ่มตัวแปรที่มีนัยสำคัญในการพยากรณ์ค่า $I(t+1)$ ประกอบด้วย

$$I(t), I(t-3), I(t-5), I(t-6), I(t-7), I^{(d-1)}(t+1), RH(t), UV(t), \cos(\theta(t+1)) \quad (36)$$

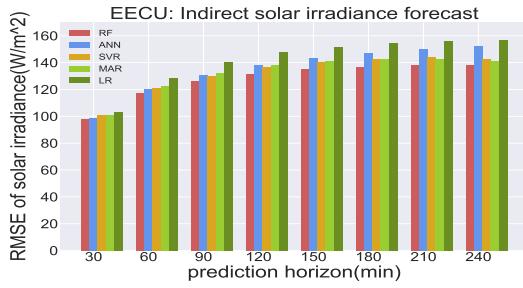
ซึ่งสอดคล้องกับผลลัพธ์จากการสร้างสมการด้วยวิธีการคัดโดยใช้เส้นแบบขั้นตอน ในขั้นตอนคัดเลือกคุณลักษณะจะได้ข้อสรุปว่าตัวแปรที่มีนัยสำคัญสำหรับการพยากรณ์ความเข้มแสลงอาทิตย์ ประกอบด้วย ความเข้มแสลงอาทิตย์ย้อนหลังในวันเดียวกัน ความเข้มแสลงอาทิตย์ย้อนหลังในวันก่อนหน้าที่เวลาเดียวกัน ความเข้มแสลงอาทิตย์ และตัวนิรังสีอัลตราไวโอเลต ส่วนตัวแปรที่มีนัยสำคัญต่ำประกอบด้วย ความเร็วลม อุณหภูมิ และความเข้มแสลงอาทิตย์ย้อนหลังในบางช่วงเวลา

ต่อไปจะนำเสนอผลลัพธ์การพยากรณ์ค่ากำลังผลิตไฟฟ้า โดยใช้ปัจจัยต่างๆ ที่เกี่ยวข้องดังที่ได้สรุปข้างต้น ที่แบ่งออกเป็น 2 แบบคือ การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสลงอาทิตย์ กล่าวคือ เราจะพยากรณ์ความเข้มแสลงอาทิตย์ก่อนจากนั้นจึงใช้แบบจำลองระบบผลิตกำลังไฟฟ้าในการแปลงค่าพยากรณ์ความเข้มแสลงอาทิตย์ไปเป็นค่ากำลังผลิตไฟฟ้า และการพยากรณ์ค่ากำลังผลิตไฟฟ้าโดยตรงโดยตรง

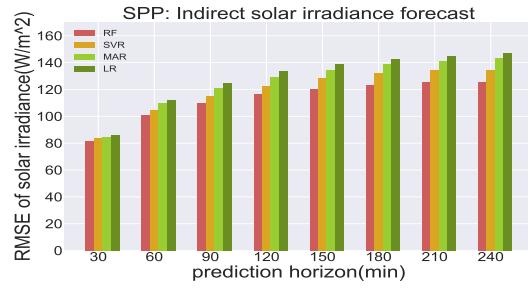
6.2 การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสลงอาทิตย์

ในส่วนนี้จะเสนอสมรรถนะของการพยากรณ์ค่าแสลงอาทิตย์ก่อน จากจึงนำเสนอสมรรถนะ RMSE และ MBE ของวิธีการพยากรณ์เมื่อแปลงค่าความเข้มแสลงอาทิตย์ไปเป็นค่ากำลังผลิตไฟฟ้า โดย RMSE จะบ่งบอกว่าค่าพยากรณ์ผิดพลาดไปจากค่าจริงมากน้อยเพียงใด ในขณะที่ MBE จะบ่งบอกว่าค่าพยากรณ์มีค่ามากหรือน้อยกว่าค่าจริงอย่างไร ซึ่งมีความสำคัญในเรื่องการบริหารจัดการพลังงาน

จากการทดลองพยากรณ์ค่าความเข้มแสลงอาทิตย์โดยใช้แบบจำลองลดด้อยเชิงเส้น MARS, SVR, Random Forest เมื่อพิจารณา RMSE และ MBE แยกตามระยะพยากรณ์ได้ผลลัพธ์ดังรูปที่ 15 และรูปที่ 16 แยกตามเวลาของค่าพยากรณ์ได้ผลลัพธ์ดังรูปที่ 17 และรูปที่ 18

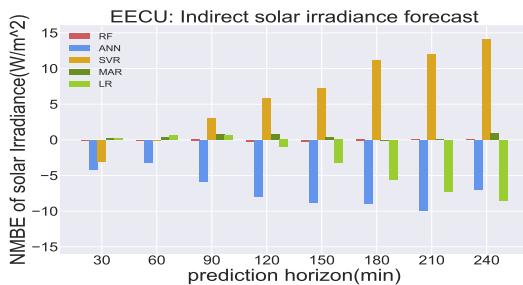


(a) การพยากรณ์ ณ ตีกภาควิศวกรรมไฟฟ้า

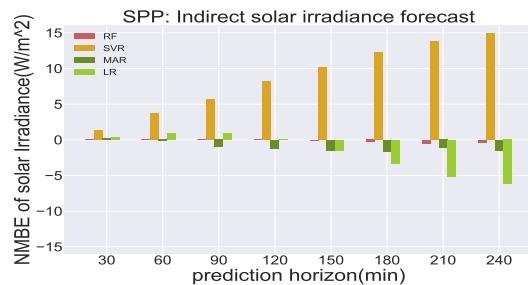


(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 15: RMSE ของความเข้มแสงอาทิตย์ในแต่ละระยะการพยากรณ์

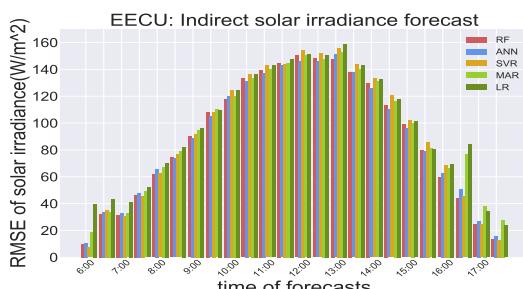


(a) การพยากรณ์ ณ ตึกภาควิชวกรรมไฟฟ้า

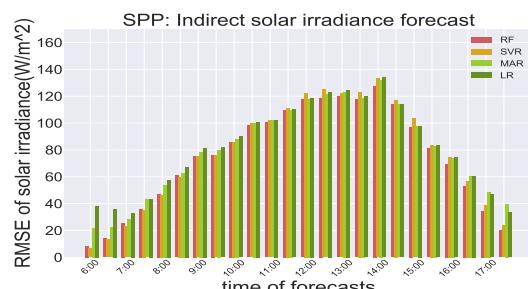


(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 16: MBE ของความเข้มแสงอาทิตย์ในแต่ละระบบการพยากรณ์



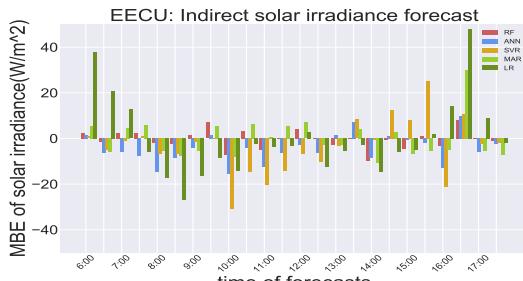
(a) การพยากรณ์ ณ ตึกภาควิชวกรรมไฟฟ้า



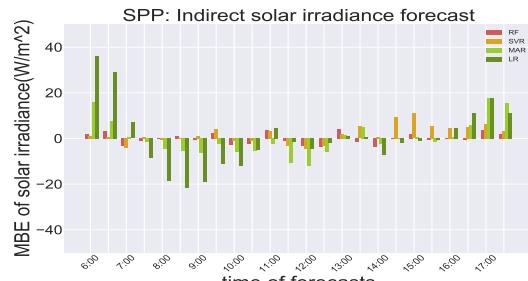
(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 17: RMSE ของความเข้มแสงอาทิตย์ในการพยากรณ์ที่แต่ละจุดเวลา

ผลของการพยากรณ์ความเข้มแสลงอาทิตย์จากทุกจุดเวลา เป็นการให้ข้อมูลสมรรถนะของแบบจำลองแต่ละชนิด เรารสามารถวิเคราะห์ค่าความผิดพลาด MBE ที่สูงถึงว่ามีการประเมินเกิน (overestimate) หรือประเมินขาด (underestimate) หรือไม่และวิเคราะห์ความผิดพลาด RMSE นี้แยกกันในแต่ละช่วงเวลา ผลการทดลองซึ่งให้เห็นว่าค่า RMSE ในช่วงกลางวันจะมีค่ามากเนื่องจากในช่วงเวลาเช้านี้มีโอกาสสูงที่ความเข้มแสลงจะมีการผันผวน (จากรูปที่ 9) และ RMSE จะมีค่าน้อยมากในช่วงเช้าเนื่องจากค่าวัดความเข้มแสลงจะต่ำหากวิเคราะห์ดู MBE ของการพยากรณ์ ณ ตึกภาควิชาระมีไฟฟ้าจะพบว่าวิธี SVR มีการประเมินขาดในเวลา 10.00 น.-12.00 น. และประเมินเกินในช่วงเวลา 14.30 น.-16.00 น. ส่วนวิธี RF ให้ค่า MBE ค่อนข้างเท่ากันทุกๆช่วงเวลาในขณะที่เมื่อสังเกตผลลัพธ์ของการพยากรณ์ ณ โรงไฟฟ้าในภาคกลางพบว่ามีค่าความผิดพลาดน้อยกว่าการพยากรณ์ ณ ตึกภาควิชาระมีไฟฟ้า ตามตารางที่ 2 จะเห็นว่าส่วนเปี่ยงเบนมาตรฐานของความเข้มแสลงอาทิตย์ใน 2 สถานที่มีความแตกต่างกัน โดยเฉพาะอย่างยิ่งในช่วงเวลากลางวันสังเกตว่าค่าส่วนเปี่ยงเบนมาตรฐานมีโรงไฟฟ้าในภาคกลางนี้ค่อนข้องกว่าความเข้มแสลงอาทิตย์ ณ ตึกภาควิชาระมีไฟฟ้า ดังนั้นการพยากรณ์ ณ โรงไฟฟ้าในภาคกลางจึงมีสมรรถนะที่ดีกว่าโดยเฉพาะในช่วงเวลากลางวันตามรูปที่ 21 และพบว่าโดยรวมแล้ววิธี SVR และ RF ให้ค่าความผิดพลาดทั้ง RMSE และ MBE ที่น้อยกว่าแบบจำลองของฐาน (Baseline) ทั้งสองแบบจำลอง



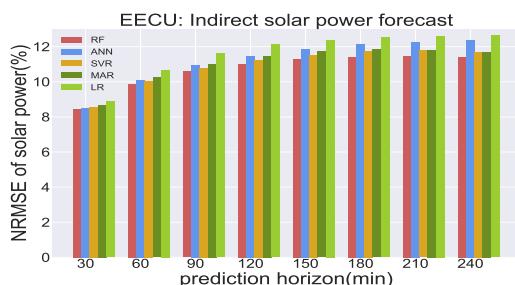
(a) การพยากรณ์ ณ ตีกภาควิเคราะห์รูปไฟฟ้า



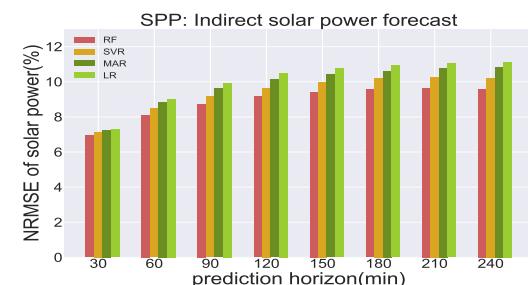
(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 18: MBE ของความเข้มแสงอาทิตย์ในการพยากรณ์ที่แต่ละจุดเวลา

เมื่อเรานำความเข้มแสงอาทิตย์ที่พยากรณ์ได้แปลงเป็นกำลังผลิตไฟฟ้าโดยใช้แบบจำลองในหัวข้อที่ 5.2 เราจะได้ผลลัพธ์สมรรถนะดังรูปที่ 19, รูปที่ 20, รูปที่ 21 และรูปที่ 22 จะเห็นว่าผลลัพธ์ที่ได้เป็นไปในทำนองเดียวกับค่าความผิดพลาดที่เกิดจากการพยากรณ์ความเข้มแสงอาทิตย์ ซึ่งชี้ให้เห็นว่าสมรรถนะในการพยากรณ์กำลังผลิตไฟฟ้าขึ้นอยู่กับสมรรถนะในการพยากรณ์ความเข้มแสงอาทิตย์กล่าวคือแบบจำลอง RF ให้ค่าความผิดพลาดทั้ง RMSE และ MBE ที่น้อยกว่าแบบจำลองอื่น

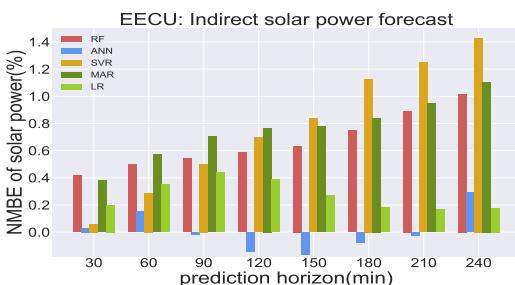


(a) การพยากรณ์ ณ ตีกภาควิเคราะห์รูปไฟฟ้า

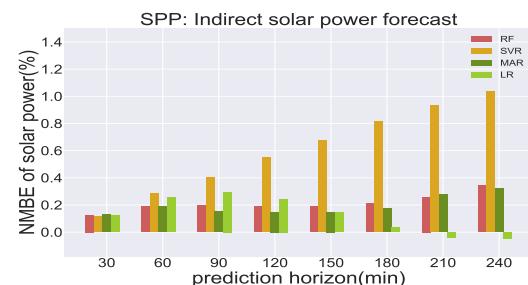


(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 19: NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์ผ่านความเข้มแสงอาทิตย์

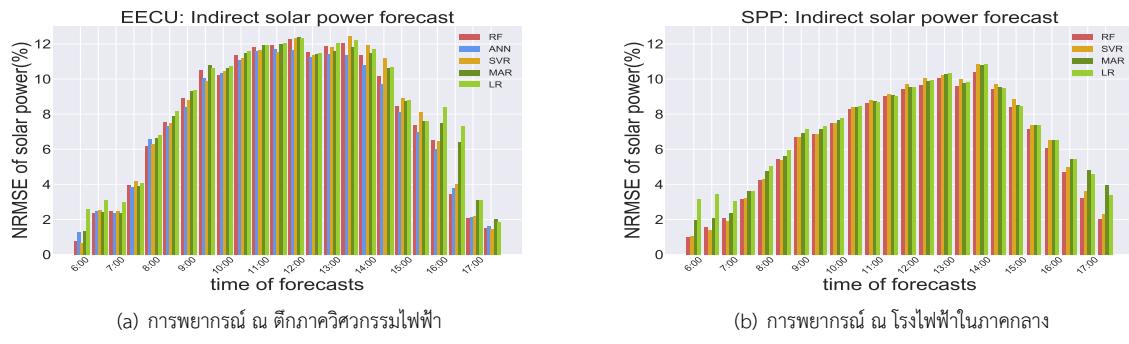


(a) การพยากรณ์ ณ ตีกภาควิเคราะห์รูปไฟฟ้า

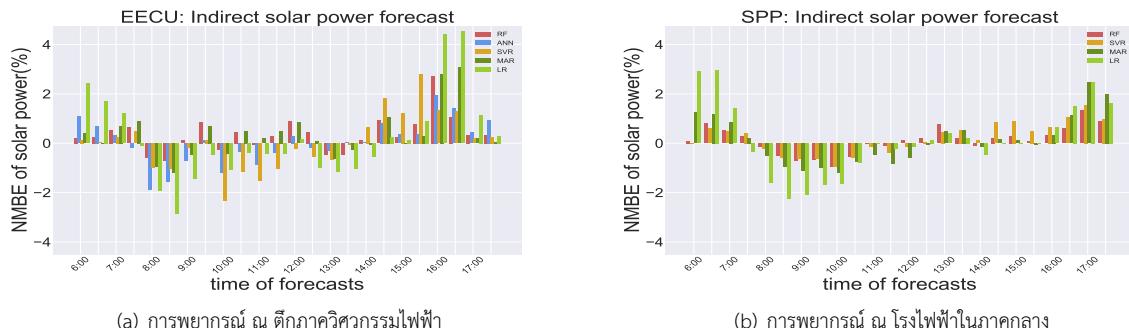


(b) การพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง

รูป 20: NMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์ผ่านความเข้มแสงอาทิตย์



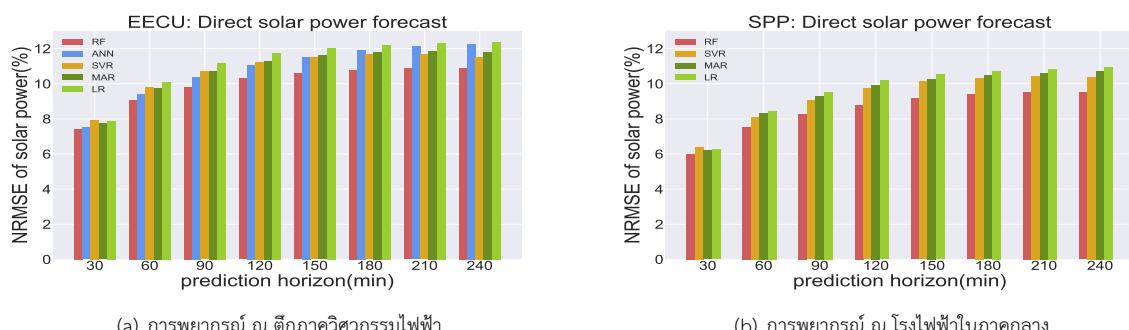
รูป 21: NRMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์ผ่านความเข้มแสงอาทิตย์ที่แต่ละจุดเวลา



รูป 22: NMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์ผ่านความเข้มแสงอาทิตย์ที่แต่ละจุดเวลา

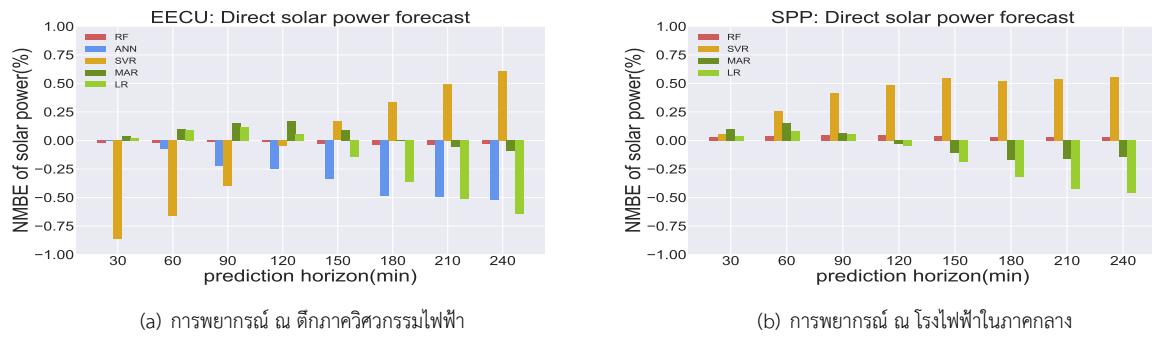
6.3 การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง

ในส่วนนี้จะแสดงผลลัพธ์สมรรถนะของวิธีการพยากรณ์ตามที่ได้นำเสนอในหัวข้อที่ 5.3 จากการทดลองพยากรณ์ค่ากำลังผลิตไฟฟ้าโดยใช้แบบจำลองถดถอยเชิงเส้น, MARS, SVR, Random Forest ได้ผลลัพธ์ดังรูปที่ 23 ถึง รูปที่ 26

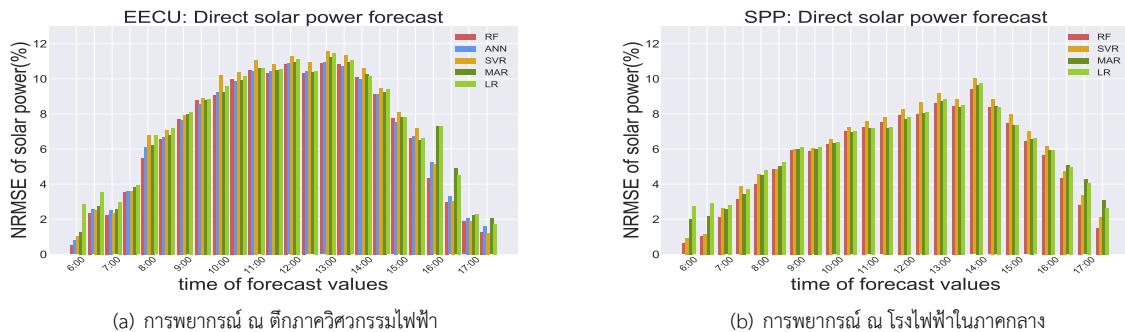


รูป 23: NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์โดยตรง

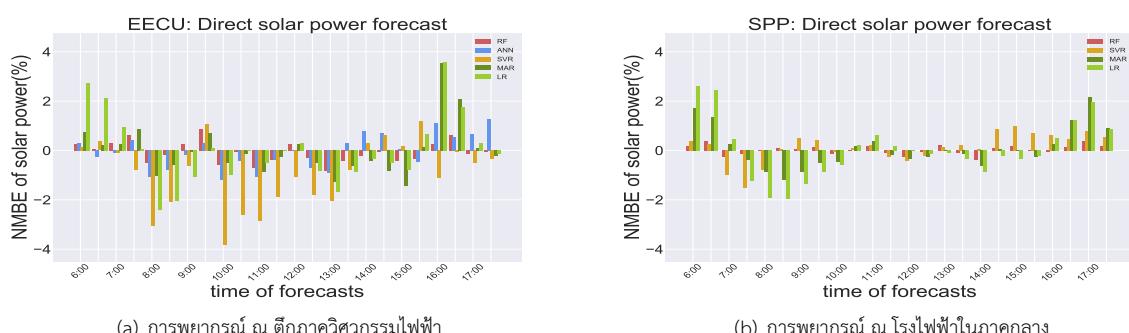
ผลการทดลองที่ให้เห็นว่า NRMSE ในช่วงกลางวันจะมีค่ามากเนื่องจากในช่วงเวลานี้ความเข้มแสงอาทิตย์จะความผันผวนสูง (จากรูปที่ 9) และในทางตรงกันข้าม NRMSE จะมีค่าน้อยในช่วงเช้าและช่วงเวลาค่ำ และจากผลลัพธ์โดยภาพรวมจะเห็นว่าการพยากรณ์กำลังผลิตไฟฟ้าโดยตรงให้ค่าความผิดพลาดที่ต่ำกว่าการพยากรณ์ผ่านความเข้มแสงอาทิตย์ซึ่งจะมีค่าความผิดพลาดเกิดเพิ่มขึ้นในส่วนของขั้นตอนการแปลงความเข้มแสงอาทิตย์เป็นกำลังผลิตไฟฟ้า



รูป 24: NMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะการพยากรณ์โดยตรง



รูป 25: NRMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์โดยตรงที่แต่ละจุดเวลา



รูป 26: NMSE ของกำลังผลิตไฟฟ้าในการพยากรณ์โดยตรงที่แต่ละจุดเวลา

6.4 ตัวอย่างผลลัพธ์การพยากรณ์

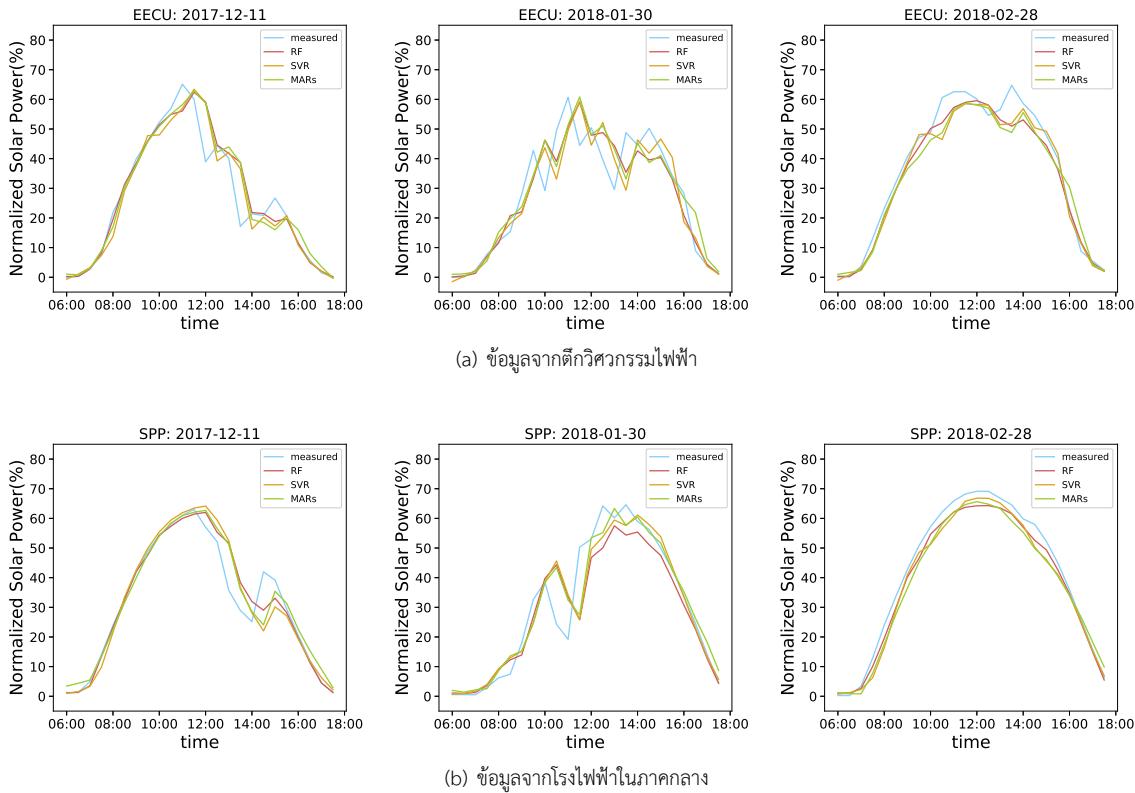
ในส่วนนี้จะแสดงตัวอย่างผลลัพธ์การพยากรณ์ทางตรงในระยะ 30 นาทีด้วยแบบจำลอง Random Forest, SVR, MARs ณ ตีกภาควิศวกรรมไฟฟ้าและโรงไฟฟ้าในภาคกลางโดยสุ่มเลือกวันมาแสดงจำนวน 3 วัน ดังแสดงในรูปที่ 27

6.5 การเปรียบเทียบความซับซ้อนในการคำนวณ

ความซับซ้อนในการคำนวณสามารถแบ่งออกได้เป็น 3 ส่วนหลักดังนี้

1. การคำนวณคุณลักษณะ (feature computation)

ความซับซ้อนส่วนนี้เกิดขึ้นในขั้นตอนการคำนวณตัวแปรขาเข้า ในโครงงานนี้คุณลักษณะที่ใช้เป็นข้อมูลขาเข้าของแบบจำลองเป็นค่าที่เก็บได้จากการวัด และอาจผ่านการคำนวณที่ไม่ซับซ้อน เช่น การหาค่าเฉลี่ยเคลื่อนที่ ดังนั้นความซับซ้อนในการคำนวณส่วนนี้จึงไม่ใช่ส่วนหลักที่ต้องพิจารณา



รูป 27: ตัวอย่างผลลัพธ์การพยากรณ์โดยตรงในระยะ 30 นาทีด้วยแบบจำลอง Random Forest, SVR, MARs

2. การฝึกแบบจำลอง (model training)

ความซับซ้อนส่วนนี้เกิดขึ้นในขั้นตอนการฝึกแบบจำลองโดยใช้ชุดข้อมูลฝึก ซึ่งเป็นการทำงานแบบอัปโหลด นั่นคือสามารถจัดเตรียมไว้ล่วงหน้าได้และจะต่างกันขึ้นกับวิธีการทำนาย/แบบจำลองที่ใช้

3. การพยากรณ์ (prediction)

ความซับซ้อนส่วนนี้เกิดขึ้นในขั้นตอนการคำนวณค่าพยากรณ์จากข้อมูลขาเข้าที่ได้รับ ซึ่งเป็นส่วนสำคัญในการนำแบบจำลองไปใช้งานจริงแบบเวลาจริง โดยเฉพาะอย่างยิ่งการพยากรณ์ในระยะสั้นมาก เพราะเราจำเป็นต้องได้ข้อมูลค่าพยากรณ์ที่รวดเร็วเพื่อใช้ในการบริหารจัดการระบบผลิตไฟฟ้า

กำหนดให้ n แทนจำนวนจุดข้อมูลทั้งหมดของชุดข้อมูลฝึก, p แทนจำนวนคุณลักษณะทั้งหมดของตัวแปรต้น เราสามารถสรุปความซับซ้อนในการคำนวณได้ดังนี้

Linear regression

ความซับซ้อนของการคำนวณในขั้นตอนฝึกขึ้นอยู่กับขั้นตอนวิธีที่ใช้ในการหาค่าตอบของสมการ ($A^T A)x = A^T y$ โดยที่ A เป็นเมตริกขนาด $n \times p$, x เป็นคอลัมน์เวกเตอร์ขนาด p , y เป็นคอลัมน์เวกเตอร์ขนาด n ซึ่งความซับซ้อนในการคำนวณมีค่าเท่ากับ $n^2 p + p^3$ ในทางปฏิบัติ $p \ll n$ จึงสรุปได้ว่าความซับซ้อนในขั้นตอนฝึกเท่ากับ $\mathcal{O}(n^2 p)$

สำหรับในขั้นตอนการพยากรณ์นั้นความซับซ้อนในการคำนวณที่เกิดขึ้นอยู่ในขั้นตอนการคำนวณผลคูณภายใน (inner product) ระหว่างคอลัมน์เวกเตอร์ของพารามิเตอร์กับคอลัมน์เวกเตอร์ของตัวแปรต้นซึ่งขนาดเท่ากับ p จะได้ความซับซ้อนในการคำนวณเท่ากับ $2p - 1$

Multivariate Adaptive Regression Splines

ความซับซ้อนของการคำนวณในขั้นตอนฝึกขึ้นอยู่กับปัจจัยหลักคือ n , p และจำนวนฟังก์ชันเชิงเส้นแบบเป็นช่วง M ที่มีค่าเลือกเป็น $M \approx 10$ ในทางปฏิบัติ โดยโครงงานฉบับนี้ใช้ขั้นตอนวิธี Fast MARS ของ Jerome Friedman [Fri93] ในการประมาณพารามิเตอร์ของแบบจำลองซึ่งสามารถแสดงได้ว่าความซับซ้อนในการคำนวณในขั้นตอนฝึกเท่ากับ $\mathcal{O}(npM^3)$ [Fri93]

สำหรับในขั้นตอนการพยากรณ์ความซับซ้อนในการคำนวณที่เกิดขึ้นอยู่ในขั้นตอนคำนวณโดยการแทนค่าตัวแปรตันลงในสมการเชิงเส้นทั้ง M สมการแล้วหาผลรวม ซึ่งมีความซับซ้อนในการคำนวณเท่ากับ $\mathcal{O}(pM)$ [Fri93]

Artificial Neural Network

กำหนดรูปแบบโครงสร้างเครือข่ายประสาทเทียมมี q ชั้นซ่อนตัว (hidden layer) และในแต่ละชั้นซ่อนตัวมี r นิวรอน (neuron) ชั้นสัญญาณเข้า (input later) มี p นิวรอน ชั้นสัญญาณออก (output later) มี t นิวรอน พิงก์ชั้นกระตุ้น (activation function) ในชั้นซ่อนและชั้นสัญญาณออก (output later) ทั้งหมดเป็นฟังก์ชัน ReLU ในการทดลองนี้ใช้ค่า $q = 5, r = 128, t = 8$

ความซับซ้อนในการคำนวณในขั้นตอนฝึกหัดอยู่กับสองส่วนหลักคือ ขั้นตอนการส่งผ่านไปข้างหน้า (forward propagation) และขั้นตอนการส่งค่าย้อนกลับ (back propagation) ซึ่งมีความซับซ้อนในการคำนวณสูงมากโดยขึ้นกับ n , รูปแบบโครงสร้างของเครือข่ายประสาทเทียม และขั้นตอนวิธีที่ใช้ในการคำนวณแต่ละรอบของการวนซ้ำ [RSO14]

สำหรับในขั้นตอนการพยากรณ์จะเห็นว่าความซับซ้อนในการคำนวณขึ้นมาจากความซับซ้อนในการคำนวณสูงมากโดยขึ้นกับ n , รูปแบบโครงสร้างของเครือข่ายประสาทในกรณีใช้ฟังก์ชันกระตุ้นเป็นฟังก์ชัน ReLU ความซับซ้อนหลักเกิดจากความซับซ้อนในส่วนการคูณเมติกซ์นำหน้า ซึ่งมีความซับซ้อนในการคำนวณเท่ากับ $\mathcal{O}(pr + r(q - 1) + rt)$

Support Vector Regression

ความซับซ้อนของการคำนวณในขั้นตอนฝึกหัดอยู่กับปัจจัยหลักคือ n โดยในการเลือกใช้ฟังก์ชันเครื่อร์เนลใดๆ ค่าความซับซ้อนของการคำนวณในขั้นตอนฝึกหัดอยู่ระหว่าง $\mathcal{O}(n^2)$ ถึง $\mathcal{O}(n^3)$ [BL07]

สำหรับในขั้นตอนพยากรณ์จาก (18) จะเห็นว่าความซับซ้อนในการคำนวณมาจากการคำนวณของเวกเตอร์สนับสนุน และการคำนวณค่าของฟังก์ชันเครื่อร์เนล ดังนั้นความซับซ้อนในการคำนวณในขั้นตอนนี้จึงขึ้นอยู่กับชนิดของฟังก์ชันเครื่อร์เนลที่เลือกใช้ สำหรับ RBF Kernel ที่เลือกใช้ในรายงานฉบับนี้มีความซับซ้อนในการคำนวณเท่ากับ $\mathcal{O}(Sp)$ [Joh99]

Random forest

ความซับซ้อนของการคำนวณในขั้นตอนฝึกหัดอยู่กับปัจจัยหลักคือ n_{tree} และ n โดยเราสามารถแสดงได้ว่าความซับซ้อนในการคำนวณในขั้นตอนฝึกกรณีที่แยกตัวกัน $\mathcal{O}(n_{\text{tree}}mn^2 \log n)$ [CM16] และเมื่อเพิ่มเงื่อนไขจำนวนระดับหรือความลึกมากสุดของต้นไม้ที่ยอมรับได้ (d) ซึ่งเป็นการลดความซับซ้อนของแบบจำลอง ความซับซ้อนของการคำนวณในกรณีที่แยกตัวกันจะมีค่าลดลงเหลือ $\mathcal{O}(n_{\text{tree}}mdn \log n)$ นอกจากนี้หากใช้วิธี bootstrap ร่วมด้วยในขั้นตอนฝึกความซับซ้อนในการคำนวณจะลดลงเหลือ $\mathcal{O}(n_{\text{tree}}md\tilde{n} \log \tilde{n})$ (\tilde{n} คือจำนวนจุดข้อมูลที่ใช้ในขั้นตอนฝึกสำหรับวิธี bootstrap โดยทั่วไป $\tilde{n} \approx 0.632n$) [CM16]

สำหรับในขั้นตอนการพยากรณ์ดังที่เห็นว่าการคำนวณค่าพยากรณ์ในแบบจำลองต้นไม้แต่ละแบบจำลองเป็นการตรวจสอบเงื่อนไขของตัวแปรตัวนั้นในแต่ละระดับความลึกของต้นไม้ (d) ดังนั้นความซับซ้อนของการคำนวณในขั้นตอนทำนาย กรณีที่แยกตัวกัน $\mathcal{O}(n_{\text{tree}}d)$

ตาราง 4: ความซับซ้อนในการคำนวณของวิธีพยากรณ์ด้วยแบบจำลองต่างๆ

วิธี	ขั้นตอนฝึก	ขั้นตอนทำนาย (ต่อหนึ่งจุดพยากรณ์)
Linear Regression	$\mathcal{O}(n^2p)$	$\mathcal{O}(p)$
Multivariate Adaptive Regression Splines	$\mathcal{O}(npM^3)$	$\mathcal{O}(pM)$
Support Vector Regression	$\mathcal{O}(n^3)$	$\mathcal{O}(Sp)$
Random forest	$\mathcal{O}(n_{\text{tree}}md\tilde{n} \log \tilde{n})$	$\mathcal{O}(n_{\text{tree}}d)$
Artificial Neural Network	-	$\mathcal{O}(pr + r(q - 1) + rt)$

7 การวิเคราะห์และวิจารณ์ผลลัพธ์ของโครงงาน

สมรรถนะของการพยากรณ์แต่ละระยะพยากรณ์

จากรูปที่ 23 จะเห็นว่าในการพยากรณ์ระยะใกล้ (30 นาทีล่วงหน้า) สมรรถนะการพยากรณ์ของทั้ง 4 แบบจำลองมีความใกล้เคียงกัน แต่เมื่อพิจารณาการพยากรณ์ที่ระยะไกลออกไปเราจะเห็นความแตกต่างของ RMSE มากขึ้นเนื่องจากในการพยากรณ์ในระยะที่ไกลขึ้นค่าความเบี่ยงเบนจะมีความผันผวนมากขึ้น ผลลัพธ์ซึ่งให้เห็นว่าแบบจำลองจะมีสมรรถนะลดลงเมื่อพยากรณ์ในระยะไกลขึ้น โดยแบบจำลอง RF มีสมรรถนะดีที่สุดทุกระยะของการพยากรณ์

นอกจากนี้รูปที่ 15 และรูปที่ 23 แบบจำลอง Random forest, SVR และ MARs มีสมรรถนะการในระยะใกล้ (ระยะ 180 นาทีล่วงหน้าเป็นต้นไป) ที่ดีกว่าแบบจำลอง ANN อย่างเห็นได้ชัดเนื่องจากการพยากรณ์ในระยะใกล้กับน้ำหนักการใช้ตัวแปรตันที่สะท้อนถึงแนวโน้มของค่าพยากรณ์ที่เวลาันนๆ ซึ่งใน Random forest, SVR และ MARs นั้นใช้ $I_{\text{clr}}(t + 1), \dots, I_{\text{clr}}(t + 8)$ ซึ่งสะท้อนค่าแนวโน้มความเข้มแสงอาทิตย์ในสภาวะท้องฟ้าใส ณ เวลาของค่าพยากรณ์เป็นหนึ่งในตัวแปรตัน ทำให้แบบจำลองในกลุ่มนี้มีสมรรถนะการพยากรณ์ในระยะใกล้ที่ดีกว่าแบบจำลอง ANN ซึ่งที่เพียง $I_{\text{ema}}^{(d-1)}(t + 1), \dots, I_{\text{ema}}^{(d-1)}(t + 8)$ ในการอธิบายค่าแนวโน้มความเข้มแสงอาทิตย์ ณ เวลาของค่า

พยากรณ์โดยอาศัยค่าจากวันที่ผ่านมา นอกจากนี้สังเกตว่าวิธี Linear regression มีสมรรถนะแย่ที่สุด ซึ่งให้เห็นถึงความสัมพันธ์แบบไม่เป็นเชิงเส้นของตัวแปรต้นและค่าพยากรณ์

สมรรถนะของการพยากรณ์แต่ละชุดเวลา

จากรูปที่ 21 และรูปที่ 22 แบบจำลองมีสมรรถนะในช่วงเวลาเข้าและเย็นดีกว่าในช่วงเวลากลางวันโดยจากการทดลองพยากรณ์ ณ ตึกภาควิชาระมีไฟฟ้า และโรงไฟฟ้าในภาคกลางให้ผลลัพธ์ตรงกันคือช่วงเวลาที่ RMSE มีค่าสูงสุดคือ ระหว่างช่วงเวลา 13.00-14.00 น. เนื่องจากเป็นช่วงเวลาที่ความเข้มแสงอาทิตย์มีความผันผวนตามรูปที่ 9 นอกจากนี้ในการพยากรณ์ ณ ตึกภาควิชาระมีไฟฟ้าด้วยวิธี Linear regression และ MARS ซึ่งไม่ได้มีการแยกแบบจำลองตามช่วงเวลา/กลางวัน/เย็น พบว่า RMSE มีค่าสูงในช่วงเวลา 16.00 - 16.30 น. เนื่องจากเป็นช่วงเวลาที่ความเข้มแสงอาทิตย์มีค่าลดลงอย่างรวดเร็วแตกต่างจากช่วงเวลาอื่นๆ แสดงให้เห็นว่าการแยกแบบจำลองตามช่วงเวลาตามวิธีที่ใช้ในแบบจำลอง SVR และ RF มีความเหมาะสม

การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง

จากตารางที่ 5 และตารางที่ 6 จะเห็นว่าวิธีการพยากรณ์กำลังผลิตไฟฟ้าโดยตรงมีสมรรถนะการพยากรณ์โดยรวมที่ดีกว่าการพยากรณ์ผ่านความเข้มแสงอาทิตย์ในทุกแบบจำลองสอดคล้องกับข้อมูลทั้ง 2 แหล่ง เนื่องจากการพยากรณ์ผ่านความเข้มแสงอาทิตย์มีการสะสมความคลาดเคลื่อนใน 2 ขั้นตอนกล่าวคือในขั้นตอนการพยากรณ์ค่าความเข้มแสง และขั้นตอนการแปลงค่าความเข้มแสงอาทิตย์เป็นกำลังผลิตไฟฟ้าต่อจาก การพยากรณ์กำลังผลิตไฟฟ้าโดยตรงตามรูปที่ 12 ซึ่งมีความคลาดเคลื่อนเกิดขึ้นเพียงขั้นตอนเดียวจากการพยากรณ์ทำให้วิธีการพยากรณ์ กำลังผลิตไฟฟ้าโดยตรงให้สมรรถนะการพยากรณ์ที่ดีกว่าการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์อย่างดังนี้

ให้ e_x แทนค่าความผิดพลาดในการประมาณตัวแปร x

สมมติให้ในขั้นตอนพยากรณ์ความเข้มแสงอาทิตย์เกิดค่าความผิดพลาดเท่ากับ e_I นั่นคือ $\hat{I} = I + e_I$ และเมื่อเราคำนวณความเข้มแสงอาทิตย์ที่พยากรณ์ได้ผ่านแบบจำลองแปลงความเข้มแสงอาทิตย์เป็นพลังงานไฟฟ้าตามสมการ

$$\hat{P}(\hat{I}) = a_1 \hat{I} + a_2 \hat{I}^2 + a_3 \hat{I}^3$$

อย่างไรก็ตามค่าพารามิเตอร์ a_1, a_2, a_3 ในแบบจำลองเป็นค่าที่ถูกประมาณโดยใช้ชุดข้อมูลฝึก นอกจากนี้โครงสร้างของแบบจำลองถูกกำหนดขึ้นอย่างง่าย และอาจมีตัวแปรอื่นๆ ที่เกี่ยวข้องกับค่ากำลังผลิตไฟฟ้าเพิ่มเติมซึ่งไม่ได้อยู่ในโครงสร้างแบบจำลองข้างต้นจึงมีค่าความผิดพลาดที่เกิดขึ้นในส่วนนี้ด้วย นั่นคือค่าความผิดพลาดจะเกิดจากค่าความผิดพลาดจากโครงสร้างแบบจำลอง (model error) และค่าความผิดพลาดจากการประมาณ (estimation error) ดังนั้นเราจึงเรียกชื่อสมการได้ว่า

$$P(\hat{I}) = (a_1 + e_{a_1}) \hat{I} + (a_2 + e_{a_2}) \hat{I}^2 + (a_3 + e_{a_3}) \hat{I}^3 + e_{\text{structure}}$$

โดยที่ $e_{\text{structure}}$ เป็นค่าความผิดพลาดที่เกิดจากการเลือกโครงสร้างแบบจำลอง

ดังนั้นความผิดพลาดในการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์จึงเขียนได้เป็น $e_{P_{\text{indirect}}} = f(e_I, e_a, e_{\text{structure}})$ ขณะที่การพยากรณ์โดยตรงความผิดพลาดในการพยากรณ์เกิดขึ้นในขั้นตอนพยากรณ์กำลังผลิตไฟฟ้าเพียงอย่างเดียว สมมติให้เกิดค่าความผิดพลาดเท่ากับ $e_{P_{\text{direct}}}$ นั่นคือ $\hat{P} = P + e_{P_{\text{direct}}}$

ผลการทดลองแสดงให้เห็นว่า $e_{P_{\text{direct}}}$ มีค่าน้อยกว่า $e_{P_{\text{indirect}}}$ สำหรับทุกๆ แบบจำลองพยากรณ์ นั่นคือวิธีการพยากรณ์กำลังผลิตไฟฟ้าโดยตรงให้สมรรถนะการพยากรณ์ที่ดีกว่าการพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์

ความซับซ้อนในการคำนวณ

ในที่นี้เราจะให้ความสำคัญกับความซับซ้อนในการคำนวณในขั้นตอนการทำนาย เนื่องจากจะเป็นส่วนที่เกิดขึ้นเมื่อเราคำนวณ ซึ่งจากผลลัพธ์ในหัวข้อที่ 6.5 พบว่าความซับซ้อนในการคำนวณในขั้นตอนพยากรณ์ของแต่ละวิธีไม่แตกต่างกันอย่างมีนักกันในเชิงการนำไปใช้งาน โดยวิธีที่มีความซับซ้อนสูงสุดคือ ANN รองลงมาคือ RF, SVR, MARS และ Linear Regression ด้วยความซับซ้อนในการคำนวณ $\mathcal{O}(pr + r(q-1) + rt), \mathcal{O}(n_{\text{tree}}d), \mathcal{O}(Sp), \mathcal{O}(pM), \mathcal{O}(p)$ ตามลำดับ

แบบจำลองที่มีสมรรถนะดีที่สุด

จากการทดลองพยากรณ์กำลังผลิตไฟฟ้า ณ ตึกภาควิชาระมีไฟฟ้าและโรงไฟฟ้าในภาคกลาง ได้ผลลัพธ์สรุปดังตารางที่ 5 และตารางที่ 6 จะเห็นว่าจากการทดลองบนข้อมูลทั้ง 2 แหล่งให้ผลลัพธ์ที่สอดคล้องกันว่าแบบจำลอง random forest มีสมรรถนะการพยากรณ์ที่ดีที่สุดในทุกระยะการพยากรณ์ โดยเมื่อพิจารณารูปที่ 18 และรูปที่ 26 ซึ่งแสดงค่า MBE และ NMSE ในแต่ละชุดเวลาของค่าพยากรณ์ของข้อมูลทั้ง 2 แหล่งจะเห็นว่าผลการพยากรณ์จากแบบจำลอง random forest นั้นมีค่าใกล้เคียงศูนย์ในทุกๆ ชุดเวลา สะท้อนให้เห็นถึงความสามารถในการจำแนกเวลาในการพยากรณ์ของแบบจำลองได้สอดคล้องกับผลการวิเคราะห์ข้อมูลในหัวข้อที่ 4 ซึ่งพบการกระจายตัวของข้อมูลที่แตกต่างกันในแต่ละช่วงเวลา ด้วยคุณสมบัตินี้ของแบบจำลองทำให้แบบจำลอง random forest มีสมรรถนะการพยากรณ์ที่ดีที่สุดแต่ต้องแลกมาด้วยความซับซ้อนในการคำนวณที่สูง เช่นกัน อย่างไรก็ตามแบบจำลอง ANN ที่มีความซับซ้อนในขั้นตอนผึกสูงสุดแต่สมรรถนะไม่ได้ไปกว่าแบบจำลอง random forest นัก ซึ่งให้เห็นถึงความซับซ้อนของแบบจำลองที่เกินความจำเป็น

ตาราง 5: NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามระยะพยากรณ์ ณ ตึกวิศวกรรมไฟฟ้า

แบบจำลองที่ใช้พยากรณ์	NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะพยากรณ์							
	$\hat{P}(t+1)$	$\hat{P}(t+2)$	$\hat{P}(t+3)$	$\hat{P}(t+4)$	$\hat{P}(t+5)$	$\hat{P}(t+6)$	$\hat{P}(t+7)$	$\hat{P}(t+8)$
LR direct method	7.876	10.085	11.169	11.762	12.048	12.229	12.317	12.396
LR indirect method	8.873	10.683	11.617	12.117	12.369	12.522	12.597	12.635
MARs direct method	7.745	9.746	10.719	11.293	11.607	11.807	11.857	11.814
MARs indirect method	8.683	10.241	11.008	11.474	11.729	11.843	11.814	11.684
SVR direct method	7.920	9.800	10.702	11.224	11.497	11.671	11.672	11.536
SVR indirect method	8.560	10.051	10.774	11.259	11.544	11.748	11.781	11.686
RF direct method	7.440	9.078	9.836	10.298	10.588	10.782	10.872	10.897
RF indirect method	8.446	9.893	10.587	11.017	11.292	11.424	11.461	11.417
ANN direct method	7.518	9.403	10.384	11.055	11.536	11.926	12.127	12.274
ANN indirect method	8.519	10.114	10.938	11.487	11.870	12.133	12.275	12.385

ตาราง 6: NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามระยะพยากรณ์ ณ โรงไฟฟ้าภาคกลาง

แบบจำลองที่ใช้พยากรณ์	NRMSE ของกำลังผลิตไฟฟ้าในแต่ละระยะพยากรณ์							
	$\hat{P}(t+1)$	$\hat{P}(t+2)$	$\hat{P}(t+3)$	$\hat{P}(t+4)$	$\hat{P}(t+5)$	$\hat{P}(t+6)$	$\hat{P}(t+7)$	$\hat{P}(t+8)$
LR direct method	6.252	8.460	9.523	10.192	10.537	10.736	10.864	10.960
LR indirect method	7.291	9.030	9.920	10.485	10.763	10.924	11.036	11.140
MARs direct method	6.196	8.320	9.315	9.932	10.267	10.480	10.628	10.711
MARs indirect method	7.242	8.851	9.615	10.127	10.423	10.626	10.773	10.842
SVR direct method	6.398	8.095	9.043	9.728	10.137	10.345	10.436	10.372
SVR indirect method	7.139	8.472	9.167	9.644	9.976	10.201	10.275	10.204
RF direct method	6.018	7.547	8.285	8.813	9.158	9.389	9.521	9.511
RF indirect method	6.931	8.120	8.719	9.159	9.406	9.556	9.631	9.572

8 บทสรุป

ในโครงการนี้เรานำเสนอการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 30, 60, 90, ..., 240 นาทีล่วงหน้า พยากรณ์ตั้งแต่เวลา 5:30-17:00 น. ทุกๆ 30 นาที ณ ตึกวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และโรงไฟฟ้าในภาคกลาง

ข้อมูลที่ใช้ในการทดลองเก็บในช่วงวันที่ 1 มกราคม พ.ศ. 2560 จนถึงวันที่ 31 ธันวาคม พ.ศ. 2561 โดยข้อมูลทั้งหมดผ่านการจัดการข้อมูลที่สูงหาย และลดอัตราสุ่มเหลือ 30 นาที

วิธีการพยากรณ์แบ่งออกเป็น 2 วิธีคือ การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์ และการพยากรณ์กำลังผลิตไฟฟ้าโดยตรง โดยเปรียบเทียบสมรรถนะของแบบจำลองพยากรณ์ซึ่งใช้เทคนิคที่แตกต่างกันทั้ง 5 แบบประกอบด้วย Linear regression, Multivariate adaptive regression splines, Artificial neural network, Support vector regression และ Random forest ทั้งในแง่ของความผิดพลาดในการพยากรณ์และความซับซ้อนในการคำนวณ ผลลัพธ์ของโครงการแบ่งออกได้ดังนี้

การคัดเลือกคุณลักษณะ

คุณลักษณะที่มีความสำคัญสำหรับการพยากรณ์ความเข้มแสงอาทิตย์ ประกอบด้วย ความเข้มแสงอาทิตย์ย้อนหลังในวันเดียวกัน, ความเข้มแสงอาทิตย์ย้อนหลังในวันก่อนหน้าที่เวลาเดียวกัน, ความชื้นสัมพัทธ์ และดัชนีรังสีอัลตราไวโอเลต ซึ่งสามารถใช้คุณลักษณะเหล่านี้ในการแบ่งออกแบบจำลองและพัฒนาวิธีการพยากรณ์ต่อไป

การเปรียบเทียบสมรรถนะของแบบจำลอง

แบบจำลองพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ทั้ง 2 วิธีคือการพยากรณ์ผ่านความเข้มแสงอาทิตย์และการพยากรณ์กำลังผลิตไฟฟ้าโดยตรง และแบ่งออกเป็น 5 เทคนิคการพยากรณ์ ผลการทดลองแสดงให้เห็นว่าการพยากรณ์แบบโดยตรงมีสมรรถนะที่ดีกว่าการพยากรณ์ผ่านความเข้มแสงอาทิตย์ในทุกๆ เทคนิคการพยากรณ์ และการพยากรณ์ในระยะใกล้มีสมรรถนะสูงกว่าการพยากรณ์ในระยะไกลซึ่งเป็นเรื่องที่สมเหตุสมผล เพราะการพยากรณ์กำลังผลิตล่วงหน้าในอนาคตที่ยาวนานกว่าอยู่มีความไม่แน่นอนที่มากกว่า นอกจากนี้ตารางที่ 7 และตารางที่ 8 สมรรถนะการพยากรณ์ในช่วงเวลาเช้าและเย็นจะสูงกว่าช่วงเวลากลางวันซึ่งเป็นช่วงเวลาที่ความเข้มแสงอาทิตย์มีการกระจายตัวสูง

การเปรียบเทียบสมรรถนะของแต่ละแบบจำลองตามเทคนิคการพยากรณ์ได้ผลลัพธ์สอดคล้องกันทั้งในการพยากรณ์ ณ ตึกวิศวกรรมศาสตร์ และโรงไฟฟ้าพลังงานแสงอาทิตย์ในภาคกลางว่าแบบจำลอง SVR และ RF มีสมรรถนะที่ดีกว่าแบบจำลองอื่นๆ รวมทั้ง ANN โดยจะเห็นความแตกต่างที่ชัดเจนในการพยากรณ์ระยะไกล

ตาราง 7: NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามเวลา ณ ตีกวิศวกรรมไฟฟ้า

เวลาของค่าพยากรณ์	การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์					การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง				
	RF	ANN	SVR	MARs	LR	RF	ANN	SVR	MARs	LR
06:00	0.779	1.309	0.663	1.317	2.587	0.510	0.797	1.021	1.287	2.864
06:30	2.382	2.484	2.534	2.393	3.111	2.363	2.560	2.524	2.752	3.533
07:00	2.465	2.354	2.472	2.389	2.979	2.244	2.518	2.331	2.559	2.985
07:30	3.973	3.846	4.179	3.926	4.083	3.546	3.613	3.614	3.820	3.925
08:00	6.187	6.585	6.307	6.622	6.814	5.508	6.134	6.778	6.244	6.766
08:30	7.534	7.310	7.493	7.890	8.195	6.565	6.676	7.092	6.789	7.214
09:00	8.895	8.384	8.798	9.333	9.369	7.711	7.666	7.934	7.973	8.120
09:30	10.52	10.04	9.888	10.78	10.59	8.790	8.561	8.903	8.813	8.835
10:00	10.21	10.31	10.44	10.59	10.76	9.062	9.249	10.23	9.258	9.591
10:30	11.34	11.09	11.22	11.48	11.56	10.00	9.857	10.40	9.954	10.17
11:00	11.83	11.59	11.66	11.92	11.95	10.52	10.42	11.05	10.58	10.63
11:30	11.92	11.71	11.52	11.99	12.07	10.32	10.43	10.84	10.49	10.54
12:00	12.28	11.65	12.34	12.36	12.34	10.83	10.87	11.28	10.95	11.10
12:30	11.52	11.22	11.35	11.40	11.47	10.34	10.41	10.93	10.35	10.45
13:00	11.85	11.42	11.79	11.58	12.05	10.87	10.92	11.58	11.22	11.47
13:30	12.02	11.34	12.42	11.84	12.23	10.85	10.74	11.37	10.96	11.04
14:00	11.37	10.78	11.93	11.50	11.70	10.12	9.988	10.61	10.25	10.17
14:30	10.18	9.692	11.18	10.62	10.68	9.148	9.148	9.492	9.250	9.394
15:00	8.476	8.116	8.920	8.719	8.784	7.743	7.550	8.075	7.832	7.828
15:30	7.383	6.998	8.105	7.598	7.608	6.639	6.743	7.204	6.518	6.647
16:00	6.495	6.033	6.476	7.505	8.425	4.336	5.271	5.121	7.316	7.323
16:30	3.462	3.783	3.997	6.422	7.329	2.974	3.293	3.048	4.886	4.525
17:00	2.093	2.128	2.187	3.122	3.076	1.865	2.062	1.891	2.243	2.290
17:30	1.481	1.649	1.470	2.037	1.873	1.260	1.631	1.221	2.044	1.713

แบบจำลองพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์

จากแบบจำลองพยากรณ์กำลังผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ทั้ง 2 วิธีคือ การพยากรณ์ผ่านความเข้มแสงอาทิตย์และการพยากรณ์กำลังผลิตไฟฟ้าโดยตรง โดยผลการทดลองแสดงให้เห็นว่า สำหรับการพยากรณ์ 30 นาทีล่วงหน้าแบบจำลอง Random Forest ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า NRMSE เท่ากับ 6.93% และ 6.02% ตามลำดับ (ผลพยากรณ์ ณ โรงไฟฟ้าในภาคกลาง) และยังเป็นแบบจำลองที่มีความซับซ้อนในขั้นตอนพยากรณ์ไม่สูงมากนักเทียบกับวิธีการอื่นๆ เช่น ANN อีกด้วย

เมื่อนำผลลัพธ์ที่ได้เปรียบเทียบกันงานในอดีตตามหัวข้อที่ 1 [VKS16] ใช้วิธี SARIMA และ [BMP13] ใช้วิธี SARIMA-SVR ในการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 1 ชั่วโมงล่วงหน้าให้ผลลัพธ์ค่า NRMSE เท่ากับ 8.12 และ 9.40% ตามลำดับซึ่งมีค่าใกล้เคียงกับผลลัพธ์จากแบบจำลองที่ดีที่สุดที่ได้นำเสนอในโครงการฉบับนี้ หรือจากการของ [XCS12] ที่ใช้วิธี SVR ร่วมกับการวิเคราะห์ความคล้ายกันของแต่ละวัน ในการพยากรณ์กำลังผลิตไฟฟ้าในระยะ 2 ชั่วโมง ให้ผลลัพธ์ค่า NRMSE เท่ากับ 9.34% เทียบกับผลลัพธ์จากการแบบจำลองที่ดีที่สุดในโครงการฉบับนี้ที่ระยะในการพยากรณ์เดียวกัน ให้ผลลัพธ์ค่า NRMSE เท่ากับ 8.81% ซึ่งมีสมรรถนะสูงกว่า อย่างไรก็ตามค่าความผิดพลาดที่เกิดขึ้นยังขึ้นอยู่กับลักษณะของข้อมูลที่ใช้ในการทดลอง เช่น การกระจายตัว และสภาพภูมิอากาศที่แตกต่างกันไปในแต่ละพื้นที่ นอกจากนี้ผลลัพธ์ที่ได้จากการทดลองสอดคล้องกับงานของ [AMR18] ที่เปรียบเทียบการใช้วิธี SVR และ RF เพื่อพยากรณ์กำลังผลิตไฟฟ้าในระยะ 1 ชั่วโมงและได้ผลลัพธ์ว่าวิธี RF มีสมรรถนะที่ดีกว่าวิธี SVR

ในการพัฒนาเทคนิคการพยากรณ์ในอนาคต เรา秧สามารถผลลัพธ์การพยากรณ์ที่ได้ไปใช้ร่วมกับผลการพยากรณ์ด้วยวิธีการอื่นๆ ซึ่งมีลักษณะเฉพาะแตกต่างกันได้ เช่น ใช้แบบจำลองซึ่งใช้ข้อมูลขาเข้าที่แตกต่างกันออกไป โดยใช้เทคนิคเทคนิคการเรียนรู้ร่วมกัน (Ensemble Method) จะทำให้ได้วิธีการพยากรณ์ที่มีสมรรถนะดียิ่งขึ้นต่อไป

9 กิตติกรรมประกาศ

โครงการฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์ด้วยความกรุณาอย่างยิ่งจาก ผศ.ดร. จิตโภนุท ส่งศรี ที่ได้สละเวลาอันมีค่าแก่คณะผู้จัดทำ เพื่อให้คำปรึกษาและแนะนำตลอดจนตรวจทานแก่ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่เป็นอย่างยิ่ง ตลอดระยะเวลาการดำเนินโครงการที่ทำโครงการฉบับนี้จนสำเร็จสมบูรณ์ลุล่วงได้ด้วยดี คณะผู้จัดทำขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี่

ตาราง 8: NRMSE ของการพยากรณ์กำลังผลิตไฟฟ้าแยกตามเวลา ณ โรงไฟฟ้าภาคกลาง

เวลาของค่าพยากรณ์	การพยากรณ์กำลังผลิตไฟฟ้าผ่านความเข้มแสงอาทิตย์				การพยากรณ์กำลังผลิตไฟฟ้าโดยตรง			
	RF	SVR	MARs	LR	RF	SVR	MARs	LR
06:00	0.988	1.035	1.949	3.169	0.620	0.922	1.999	2.726
06:30	1.541	1.416	2.070	3.437	1.047	1.143	2.189	2.927
07:00	2.071	1.886	2.366	3.035	2.147	2.626	2.557	2.804
07:30	3.176	3.222	3.617	3.624	3.129	3.907	3.430	3.697
08:00	4.262	4.278	4.767	5.029	4.006	4.567	4.506	4.819
08:30	5.431	5.388	5.634	5.976	4.847	4.867	5.046	5.256
09:00	6.694	6.691	6.919	7.146	5.936	5.993	5.983	6.105
09:30	6.836	6.884	7.141	7.293	5.877	6.031	5.990	6.100
10:00	7.463	7.517	7.664	7.773	6.272	6.580	6.318	6.411
10:30	8.301	8.419	8.425	8.440	7.028	7.246	6.936	7.010
11:00	8.614	8.793	8.751	8.709	7.273	7.571	7.176	7.219
11:30	9.045	9.141	9.085	9.039	7.514	7.823	7.178	7.267
12:00	9.452	9.732	9.529	9.529	7.937	8.273	7.704	7.831
12:30	9.654	10.05	9.882	9.940	8.006	8.656	8.037	8.124
13:00	10.04	10.24	10.30	10.34	8.640	9.180	8.742	8.862
13:30	9.594	10.00	9.793	9.814	8.434	8.860	8.414	8.510
14:00	10.41	10.86	10.78	10.85	9.386	10.02	9.645	9.765
14:30	9.418	9.716	9.521	9.463	8.370	8.867	8.421	8.408
15:00	8.379	8.853	8.499	8.435	7.492	7.995	7.362	7.378
15:30	7.173	7.398	7.382	7.352	6.435	7.001	6.591	6.606
16:00	6.083	6.549	6.540	6.526	5.677	6.194	5.929	5.950
16:30	4.688	4.993	5.420	5.432	4.363	4.768	5.061	4.993
17:00	3.237	3.624	4.809	4.601	2.828	3.359	4.267	4.065
17:30	2.015	2.320	3.974	3.396	1.520	2.134	3.096	2.633

ขอขอบคุณภาควิชาชีวกรรมไฟฟ้าที่ให้การสนับสนุนทั้งในด้านสถานที่และข้อมูลที่ใช้ในการจัดทำโครงงานฉบับนี้จากทีมวิจัยสมาร์ทกริด จุฬาลงกรณ์มหาวิทยาลัย[SGR]

ขอขอบคุณกรรมการไฟฟ้าฝ่ายผลิตแห่งประเทศไทยและโรงไฟฟ้าพลังงานแสงอาทิตย์ในภาคกลางที่ให้การสนับสนุนข้อมูลกำลังผลิตไฟฟ้า และความเข้มแสงอาทิตย์ที่ใช้ในการจัดทำโครงงานฉบับนี้

สุดท้ายนี้ขอขอบคุณที่ๆ และเพื่อนๆ ในภาควิชาชีวกรรมไฟฟ้า ที่ให้ความช่วยเหลือและกำลังใจในการจัดทำโครงงานฉบับนี้เสมอมา

เอกสารอ้างอิง

- [AMR18] W.M. Ahmad, M. Mourshed, and Y. Rezgu. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy*, 164:465–474, 2018.
- [AOE⁺16] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez de Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136:78–111, 2016.
- [BA01] S. Bernhard and S. Alexander. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
- [BEO16] W. Björn, L. Elke, and K. Oliver. Statistical learning for short-term photovoltaic power predictions. In *Computational sustainability*, pages 31–45. Springer, 2016.

- [Ber45] H. Bernhard. Insolation in relation to cloudiness and cloud density. *Journal of meteorology*, 2(3):154–166, 1945.
- [BL07] L. Bottou and C. Lin. Support vector machine solvers. *Large scale kernel machines*, 3(1):301–320, 2007.
- [BMP13] M. Bouzerdoum, A. Mellit, and A.M. Pavan. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 98:226–235, 2013.
- [CL11] C. Chang and C. Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.
- [CM16] G. Callejas and A. Miguel. The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, 326:4–12, 2016.
- [CV95] C. Corinna and V. Vladimir. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [DE12] R. Djamila and M. Ernest. *Optimization of photovoltaic power systems: modelization, simulation and control*. Springer Science & Business Media, 2012.
- [FAGJ15] M. Francisco, T. Alicia, C. Gualberto, and R. José. A survey on data mining techniques applied to energy time series forecasting. *Energies*, In press, 11 2015.
- [FHT01] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [Fri93] J. Friedman. Fast MARS. Technical report, Technical Report, 1993.
- [IPC13] R.H. Inman, H. Pedro, and C.F.M. Coimbra. Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, 39(6):535–576, 2013.
- [Jer91] F. Jerome. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.
- [Joh99] P. John. Fast training of support vector machines using sequential minimal optimization. advances in kernel methods—support vector learning (pp. 185–208). *AJ, MIT Press, Cambridge, MA*, 1999.
- [JWHT13] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to statistical learning*, volume 112. Springer, 2013.
- [MEPV12] P. Marius, P. Eugenia, G. Paul, and B. Viorel. *Weather modeling and forecasting of PV systems operation*. Springer Science & Business Media, 2012.
- [MIG16] R. Mashud, K. Irena, and A. Vassilos G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Conversion and Management*, 121:380–390, 2016.
- [PC07] A. Pansak and S. Chumnong. An assessment of the ashrae clear sky model for irradiance prediction in thailand. *Asian J. Energy Environ*, 8(02):523–532, 2007.
- [PR02] I. Pierre and P. Richard. A new airmass independent formulation for the linke turbidity coefficient. *Solar Energy*, 73(3):151–157, 2002.
- [RSO14] L. Roi, S. Shwartz, and S. Ohad. On the computational efficiency of training neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 855–863. Curran Associates, Inc., 2014.
- [SGR] Smart Grid Research Unit: SGRU. Pv measurement data set. <http://www.sgru.eng.chula.ac.th/>. Accessed on Apr 28, 2020.
- [SS04] A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

- [SSY] M. Samanta, B. Srikanth, and J. Yerrapragada. Short-term power forecasting of solar pv systems using machine learning techniques.
- [Vap99] V. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.
- [Vio97] B. Viorel. Verification of some very simple clear and cloudy sky models to evaluate global solar irradiance. *Solar Energy*, 61(4):251–264, 1997.
- [VKS16] S. Vagopoulos, G. Chouliarasand . Kardakosand, C. Simoglou, and A. Bakirtzis. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In *2016 IEEE International Energy Conference (ENERGYCON)*, pages 1–6. IEEE, 2016.
- [XCS12] R. Xu, H. Chen, and X. Sun. Short-term photovoltaic power forecasting with weighted support vector machine. In *2012 IEEE International Conference on Automation and Logistics*, pages 248–253. IEEE, 2012.
- [จ57] เสริม จันทร์ฉาย. รังสีอาทิตย์. หน่วยวิจัยพลังงานแสงอาทิตย์ ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, 2557.

10 ภาคผนวก

ในภาคผนวกนี้จะนำเสนอที่มาของค่าพารามิเตอร์ที่เลือกใช้ในแบบจำลอง Support Vector Regression และ Random Forest และเปรียบเทียบสมรรถนะของแบบจำลองที่เปลี่ยนแปลงไปเมื่อปรับค่าพารามิเตอร์ของแบบจำลอง

10.1 ผลการปรับค่าพารามิเตอร์ของแบบจำลอง Support Vector Regression

สมรรถนะของ SVR นั้นขึ้นอยู่กับชนิดของเครื่องเรียนฟังก์ชันที่เลือกใช้และพารามิเตอร์ของฟังก์ชันเครื่องเรียนนั้นๆ โดยในรายงานฉบับนี้เลือกใช้ Radial-basis function (RBF) kernel เมื่อจากการคำนวณค่า C ที่ตั้งไว้ในแบบจำลองที่ต่างกันนั้นกับการคำนวณผลคุณแบบจุดของข้อมูลขาเข้าที่อยู่ในปริภูมิมิติอนันต์ทำให้สามารถรับมือกับความสัมพันธ์ที่ไม่เป็นเส้นได้ [SS04] โดยมีพารามิเตอร์ของฟังก์ชันเครื่องเรียนดังนี้

1. สัมประสิทธิ์ของฟังก์ชันลงโทษ (C) เป็นค่าที่ควบคุมความสมดุลระหว่าง การยอมรับความคลาดเคลื่อนที่มากกว่า ε ในชุดข้อมูลฝึก และความซับซ้อนของแบบจำลอง [BA01] การปรับค่า C ให้มีค่าน้อยเป็นการยอมให้เกิดความคลาดเคลื่อนในชุดข้อมูลฝึกได้มากซึ่งสามารถนำไปสู่เกิดปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลน้อยเกินไป (Under-fitting) ในทางตรงกันข้ามการปรับค่า C ให้มีค่ามากจะเป็นการบังคับให้ฟังก์ชันวัดถูประสกนรุ่งเงินที่จะลด empirical risk ให้ต่ำที่สุดในชุดข้อมูลฝึกซึ่งสามารถนำไปสู่แบบเกิดปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลมากเกินไป (Over-fitting)
2. สัมประสิทธิ์เครอร์เนล (γ) เป็นค่าที่แสดงถึงระยะของอิทธิพลของจุดข้อมูลฝึกหนึ่งๆ โดยถ้ากำหนดให้ γ มีค่าน้อย แสดงถึงอิทธิพลของจุดข้อมูลฝึกหนึ่งๆ มีระยะใกล้ ในทางตรงกันข้ามถ้ากำหนดให้ γ มีค่ามาก แสดงถึงอิทธิพลของจุดข้อมูลฝึกหนึ่งๆ มีระยะไกล
3. ค่าความคลาดเคลื่อนที่ยอมรับได้ (ε) เป็นพารามิเตอร์ที่กำหนดขนาดของบริเวณที่ยอมให้เกิดความคลาดเคลื่อนระหว่างค่าพยากรณ์จากฟังก์ชันและค่าจริง โดยการปรับค่า (ε) สูงจะเป็นการลดความแม่นยำของการพยากรณ์ในชุดข้อมูลฝึก

ในการปรับค่าพารามิเตอร์ทั้ง 3 จะเริ่มจากกำหนดให้ค่า $\gamma = 1/p$ โดยที่ p แทนจำนวนคุณลักษณะทั้งหมดของตัวแปรต้น ตามที่มีการเสนอใน [CL11] ซึ่งในที่นี้ $\gamma = 1/11$ และค่า $\varepsilon = 0.1$ จากนั้นปรับค่า C ระหว่างช่วง 2^{-3} ถึง 2^9 จากผลในตารางที่ 9 พบว่าสำหรับการเพิ่มค่า C ในตอนต้น สมรรถนะของแบบจำลองในชุดข้อมูลตรวจสอบจะเพิ่มขึ้นอย่างมีนัยสำคัญแต่เมื่อเพิ่มค่า C ถึงค่าหนึ่งสมรรถนะแบบจำลองในชุดข้อมูลตรวจสอบจะค่อนข้างคงที่ในขณะที่สมรรถนะแบบจำลองในชุดข้อมูลฝึกยังคงเพิ่มขึ้นอย่างต่อเนื่องจึงสรุปได้ว่าการเพิ่มค่า C ต่อไปจากค่าดังกล่าวสามารถนำไปสู่ปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลมากเกินไป ดังนั้นจึงเลือกค่า $C = 128$ หลังจากนั้นปรับค่า γ ระหว่างช่วง 2^{-7} ถึง 2^2 ในขณะที่กำหนดค่า $C = 16$ และ $\varepsilon = 0.1$ จากผลในตารางที่ 10 พบว่าค่า γ ที่ทำให้ RMSE ต่ำที่สุดคือ $\gamma = 0.125$ ดูท้ายปรับค่า ε ระหว่างช่วง 2^{-4} ถึง 2^7 ผลในตารางที่ 11 พบว่าสำหรับค่า ε ที่น้อยกว่า 16 การปรับค่า ε นั้นไม่ส่งผล อย่างมีนัยสำคัญต่อมัตรฐานของแบบจำลองสะท้อนให้เห็นว่าในการประมาณฟังก์ชันได้�่อมมีความคลาดเคลื่อนที่ไม่สามารถทำให้ลดลงได้ (irreducible error) เกิดขึ้นเสมอในที่นี้เลือกค่า $\varepsilon = 4$

สรุปค่าพารามิเตอร์ที่เลือกใช้คือ $C = 128$, $\gamma = 0.125$ และ $\varepsilon = 4$

ตาราง 9: สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า C

C	RMSE	
	training set	validation set
2^{-2}	137.4814	131.0199
2^{-1}	121.6232	121.6037
2^0	113.5810	116.3607
2^1	109.0157	112.6498
2^2	106.1898	110.1800
2^3	104.5957	108.7586
2^4	103.4812	107.9071
2^5	102.6484	107.3979
2^6	101.8962	107.4631
2^7	101.0159	107.3815
2^8	100.1987	107.4247
2^9	99.2640	107.5412

หมายเหตุ ทดลองปรับค่า C ในขณะที่ค่า $\gamma = 1/11$ และ $\varepsilon = 0.1$

ตาราง 10: สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า γ

γ	RMSE	
	training set	validation set
2^2	197.1516	201.7977
2^1	164.1737	165.2336
2^0	132.5032	132.7949
2^{-1}	113.8153	115.7224
2^{-2}	106.4343	109.5702
2^{-3}	104.7495	108.6500
2^{-4}	104.7007	109.1461
2^{-5}	105.1972	110.0012
2^{-6}	106.2430	111.4005
2^{-7}	108.2354	113.1515

หมายเหตุ ทดลองปรับค่า γ ในขณะที่ค่า $C = 2^7$ และ $\varepsilon = 0.1$

ตาราง 11: สมรรถนะของแบบจำลอง SVR เมื่อปรับค่า ε

ε	RMSE	
	training set	validation set
2^7	115.3209	118.9872
2^6	106.2897	110.4693
2^5	104.9467	109.0708
2^4	104.7150	108.6971
2^3	104.7013	108.5600
2^2	104.6915	108.6042
2^1	104.7177	108.6412
2^0	104.7319	108.6539
2^{-1}	104.7444	108.6456
2^{-2}	104.7503	108.6537
2^{-3}	104.7491	108.6510
2^{-4}	104.7499	108.6464

หมายเหตุ ทดลองปรับค่า ε ในขณะที่ค่า $C = 2^7$ และ $\gamma = 2^{-3}$

10.2 ผลการปรับค่าพารามิเตอร์ของแบบจำลอง Random Forest

พารามิเตอร์สำคัญที่เป็นตัวกำหนดเงื่อนไขของแบบจำลอง และยังส่งต่อประสิทธิภาพ/ความซับซ้อนในการคำนวณของการพยากรณ์มีดังนี้

1. จำนวนแบบจำลองต้นไม้ทั้งหมดภายในป่า เขียนแทนด้วย n_{tree}
เป็นพารามิเตอร์ที่ส่งผลโดยตรงต่อกำลังการคำนวณที่ใช้ในทั้งการฝึกแบบจำลองและขั้นตอนการพยากรณ์โดยถึง n_{tree} มีค่ามาก แบบจำลองจะมีความแปรปรวนต่อการเปลี่ยนแปลงชุดข้อมูลฝึกัดลง แต่ในขณะเดียวกันในการคำนวณจะใช้กำลังการคำนวณที่มากขึ้น
2. จำนวนระดับหรือความลึกมากสุดของต้นไม้ที่ยอมรับได้ เขียนแทนด้วย d
เป็นพารามิเตอร์ที่กำหนดความซับซ้อนของแบบจำลองและการเข้ากันได้ในชุดข้อมูลฝึก การปรับค่า d ให้มีค่ามากเกินไปจะนำไปสู่เกิดปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลฝึกมากเกินไป (Over-fitting)
3. จำนวนตัวอย่างจากชุดข้อมูลฝึกน้อยสุดภายในปริภูมิ ที่ยินยอมให้มีการเริ่มต้นการแบ่งปริภูมิ เขียนแทนด้วย $n_{min_samples_split}$
เป็นพารามิเตอร์ที่ควบคุมความสมดุลระหว่างค่า RMSE ในชุดข้อมูลฝึก และความซับซ้อนของแบบจำลอง การปรับค่า $n_{min_samples_split}$ ให้มีค่าน้อยเกินไปเมื่อโอกาสที่จะทำให้จำนวนระดับหรือความลึกของต้นไม้มีค่ามาก และนำไปสู่เกิดปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลฝึกมากเกินไป (Over-fitting)
4. จำนวนตัวอย่างจากชุดข้อมูลฝึกน้อยสุดที่ยินยอมมีในแต่ละปริภูมิอยู่ เขียนแทนด้วย $n_{min_samples_leaf}$
เป็นพารามิเตอร์ที่ควบคุมความสมดุลระหว่างค่า RMSE ในชุดข้อมูลฝึก และความซับซ้อนของแบบจำลอง การปรับค่า $n_{min_samples_leaf}$ ให้มีค่าน้อยเกินไปเมื่อโอกาสที่จะทำให้จำนวนระดับหรือความลึกของต้นไม้มีค่ามาก และนำไปสู่เกิดปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลฝึกมากเกินไป (Over-fitting) เป็นพารามิเตอร์ที่มีความสัมพันธ์กันกับ $n_{min_samples_split}$
5. จำนวนคุณลักษณะของตัวแปรต้นໃใช้ในแต่ละปมของแบบจำลองต้นไม้ เขียนแทนด้วย m
เป็นพารามิเตอร์ที่มีผลกับความแตกต่างกันระหว่างแบบจำลองต้นไม้ย่อยแต่ละแบบจำลอง เมื่อเราปรับค่า m ให้น้อยลง แบบจำลองต้นไม้ย่อยแต่ละแบบจำลองจะมีความแตกต่างกันมากขึ้น แต่การเข้ากันระหว่างแบบจำลองและชุดข้อมูลฝึกจะลดลง

n_{tree} เป็นพารามิเตอร์ที่ส่งผลโดยตรงต่อกำลังการคำนวณที่ใช้ในทั้งการฝึกแบบจำลองและขั้นตอนการพยากรณ์ ซึ่งในการทดลองนี้เราจะกำหนด $n_{tree} = 1000$ จากนั้นจะทำการคำนวณค่า RMSE ในชุดข้อมูลฝึกและชุดข้อมูลตรวจสอบ เมื่อปรับพารามิเตอร์ $n_{min_samples_split}$ และ $n_{min_samples_leaf}$ โดยไม่พิจารณาเงื่อนไขของ d และกำหนด $m = p = 25$

ตาราง 12: สมรรถนะของแบบจำลอง RF เมื่อปรับค่า $n_{min_samples_split}$ และ $n_{min_samples_leaf}$

$n_{min_samples_split}$	$n_{min_samples_leaf}$	RMSE	
		training set	validation set
28	16	83.4965	105.3259
32	16	83.4965	105.3259
34	16	83.5903	104.7497
36	16	84.2358	105.2982
40	16	84.9581	105.2853
44	16	85.6133	105.2779
34	10	80.1608	104.8592
34	12	81.2877	104.8138
34	14	82.2806	104.7892
34	16	83.5903	104.7497
34	18	84.3997	104.7647
34	20	85.7072	104.6964

หมายเหตุ ทดลองปรับค่าค่า $n_{min_samples_split}$ และ $n_{min_samples_leaf}$ ในขณะที่ $m = 25$

จากผลลัพธ์ดังตารางที่ 12 พบร่วงการปรับค่า $n_{min_samples_split}$ และ $n_{min_samples_leaf}$ จะส่งผลต่อค่า RMSE ในชุดข้อมูลฝึกคือเมื่อเราเพิ่มค่า $n_{min_samples_leaf}$ หรือ $n_{min_samples_leaf}$ จะทำให้ค่า RMSE ในชุดข้อมูลฝึกมีค่าสูงขึ้น เพราะเป็นการจำกัดเงื่อนไขในการเข้ากัน (fitting) ของแบบจำลอง อย่างไรก็ตามการเพิ่มค่า $n_{min_samples_split}$ และ $n_{min_samples_split}$ เป็นการป้องกันปัญหาการเข้ากันระหว่างแบบจำลองและชุดข้อมูลมากเกินไป (Over-fitting) ซึ่งในที่นี้เราจะเลือก $n_{min_samples_split}$ และ $n_{min_samples_leaf}$ ซึ่งทำให้ค่า RMSE ในชุดข้อมูลตรวจสอบมีค่าต่ำที่สุด นั่นคือ

$$n_{\text{min_samples_split}} = 34, n_{\text{min_samples_leaf}} = 16$$

จากนั้นปรับค่า d ซึ่งเป็นพารามิเตอร์ซึ่งกำหนดความชัดช้อนของแบบจำลองได้ผลลัพธ์ตามตารางที่ 13 เลือกค่า d ซึ่งทำให้ RMSE ในชุดข้อมูลตรวจสอบมีค่าต่ำที่สุด นั่นคือ $d = 10$ สุดท้ายปรับค่า m ได้ผลลัพธ์ตามตารางที่ 14 และเลือกค่า m ซึ่งทำให้ RMSE ในชุดข้อมูลตรวจสอบมีค่าต่ำที่สุด นั่นคือ $m = 13$

ตาราง 13: สมรรถนะของแบบจำลอง RF เมื่อปรับค่า d

d	RMSE	
	training set	validation set
20	83.5903	104.7497
15	83.3568	104.7410
13	83.8632	104.7304
12	84.4242	104.7097
11	85.3068	104.6863
10	86.6353	104.6413
9	88.5226	104.6944
8	90.9809	104.7087
7	93.8837	104.7113

หมายเหตุ ทดลองปรับค่า d ในขณะที่ $n_{\text{min_samples_split}} = 34, n_{\text{min_samples_leaf}} = 16, m = 25$

ตาราง 14: สมรรถนะของแบบจำลอง RF เมื่อปรับค่า m

m	RMSE	
	training set	validation set
25	86.6353	104.6413
20	87.0418	104.6347
15	87.3493	104.5573
14	87.8009	104.5927
13	87.9583	104.4538
12	88.1825	104.5702
11	88.4028	104.5692
10	88.6606	104.7341

หมายเหตุ ทดลองปรับค่า m ในขณะที่ $n_{\text{min_samples_split}} = 34, n_{\text{min_samples_leaf}} = 16, d = 10$

สรุปค่าพารามิเตอร์ที่เลือกคือ

$$n_{\text{tree}} = 1000, n_{\text{min_samples_split}} = 34, n_{\text{min_samples_leaf}} = 16, d = 10, m = 13 \quad (37)$$

10.3 ชุดโปรแกรมคำสั่ง

ชุดโปรแกรมคำสั่งที่จัดทำขึ้นในโครงงานนี้สามารถเข้าถึงได้จาก https://github.com/sunncyn/Intraday_solar_forecasting