

การทำนายการผลิตพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์ด้วยการเรียนรู้ของเครื่อง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2566

PREDICTION OF SOLAR CELL ENERGY PRODUCTION USING MACHINE LEARNING



Mr. Phoomiphat Hongthanakornhirun

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering

Faculty Of Engineering

Chulalongkorn University

Academic Year 2023

| | |
|---------------------------------|--|
| หัวข้อวิทยานิพนธ์ | การทำนายการผลิตพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์ด้วยการเรียนรู้ของเครื่อง |
| โดย | นายภูมิพัฒน์ หงส์ธนากรหิรัญ |
| สาขาวิชา | วิศวกรรมไฟฟ้า |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | ศาสตราจารย์ ดร.ลัญฉกร วุฒิสีทธิกุลกิจ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม | ศาสตราจารย์ ดร.วาทีต เบญจพลกุล |

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของ
ของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

| | |
|---|---------------------------------|
| | คณบดีคณะวิศวกรรมศาสตร์ |
| (ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล) | |
| คณะกรรมการสอบวิทยานิพนธ์ | ประธานกรรมการ |
| | |
| (รองศาสตราจารย์ ดร.ภูมิพัฒน์ แสงอุดมเลิศ) | |
| | อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก |
| (ศาสตราจารย์ ดร.ลัญฉกร วุฒิสีทธิกุลกิจ) | |
| | อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม |
| (ศาสตราจารย์ ดร.วาทีต เบญจพลกุล) | |
| | กรรมการ |
| (ดร.พิสิษฐ์พล จิรพวงศานานุรักษ์) | |
| | กรรมการ |
| (ผู้ช่วยศาสตราจารย์ ดร.วิทยากร อัครวิเศษ) | |

ภูมิพัฒน์ หงส์ธนากรศิริชัย : การทำนายการผลิตพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์ด้วยการเรียนรู้ของเครื่อง. (PREDICTION OF SOLAR CELL ENERGY PRODUCTION USING MACHINE LEARNING) อ.ที่ปรึกษาหลัก : ศ. ดร.ลัญจกร วุฒิสทิธิกุลกิจ, อ.ที่ปรึกษาร่วม : ศ. ดร.วาทีต เบญจพลกุล

ในวิทยานิพนธ์ฉบับนี้ ผู้เขียนได้เสนอโมเดลการเรียนรู้ของเครื่องที่เกิดจากการผสมการทำนายร่วมกันระหว่างโมเดลที่ต่างชนิดกัน เพื่อปรับปรุงประสิทธิภาพการทำนายค่าพลังงานไฟฟ้าที่ผลิตได้จากเซลล์แสงอาทิตย์ให้มีความแม่นยำยิ่งขึ้น ในการศึกษาครั้งนี้ได้นำข้อมูลจากระบบฐานข้อมูลของโรงไฟฟ้าพลังงานแสงอาทิตย์ขนาด 5 เมกะวัตต์ ของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) ตั้งอยู่ที่ อ.ทับสะแก จ.ประจวบคีรีขันธ์ และข้อมูลสภาพอากาศย้อนหลังจากเว็บไซต์ Solcast และ OpenWeatherMap มาวิเคราะห์ความสัมพันธ์ เมื่อได้ตัวแปรที่ส่งผลกระทบต่อปริมาณการผลิตจึงนำข้อมูลย้อนหลังไปเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่องชนิดต่างๆ เพื่อนำผลการทำนายที่ได้มาเปรียบเทียบประสิทธิภาพการทำนายร่วมกับโมเดลที่เสนอโดยงานวิทยานิพนธ์ฉบับนี้ได้เสนอการทำนายร่วมกันโดยใช้โมเดลการวิเคราะห์การถดถอยเชิงเส้นและโมเดลแรนดอมฟอเรสต์ เพื่อเป็นการปรับปรุงประสิทธิภาพการทำนาย ผู้วิจัยได้เสนอการปรับจูนค่าไฮเปอร์พารามิเตอร์และเพิ่มโมเดลบุสต์ชนิดต่างๆ เพื่อเพิ่มประสิทธิภาพโมเดล เมื่อกระบวนการเรียนรู้เสร็จสิ้นนำโมเดลการเรียนรู้ที่ได้ไปทำนายข้อมูลชุดตรวจสอบและชุดทดสอบ จากนั้นจึงนำผลการทำนายที่ได้ไปประเมินประสิทธิภาพการทำนายโดยหาค่า MAE, RMSE และ R^2 ระหว่างค่าพลังงานไฟฟ้าที่ผลิตได้จริงกับผลที่ได้จากการทำนาย จากผลการทำนายพบว่าโมเดลการเรียนรู้ของเครื่องที่เสนอในการทดลองนี้ให้ประสิทธิภาพสูงที่สุดโดยมีค่า R^2 มากกว่า 94% ซึ่งแสดงถึงความสามารถในการทำนายที่ดี ซึ่งส่งผลให้การนำข้อมูลผลการทำนายที่ได้ไปวิเคราะห์แนวโน้มการผลิตเพื่อเพิ่มประสิทธิภาพในการจัดการพลังงานดียิ่งขึ้น

สาขาวิชา วิศวกรรมไฟฟ้า

ปีการศึกษา 2566

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

6470045321 : MAJOR ELECTRICAL ENGINEERING

KEYWORD: Machine Learning / Prediction / Performance

Phoomiphat Hongthanakornhirun : PREDICTION OF SOLAR CELL ENERGY PRODUCTION USING MACHINE LEARNING. Advisor: Prof. LUNCHAKORN WUTTISITTIKULKIJ Co-advisor: Prof. WATIT BENJAPOLAKUL

In this thesis, the author proposed combining different machine-learning models to improve prediction performance. The author obtained solar power from the Electrical Generating Authority of Thailand 5 MW Solar Power Plant in Thap Sakae, Prachuap Khiri Khan. Weather information in this research was obtained from the Solcast and the OpenWeatherMap websites, which can get the data using latitude and longitude. First, check the data's correctness and cleanliness then find the correlation between input and target features. Next, creating and splitting the datasets of training, validating, and testing by 60, 20, and 20 percent, respectively. After that, we train the data with our proposed model, an ensemble of linear regression and random forest regressor models, and other machine-learning models. During training, tuning hyperparameters and boosting algorithms help reach higher performances. Next, try to train with the validation and test datasets. After finishing, evaluate the model by using evaluation metrics such as MAE, RMSE, and R^2 . As a result, our proposed method can reach an R^2 score of over 94% which can help electricity managers manage electricity outflows to the main electrical infrastructure with the best performances.

Field of Study: Electrical Engineering

Student's Signature

Academic Year: 2023

Advisor's Signature

Co-advisor's Signature

กิตติกรรมประกาศ

งานวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีเนื่องจากได้รับคำแนะนำและการช่วยเหลือจาก ศาสตราจารย์ ดร. ลัญฉกร วุฒิสีทธิกุลกิจ และ ศาสตราจารย์ ดร. วาทีต เบญจพลกุล สำหรับการดำเนินงานวิจัยทุกขั้นตอน

ขอขอบคุณรองศาสตราจารย์ ดร.ภูมิพัฒน์ แสงอุดมเลิศ ประธานกรรมการ ผู้ช่วยศาสตราจารย์ ดร.วิทยากร อัครวิเศษ และ อ.ดร.พิสิษฐ์พล จิรพงศานานุรักษ์ ที่ให้เกียรติมาเป็นกรรมการสอบวิทยานิพนธ์

ขอขอบคุณ นายวัฒนศักดิ์ ศรีศิริ (พี่มาร์ค) นิสิตระดับปริญญาโทผู้บัณฑิตสำหรับความช่วยเหลือในการเข้าไปเก็บข้อมูลและคำแนะนำในการแก้ปัญหาระหว่างการทดลองงานวิจัย

ขอขอบคุณทางการไฟฟ้าฝ่ายผลิตแห่งประเทศไทยที่ให้ความกรุณาในการเข้าไปเก็บข้อมูลสำหรับงานวิจัยฉบับนี้ที่โรงไฟฟ้าพลังงานแสงอาทิตย์ อ.ทับสะแก จ.ประจวบคีรีขันธ์ ตลอดจนบุคลากรที่ให้ความช่วยเหลือและคำแนะนำในการเก็บรวบรวมข้อมูล

สุดท้ายนี้ขอขอบคุณทางครอบครัวของผู้วิจัยที่ให้การสนับสนุนตลอดการศึกษานี้

ภูมิพัฒน์ หงส์ธนากรศิริณู

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

หน้า

| | |
|--|----|
| | ค |
| บทคัดย่อภาษาไทย..... | ค |
| | ง |
| บทคัดย่อภาษาอังกฤษ..... | ง |
| กิตติกรรมประกาศ..... | จ |
| สารบัญ..... | ฉ |
| สารบัญตาราง..... | ฉ |
| สารบัญรูปภาพ..... | ญ |
| บทที่ 1 | 1 |
| บทนำ..... | 1 |
| 1.1 ที่มาและความสำคัญ..... | 1 |
| 1.2 การทบทวนวรรณกรรม..... | 2 |
| 1.3 วัตถุประสงค์ของวิทยานิพนธ์..... | 10 |
| 1.4 ขอบเขตของวิทยานิพนธ์..... | 10 |
| 1.5 ขั้นตอนการดำเนินการ..... | 11 |
| 1.6 ประโยชน์ที่คาดว่าจะได้รับ..... | 11 |
| บทที่ 2 | 12 |
| ความรู้พื้นฐานและทฤษฎีที่เกี่ยวข้อง | 12 |
| 2.1 การเตรียมข้อมูลสำหรับการทำนายด้วยโมเดลการเรียนรู้ของเครื่อง..... | 12 |
| 2.1.1 การวิเคราะห์ความสัมพันธ์ของข้อมูล..... | 12 |
| 2.1.2 การกำหนดมาตรฐานของข้อมูล (Standardization)..... | 13 |

| | | |
|---------|--|----|
| 2.1.3 | กระบวนการนอร์มัลไลเซชัน (Normalization) | 14 |
| 2.2 | โมเดลการเรียนรู้ของเครื่องสำหรับการทำนายพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์ | 14 |
| 2.2.1 | การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) | 14 |
| 2.2.2 | แรนดอมฟอเรสต์สำหรับการถดถอย (Random Forest for Regression) | 15 |
| 2.2.3 | การปรับจูนค่าไฮเปอร์พารามิเตอร์ (Hyperparameter Optimization) | 16 |
| 2.2.4 | การค้นหาชุดของตัวปรับจูนข้อมูลของไฮเปอร์พารามิเตอร์ | 16 |
| 2.2.5 | ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) | 18 |
| 2.2.6 | อาร์มา (Auto Regressive Integrated Moving Average) | 19 |
| 2.3 | การประเมินความแม่นยำของข้อมูลการทำนายด้วยวิธีการทางสถิติ | 20 |
| บทที่ 3 | | 21 |
| | การออกแบบและพัฒนาโมเดลการเรียนรู้ของเครื่องสำหรับการทำนายพลังไฟฟ้าที่ผลิตได้จากเซลล์แสงอาทิตย์ | 21 |
| 3.1 | ภาพรวมขั้นตอนการทำงานของงานวิจัย | 21 |
| 3.2 | รายงานวิธีการดำเนินการวิจัย | 22 |
| บทที่ 4 | | 34 |
| | การทดสอบและผลการวิจัย | 34 |
| 4.1 | การทดสอบหาความสัมพันธ์ระหว่างข้อมูล | 34 |
| 4.2 | กระบวนการเรียนรู้ผ่านโมเดล | 38 |
| 4.2.1 | โมเดลการวิเคราะห์การถดถอยเชิงเส้น | 38 |
| 4.2.2 | โมเดลแรนดอมฟอเรสต์ | 39 |
| 4.2.3 | โมเดลซัพพอร์ตเวกเตอร์แมชชีน | 39 |
| 4.2.4 | โมเดลอาร์มา | 40 |
| 4.2.5 | โมเดลที่เสนอ (LinearRegression+RandomForest+XgBoost+LightGBM) | 40 |
| 4.3 | ผลการประเมินประสิทธิภาพโมเดลการเรียนรู้ของเครื่อง | 41 |

| | |
|----------------------------------|----|
| บทที่ 5 | 46 |
| สรุปและวิเคราะห์ผลการทดลอง | 46 |
| ภาคผนวก..... | 48 |
| บรรณานุกรม..... | 60 |
| ประวัติผู้เขียน | 65 |



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

| | หน้า |
|--|------|
| ตารางที่ 1 สรุปการทบทวนวรรณกรรม..... | 6 |
| ตารางที่ 2 ค่าประเมินประสิทธิภาพโมเดลการเรียนรู้ของเครื่อง | 42 |
| ตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่นในการทบทวนวรรณกรรม | 43 |
| ตารางที่ 4 รายงานการตรวจสอบข้อมูลจากระบบโรงไฟฟ้าพลังงานแสงอาทิตย์ ปี 2020 | 48 |
| ตารางที่ 5 รายงานการตรวจสอบข้อมูลจากระบบโรงไฟฟ้าพลังงานแสงอาทิตย์ ปี 2021 | 50 |
| ตารางที่ 6 รายงานการตรวจสอบข้อมูลจากระบบโรงไฟฟ้าพลังงานแสงอาทิตย์ ปี 2022 | 52 |
| ตารางที่ 7 รายงานการตรวจสอบข้อมูลจากระบบโรงไฟฟ้าพลังงานแสงอาทิตย์ ปี 2023 | 54 |
| ตารางที่ 8 ผลการทำนายระหว่างวันที่ 1 ถึง 7 สิงหาคม 2566 และค่าพลังงานที่ผลิตได้จริง..... | 56 |

สารบัญรูปภาพ

| | หน้า |
|--|------|
| รูปที่ 1 แผนภาพความร้อนแสดงความสัมพันธ์ระหว่างข้อมูลตัวแปร | 13 |
| รูปที่ 2 แผนผังแสดงรูปแบบของโมเดลเรียดอมพอเรสต์ชนิดถดถอย | 16 |
| รูปที่ 3 แผนผังเสนอการปรับปรุงการทำงานของอัลกอริทึมเรียดอมพอเรสต์ชนิดถดถอย | 17 |
| รูปที่ 4 ศูนย์ควบคุมโรงไฟฟ้าพลังงานแสงอาทิตย์ | 21 |
| รูปที่ 5 ภาพรวมแสดงการดำเนินการวิจัย | 22 |
| รูปที่ 6 ภาพแผงโซลาร์เซลล์บริเวณโซน C | 23 |
| รูปที่ 7 ตัวอย่างชุดข้อมูลที่ผ่านการนอร์มัลไลเซชัน | 23 |
| รูปที่ 8 แผนภาพ flowchart แสดงการดำเนินการขั้นตอนที่ 1 | 24 |
| รูปที่ 9 แผนภาพ flowchart แสดงการดำเนินการขั้นตอนที่ 1 (ต่อ) | 25 |
| รูปที่ 10 การนำเข้าชุดคำสั่งและชุดข้อมูล | 26 |
| รูปที่ 11 โค้ดโปรแกรมการรวมชุดข้อมูล | 26 |
| รูปที่ 12 โค้ดโปรแกรมการรวมชุดข้อมูล | 27 |
| รูปที่ 13 โค้ดโปรแกรมสำหรับการแบ่งชุดข้อมูล | 27 |
| รูปที่ 14 โค้ดโปรแกรมแสดงการกำจัดตัวแปร | 27 |
| รูปที่ 15 แผนภาพ flowchart สำหรับการดำเนินการขั้นตอนที่ 2 | 28 |
| รูปที่ 16 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Linear Regression | 29 |
| รูปที่ 17 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Random Forest Regressor | 29 |
| รูปที่ 18 โค้ดโปรแกรมแสดงการค้นหาค่าไฮเปอร์พารามิเตอร์ | 30 |
| รูปที่ 19 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดลบูสท์ | 31 |
| รูปที่ 20 โค้ดโปรแกรมสำหรับการผสานโมเดล (Ensemble) | 31 |
| รูปที่ 21 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Support Vector Machine | 32 |

| | |
|--|----|
| รูปที่ 22 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดลอาร์ริมา | 32 |
| รูปที่ 23 โค้ดโปรแกรมการสร้างฟังก์ชันตัวชี้วัดสำหรับประเมินประสิทธิภาพ | 33 |
| รูปที่ 24 ตัวอย่างวิธีการหาค่าสัมประสิทธิ์ของเพียร์สัน | 34 |
| รูปที่ 25 ความสัมพันธ์ระหว่างข้อมูลจากโรงไฟฟ้าและสภาพอากาศจากเว็บไซต์ SolCast | 35 |
| รูปที่ 26 ความสัมพันธ์ระหว่างข้อมูลจากโรงไฟฟ้าและสภาพอากาศจากเว็บไซต์ OpenWeatherMap | 35 |
| รูปที่ 27 แผนภาพแสดงโมเดลการวิเคราะห์หาค่าการแผ่รังสีของดวงอาทิตย์ | 37 |
| รูปที่ 28 ผลการทำนายผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้นเปรียบเทียบกับค่าที่ผลิตได้จริง . | 38 |
| รูปที่ 29 ผลการทำนายผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้นเปรียบเทียบกับค่าที่ผลิตได้จริง . | 39 |
| รูปที่ 30 ผลการทำนายผ่านโมเดลซัพพอร์ตเวกเตอร์แมชชีนเปรียบเทียบกับค่าที่ผลิตได้จริง | 39 |
| รูปที่ 31 ผลการทำนายผ่านโมเดลอาร์ริมาเปรียบเทียบกับค่าที่ผลิตได้จริง | 40 |
| รูปที่ 32 ผลการทำนายผ่านโมเดลที่เสนอเปรียบเทียบกับค่าที่ผลิตได้จริง | 40 |
| รูปที่ 33 ผลการทำนายผ่านโมเดลที่เสนอเปรียบเทียบกับ LR และ RF | 41 |

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ในปัจจุบันพลังงานจากเซลล์แสงอาทิตย์ได้มีการใช้แพร่หลายขึ้น โดยทางองค์การพลังงานระหว่างประเทศ (IEA) ได้มีการประเมินว่าในอีก 5 ปีข้างหน้าสัดส่วนการใช้พลังงานหมุนเวียนจะมีเพิ่มขึ้นประมาณ 30% เมื่อเทียบกับปริมาณการใช้พลังงานหมุนเวียนในปัจจุบัน ซึ่งพลังงานหมุนเวียนหลักที่นำมาใช้ในการผลิตไฟฟ้าคือพลังงานลมและพลังงานแสงอาทิตย์ จากรายงาน ได้ระบุว่ามีการตั้งเป้าไว้ว่าในปี 2030 ค่ากำลังการผลิตไฟฟ้าจากเซลล์แสงอาทิตย์รวมทั่วโลกจะเพิ่มขึ้นราว 25% และในปัจจุบันมีการใช้งานการรับส่งข้อมูลกันอย่างหลากหลายวิธี สำหรับการรับส่งข้อมูลสำหรับระบบพลังงานเซลล์แสงอาทิตย์เพื่อที่จะแสดงผลข้อมูลกำลังการผลิต ข้อมูลความผิดปกติบนแผงหรือข้อมูลต่างๆ ที่ต้องการแสดงผลแบบเรียลไทม์ ทำให้ทางผู้ผลิตเซลล์พลังงานแสงอาทิตย์ต้องมีค่าใช้จ่ายที่เพิ่มขึ้น เพื่อเป็นการลดค่าใช้จ่ายในส่วนของการรับส่งข้อมูล จึงได้มีการนำระบบการจัดการข้อมูลเข้ามาใช้ในการจัดการข้อมูลต่างๆ ในระบบการผลิตพลังงานไฟฟ้าเซลล์แสงอาทิตย์ แต่การนำระบบการจัดการเข้ามาใช้ส่งผลให้ช่วงเวลาในการรับส่งข้อมูลอาจมีความล่าช้าในการส่งข้อมูล ด้วยเหตุนี้จึงส่งผลให้การทำนายปริมาณการผลิตพลังงานไฟฟ้าด้วยพลังงานแสงอาทิตย์ยากขึ้นกว่าเดิม

จากมุมมองของผู้ผลิตไฟฟ้าเพื่อให้มั่นใจความเสถียรของการเชื่อมต่อระบบโรงไฟฟ้าพลังงานแสงอาทิตย์เข้ากับโครงข่ายระบบไฟฟ้าสาธารณูปโภค การลดค่าใช้จ่ายที่ไม่จำเป็นและการรับรองความปลอดภัยในการเชื่อมต่อระบบผลิตพลังงานไฟฟ้าเซลล์แสงอาทิตย์เข้ากับระบบการไฟฟ้าหลักเนื่องจากระบบผลิตไฟฟ้าเซลล์แสงอาทิตย์นั้นเป็นพลังงานหมุนเวียนซึ่งปริมาณการผลิตในแต่ละวันมีความไม่แน่นอน ต่างจากระบบการผลิตไฟฟ้าแบบพลังงานไอน้ำที่ใช้เป็นแหล่งผลิตพลังงานไฟฟ้าหลักในปัจจุบัน ดังนั้นการพยากรณ์ปริมาณการผลิตไฟฟ้าง่วงหน้าจึงมีความจำเป็นต่อการผลิตไฟฟ้าด้วยเซลล์พลังงานแสงอาทิตย์ จาก (Porter) กล่าวว่าต้นทุนที่ใช้ในการดำเนินงานของระบบโครงข่ายไฟฟ้าจะเพิ่มขึ้นตามกำลังการผลิตที่เพิ่มขึ้นของระบบผลิตไฟฟ้าจากพลังงานแสงอาทิตย์ (Lew) ได้เสนอวิธีการขั้นสูงและมีประสิทธิภาพในการประเมินการผลิตไฟฟ้าจากเซลล์แสงอาทิตย์ เพื่อลดต้นทุนการดำเนินงานของโครงข่ายไฟฟ้า และเพื่อเป็นการลดผลกระทบที่เกิดจากการเชื่อมโครงข่ายไฟฟ้าเข้ากับเซลล์พลังงานแสงอาทิตย์ ผู้ดำเนินการด้านพลังงานจึงจำเป็นต้องมีวิธีการคาดการณ์ปริมาณการผลิต

ไฟฟ้าที่แม่นยำเพื่อเป็นการลดต้นทุน ดังนั้นสำหรับโรงไฟฟ้าพลังงานแสงอาทิตย์จึงมีความจำเป็นอย่างมากที่จะต้องมีการพยากรณ์ปริมาณการผลิตพลังงานที่แม่นยำ เพื่อเป็นประโยชน์ในการบริหารจัดการปริมาณไฟฟ้าที่ต้องจ่ายในระบบโครงข่ายได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

1.2 การทบทวนวรรณกรรม

จากการศึกษางานวิจัยที่เกี่ยวข้อง ในการผลิตไฟฟ้าโดยใช้พลังงานแสงอาทิตย์มีปัจจัยหลายๆ อย่างที่ส่งผลกระทบต่อกำลังการผลิตไฟฟ้า เช่น ค่าความเข้มของแสง อุณหภูมิบนแผงโซลาร์เซลล์ ค่าความชื้น มุมที่แสงตกกระทบกับแผงโซลาร์เซลล์ ชนิดของแผง ในการทำนายค่ากำลังการผลิตได้นำค่าของปัจจัยต่างๆ มาใช้ในการทำนายด้วยเทคนิคต่างๆ โดยใช้เทคนิคการเรียนรู้ที่เรียกว่าโมเดลการเรียนรู้ของเครื่องหรือที่รู้จักกันในภาษาอังกฤษว่า Machine Learning Model

Sonia Leva (Leva) ได้นำเทคนิค โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) มาใช้ในการทำนายค่ากำลังการผลิตแบบระยะสั้น โดยเริ่มต้นด้วยการนำข้อมูลกำลังการผลิตที่มีอยู่ในระบบมาใช้ในการเริ่มต้นเป็นค่าอินพุต จากนั้นได้นำโมเดลการวัดค่ารังสีความร้อนเมื่อท้องฟ้าไม่มีเมฆ (Clear Sky Solar Radiation Model, CSRМ) เพื่อตรวจสอบความน่าเชื่อถือของข้อมูลทุกๆ 15 นาที โดยวัดจากรังสีของดวงอาทิตย์เมื่อไม่มีก้อนเมฆจากทฤษฎีของ CSRМ ($G_{CSRМ,1/4h}^k$), ค่ารังสีของดวงอาทิตย์ที่วัดได้ ($G_{M,1/4h}^k$), และค่าของกำลังการผลิตไฟฟ้าที่ได้จากการผลิตโดยใช้พลังงานแสงอาทิตย์ ($P_{M,1/4h}^k$) ในบทความนี้ได้้นำค่าการพยากรณ์อากาศของศูนย์พยากรณ์อากาศ โดยได้ระบุพิกัด เวลา และวันที่เพื่อให้ได้ข้อมูลของตำแหน่งของดวงอาทิตย์ที่ถูกต้องที่สุดโดยตลอด การทดลองได้มีการปรับปรุงข้อมูลให้ได้เป็นปัจจุบันมากที่สุดเพื่อลดความคลาดเคลื่อนของการทดลอง จากบทความนี้ได้สรุปผลการทดลองการคาดการณ์กำลังการผลิตไฟฟ้าของโรงไฟฟ้าแสงอาทิตย์ว่า ค่าความคลาดเคลื่อนโดยรวมจากการทดสอบตลอดการทดลองมีค่าน้อยกว่าค่าความคลาดเคลื่อนที่ได้จากการทดสอบเพียงครั้งเดียว และผู้วิจัยยังระบุเน้นย้ำถึงความถูกต้องของการนำข้อมูลในอดีตที่เก็บไว้มาประมวลผลย้อนหลัง และยังได้กล่าวว่าในวันที่มีเหตุการณ์คาดการณ์นั้นจะมีความแม่นยำสูง ในขณะที่ประสิทธิภาพโดยรวมของวันที่มีเมฆบางส่วนหรือมีเมฆมากแตกต่างกันเล็กน้อย ดังนั้นผลของการคาดการณ์จึงต้องอาศัยความน่าเชื่อถือของการพยากรณ์อากาศและการเตรียมข้อมูลที่จะนำมาใช้ในการทำนายกำลังการผลิต

Marco Pierro (Pierro) ได้เสนอการผสมผสานของโมเดลเพื่อนำมาใช้ในการทำนายค่ากำลังการผลิตไฟฟ้าของโรงไฟฟ้าพลังงานแสงอาทิตย์ โดยข้อมูลที่จะนำมาใช้ในการพยากรณ์ที่ได้นำข้อมูลเชิง

ตัวเลขที่ได้จากการพยากรณ์อากาศเปรียบเทียบกับข้อมูลที่วัดได้จากภาคพื้นดินในบริเวณโรงไฟฟ้าพลังงานแสงอาทิตย์หลังจากได้ข้อมูลแล้วก็จะนำข้อมูลที่ได้ออกไปเข้าโมเดลการเรียนรู้ของเครื่องสำหรับการทำนายข้อมูลกำลังการผลิตล่วงหน้า โดย Pierro ได้ใช้ทั้งหมด 3 โมเดลในการทดลอง โมเดลแรกได้ใช้ SARIMAX โมเดลที่สองใช้ SVM และโมเดลสุดท้ายได้เลือกใช้ ANN จากการทดลองพบว่า SVMs และ ANN เป็นโมเดลที่สามารถนำมาใช้ในการทำนายค่ากำลังการผลิตไฟฟ้าได้ ในขณะที่ SARIMAX เหมาะกับการนำไปใช้แค่ทำแบบจำลองการทดลอง

Gordon Reikard (Reikard) ได้บอกว่าหัวใจสำคัญของการนำระบบไฟฟ้าพลังงานแสงอาทิตย์รวมเข้ากับเครือข่ายไฟฟ้าหลักคือการทำนายกำลังการผลิต ปัญหาหลักของบทความนี้มีสองประการคือความสัมพันธ์กับโมเดลลำดับเวลาและจะนำพารามิเตอร์มาใช้ในการทำนายระยะสั้นได้อย่างไร ซึ่งจากงานวิจัยนี้ได้เสนอวิธีการโดยปรับค่าพารามิเตอร์ให้สอดคล้องกับข้อมูลสภาพอากาศที่เปลี่ยนแปลง ในงานวิจัยชิ้นนี้ได้นำข้อมูลการพยากรณ์จากโรงไฟฟ้าพลังงานแสงอาทิตย์สองแห่งในสหรัฐอเมริกาใช้ในการทดลอง โดยในขั้นตอนการทดลองงานวิจัยชิ้นนี้ได้ นำ โมเดลอาร์มา มาใช้เป็นโมเดลหลักซึ่งเป็นโมเดลที่ใช้สำหรับแบบจำลองลำดับเวลาได้ดีที่สุดโมเดลหนึ่งและยังทดสอบร่วมกับโมเดลอื่นๆ ที่ใช้สำหรับการทำนายข้อมูล จากการทดลองพบว่าโมเดลแต่ละชนิดมีข้อดีข้อเสียต่างกัน ในส่วนของโมเดลลำดับเวลานั้นจะมีประสิทธิภาพในการทำนายแค่ในระยะเวลาอันสั้นเมื่อเวลาผ่านไปนานขึ้นประสิทธิภาพในการทำนายจะยิ่งแยลงเนื่องจากความสัมพันธ์กันของข้อมูลเมื่อเวลาผ่านไปก็จะลดลงดังนั้นจึงเป็นเหตุผลให้ต้องคอยปรับปรุงข้อมูลอินพุตให้สัมพันธ์กับค่าการพยากรณ์อากาศล่าสุดในระบบ

Guochang Wang (Wang) ได้เสนอการทำนายค่ากำลังการผลิตไฟฟ้าจากโรงไฟฟ้าพลังงานแสงอาทิตย์โดยใช้แบบจำลองการถดถอยเชิงเส้นเชิงฟังก์ชันบางส่วน ผู้เขียนได้กล่าวว่าแบบจำลองนี้เป็นลักษณะทั่วไปของแบบจำลองการถดถอยเชิงเส้นแบบดั้งเดิม แต่ช่วยให้สามารถจำลองโครงสร้างแบบไม่เชิงเส้นได้ ในการทดลองนี้ผู้เขียนได้นำแบบจำลองนี้ไปเปรียบเทียบกับวิธีการพยากรณ์แบบเดิมเช่น Multiple Linear Regression และแบบจำลองโครงข่ายประสาทเทียม โดยวิธีการแบบเดิมไม่สามารถใช้ข้อมูลที่วัดค่าพลังงานแสงอาทิตย์ ณ ปัจจุบันได้แต่จะเป็นการนำข้อมูลย้อนหลังเพื่อมาทำนายค่ากำลังการผลิตไฟฟ้าใน 1 วันข้างหน้าเท่านั้น ในขณะที่แบบจำลองที่ผู้เขียนใช้ในการทดลองได้ทำปรับปรุงพารามิเตอร์ในอัลกอริทึมให้มีความแม่นยำในการพยากรณ์มากยิ่งขึ้น

Bouchaib Zazoum (Zazoum) ได้ศึกษาความสัมพันธ์ของตัวแปรที่มีผลต่อการผลิตไฟฟ้าพลังงานแสงอาทิตย์โดยใช้โมเดลการเรียนรู้ของเครื่องที่แตกต่างกันสองวิธีคือ Support Vector Machine (SVM) และ Gaussian Process Regression (GPR) โดยพารามิเตอร์หลักที่ผู้วิจัยนำมาใช้ในการทดลองได้แก่ อุณหภูมิของแผงโซลาร์เซลล์แสงอาทิตย์ อุณหภูมิสภาพแวดล้อม พลิกซ์ของแสงอาทิตย์ เวลาและความชื้นสัมพัทธ์ จากการทดลองผลออกมาพบว่า อัลกอริทึม Matern 5/2 GPR

ให้ประสิทธิภาพการทำนายที่ดีที่สุด ในขณะที่ Cubic SVM ให้ประสิทธิภาพในการทำนายที่แย่ที่สุด และผู้เขียนยังได้ใช้ค่า Root Mean Square Error (RMSE) และค่า Mean Absolute Error (MAE) เพื่อแสดงการเปรียบเทียบประสิทธิภาพและค่าความแม่นยำของโมเดลการเรียนรู้ของเครื่องทั้งสองชนิดด้วย

C. Vennila (Vennila) ได้เสนอโมเดลที่เกิดจากการผสมผสานความรู้ด้านโมเดลการเรียนรู้ของเครื่อง เข้ากับวิธีการทางสถิติเพื่อคาดการณ์ค่ากำลังการผลิตไฟฟ้าที่ได้จากการผลิตด้วยเซลล์พลังงานแสงอาทิตย์เพื่อปรับปรุงประสิทธิภาพและความแม่นยำในการทำนาย จากผลการวิจัยพบว่า โมเดลที่สร้างขึ้นใหม่ช่วยลดต้นทุนในการจัดการระบบพลังงานไฟฟ้าเนื่องจากเมื่อเปรียบเทียบกับวิธีการเดิมโมเดลใหม่นี้ให้ประสิทธิภาพในการทำนายที่ดีกว่าการทำนายด้วยโมเดลการเรียนรู้ของเครื่องเพียงอย่างเดียว

Rachit Srivastava (Srivastava) ได้ริวถึงการทำนายค่ารังสีจากแสงอาทิตย์โดยใช้โมเดลการเรียนรู้ของเครื่องรูปแบบต่างๆ ทั้งโมเดลแบบตัวแปรเดียวและหลายตัวแปร และได้สร้างสมการทางคณิตศาสตร์เพื่อนำมาใช้คำนวณค่าพลังงานไฟฟ้าที่สามารถผลิตได้จากโรงไฟฟ้าพลังงานแสงอาทิตย์ล่วงหน้า โดยจากผลการทดลองเก็บข้อมูลและนำเรียนรู้ไปผ่านโมเดลการเรียนรู้ของเครื่องพบว่าโมเดลแรนดอมฟอเรสต์ ให้ผลออกมาดีที่สุด โดยให้ค่าความผิดพลาดน้อยที่สุดเมื่อเทียบกับโมเดลประเภทอื่นโดยมีค่าความผิดพลาดเพียง 10% เมื่อเทียบกับโมเดลประเภทอื่น

Seungbeom Nam (Nam) ได้เสนอวิธีทำนายค่าปริมาณพลังงานแสงอาทิตย์ที่จะผลิตได้ โดยใช้รูปแบบการจำลองเชิงพื้นที่และเวลาแบบผสมผสานด้วยเทคนิค kriging และได้ใช้แบบจำลองความน่าจะเป็นของ Naïve Bayes แล้วนำผลการทำนายที่ได้มาวัดค่าความผิดพลาดด้วยค่าของ Normalized mean absolute error (NMAE) ผลลัพธ์ที่ได้แสดงให้เห็นว่าการนำเทคนิค kriging หรือการนำข้อมูลจากบริเวณใกล้เคียงจุดที่สนใจมาใช้คาดการณ์ค่าพยากรณ์ทางอากาศนั้นให้ผลที่แม่นยำ และการนำข้อมูลที่คาดการณ์ได้ไปใช้ในการหาค่าความน่าจะเป็นด้วยวิธี Naïve Bayes ให้ผลออกมาดีกว่าเมื่อเปรียบเทียบกับการใช้แบบจำลองเดิม

A. Mellit (Mellit) ได้ศึกษาการทำนายค่ากำลังการผลิตไฟฟ้าของโรงไฟฟ้าขนาดใหญ่ 1 เมกะวัตต์ที่ประเทศอิตาลี โดยใช้โครงข่ายประสาทเทียมที่แตกต่างกันสามรูปแบบสำหรับวันที่มีแดดมาก วันที่มีเมฆบางส่วน และวันที่มีเมฆมาก ในบทวิจัยนี้ผู้เขียนได้แบ่งการนำโครงข่ายประสาทเทียมออกเป็นสามแบบ คือแบบแรกทำนายโดยหาความสัมพันธ์ของข้อมูลสภาพอากาศกับปริมาณพลังงานที่ผลิตได้ แบบที่สองคือการนำข้อมูลปัจจุบันกับข้อมูลในอดีตที่เก็บไว้ในระบบมาหาความสัมพันธ์กันเพื่อทำนายค่าพลังงานไฟฟ้าที่จะผลิตได้ในอนาคตโดยแบบที่สองนี้ไม่จำเป็นต้องใช้ข้อมูลทางสภาพอากาศเข้ามาหาความสัมพันธ์ร่วมด้วย และแบบสุดท้ายคือการนำแบบแรกและแบบที่สองมาผสมผสานกันโดยการทำนายกำลังการผลิตไฟฟ้าในอนาคตจะต้องหาความสัมพันธ์ของข้อมูลในอดีต

กับปัจจุบันโดยนำข้อมูลสภาพอากาศเข้ามาพร้อมด้วยเพื่อเพิ่มความแม่นยำของเครื่องมือการทำนายกำลังการผลิตไฟฟ้าโดยจากผลการทดลองก็แสดงให้เห็นว่าโมเดลแบบที่ 3 สามารถทำนายได้ดีกว่าแบบแรกและแบบที่สอง

Yuchi Sun (Sun) ได้ใช้เครื่องมือโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network, CNN) มาใช้เพื่อทำนายค่าพลังงานไฟฟ้าที่แผงโซลาร์เซลล์จะผลิตได้ในช่วงระยะเวลาสั้น โดยข้อมูลอินพุตที่ใช้ในการทำนายก็จะนำข้อมูลภาพถ่ายท้องฟ้าและประวัติการผลิตไฟฟ้าจากแสงอาทิตย์ที่บันทึกอยู่ในระบบมาทำนายค่าพลังงานไฟฟ้าที่ผลิตได้ล่วงหน้า 15 นาที จากการทดสอบในระยะเวลา 1 ปี พบว่าการใช้โมเดลนี้ในการทดสอบในวันที่มีแดดมากโมเดลนี้จะมีประสิทธิภาพการทำนายประมาณ 26.2% และในวันที่มีเมฆมากประสิทธิภาพของโมเดลจะอยู่ที่ประมาณ 16.1%

Yunjun Yu (Yu) ได้ศึกษานำโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network, RNN) มาใช้ในการทำนาย แต่เนื่องด้วยสภาพอากาศมีการเปลี่ยนแปลงอย่างมากในบางช่วงเวลาซึ่งทำให้การฝึกฝนของชุดข้อมูลมีความผันผวนซึ่งปัญหาตรงนี้ทางผู้เขียนได้นำโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้นระยะยาว (Long-Short Term Memory) ซึ่งเป็นโมเดลหนึ่งซึ่งอยู่ภายใต้ RNN มาใช้ในการปรับปรุงประสิทธิภาพในการฝึกฝนชุดข้อมูลให้มีค่าความแม่นยำที่ดีขึ้น โดยจากผลการพยากรณ์ค่าสัมประสิทธิ์ของ LSTM ในวันที่มีเมฆมากจะอยู่ที่ 0.9 ในขณะที่ค่าสัมประสิทธิ์ของ RNN จะอยู่ที่ 0.79 จากผลการพยากรณ์แสดงให้เห็นว่า LSTM มีประสิทธิภาพอย่างมากในการปรับปรุง RNN ให้ทำนายผลได้แม่นยำขึ้น

Mohammad Safayet Hossain (M. S. Hossain and H. Mahmood) ได้สร้างอัลกอริทึมสำหรับทำนายค่ากำลังการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์โดยใช้โครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้นระยะยาว (LSTM) โดยผู้วิจัยได้นำข้อมูลสถิติของค่าการแผ่รังสีของดวงอาทิตย์ในอดีตกับลักษณะของท้องฟ้าในช่วงที่เก็บข้อมูลนั้นมาใช้ อัลกอริทึม K-Mean เพื่อจัดประเภทของกลุ่มข้อมูลของค่าการแผ่รังสีกับลักษณะของท้องฟ้าซึ่งมีการเปลี่ยนแปลงอยู่ตลอดเวลาในแต่ละชั่วโมงของวัน เมื่อทำนายด้วยอัลกอริทึมนี้ผู้วิจัยกล่าวว่าช่วยลดข้อจำกัดด้านประสิทธิภาพได้เมื่อเทียบกับการใช้การทำนายร่วมกับลักษณะท้องฟ้าแบบตายตัว จากผลการวิจัยพบว่าการพยากรณ์ด้วยอัลกอริทึมดังกล่าวสามารถปรับปรุงความแม่นยำได้ถึง 33% เมื่อเทียบกับการพยากรณ์โดยใช้ข้อมูลท้องฟ้าแบบรายชั่วโมง และ มากถึง 44.6% เมื่อเทียบกับการพยากรณ์โดยใช้ข้อมูลท้องฟ้าแบบรายวัน

จากตัวอย่างงานวิจัยที่ได้นำเสนอการทำนายค่ากำลังการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์ โดยใช้โมเดลการเรียนรู้ของเครื่องชนิดต่างๆ สำหรับการทำนายมาใช้ร่วมกับปัจจัยที่ส่งผลต่อการผลิตไฟฟ้าด้วยแสงอาทิตย์ เช่น ค่าของความเข้มแสง อุณหภูมิของแผงโซลาร์เซลล์ อุณหภูมิสภาพอากาศโดยรอบ ความชื้นในอากาศ และอื่นๆ ซึ่งจากการศึกษางานวิจัยพบว่าโมเดลการเรียนรู้

ของเครื่องที่ได้รับการนิยมนำมาทำนายค่ากำลังการผลิตไฟฟ้าคือการทำโมเดลโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้นระยะยาว (LSTM) มาใช้ร่วมกับโมเดลประเภทอื่นเพื่อปรับปรุงประสิทธิภาพความแม่นยำของการทำนาย โดยงานวิจัยส่วนใหญ่จะวัดประสิทธิภาพของโมเดลโดยการเปรียบเทียบค่า Root Mean Square Error (RMSE) และค่า Mean Absolute Error (MAE) โดยสามารถสรุปเนื้อหา ข้อจำกัดและประสิทธิภาพโมเดลของแต่ละงานวิจัยได้ดังตารางที่ 1 ข้างล่างนี้

ตารางที่ 1 สรุปการทบทวนวรรณกรรม

| งานวิจัย | สรุปภาพรวมงานวิจัย | ข้อจำกัด | ประสิทธิภาพโมเดล |
|----------|--|--|--|
| [4] | บทความนี้ได้เสนอแบบจำลองโครงข่ายประสาทเทียมที่มีความแม่นยำสูงและสามารถประมวลผลข้อมูลที่มีสัญญาณรบกวนได้อย่างมีประสิทธิภาพ | - วิธีการทำนายนี้ต้องการเฟรมเวิร์คเครือข่ายขนาดใหญ่และใช้เวลาในการฝึกฝนมากขึ้น | ในวันที่มีแดด ค่า RMSE อยู่ที่ 12.5% ในวันที่มีเมฆมาก ค่า RMSE อยู่ที่ 24-36.9% |
| [5] | การเปรียบเทียบโมเดลที่ขับเคลื่อนด้วยข้อมูลที่แตกต่างกันโดยใช้ข้อมูลอินพุต Numerical Weather Prediction (NWP) ที่แตกต่างกันจากนั้นจึงสร้างชุดรูปแบบหลายโมเดลที่มีประสิทธิภาพที่ดี | - การศึกษามุ่งเน้นไปที่การคาดการณ์รังสีอาทิตย์เท่านั้นผลอาจไม่สามารถใช้กับตัวแปรอื่นได้ - ขาดการคำนึงถึงผลกระทบของเมฆต่อการคาดการณ์รังสีของแสงอาทิตย์ | โมเดลที่ดีที่สุดของบทความนี้คือ GTNN โดยมีค่า RMSE อยู่ที่ 11.2% ในขณะที่แย่ที่สุดคือ SARIMA มีค่า RMSE อยู่ที่ 12.8% และโมเดล MME พัฒนาจนได้ค่า RMSE 10.5% |
| [6] | เป็นการนำแบบจำลองของการคาดการณ์ระยะสั้นที่แตกต่างกันมาประเมินร่วมกับการนำข้อมูลทางอุตุนิยมวิทยาและข้อมูลของแบบจำลองลำดับเวลามาใช้ | - การศึกษานี้พิจารณาเฉพาะการคาดการณ์ระยะสั้นดังนั้นอาจไม่เหมาะกับการนำไปคาดการณ์ในระยะที่ยาวขึ้น | จากการนำโมเดล ARIMA ไปใช้ในการทำนายพบว่าการทำนายในเวลา 15 นาที จะให้ผลออกมาดีที่สุด โดยมีค่า RMSE เฉลี่ยอยู่ที่ 71.6 |

| งานวิจัย | สรุปภาพรวมงานวิจัย | ข้อจำกัด | ประสิทธิภาพโมเดล |
|----------|---|--|--|
| | ในการทำนาย | - ขาดการพิจารณาถึงผลกระทบของแผงหรือฮาร์ดแวร์อื่น ต่อความแม่นยำในการทำนาย | W/m^2 ซึ่งให้ประสิทธิภาพที่ดีกว่าโมเดลอื่นในบทความเดียวกัน |
| [7] | เป็นการนำฟังก์ชันการถดถอยเชิงเส้นของฟังก์ชันเบี่ยงเบนมาทำนายข้อมูลไม่เชิงเส้น โดยนำไปเปรียบเทียบกับโมเดล MLP และ RBF | - บทความนี้มุ่งเน้นเฉพาะการคาดการณ์รายวันโดยไม่คำนึงการคาดการณ์ระยะยาว - มีพารามิเตอร์มาก (~17) ส่งผลต่อการเลือกมาใช้ | ผลการทดลองแสดงให้เห็นว่าการนำฟังก์ชัน PFLRM มาใช้ช่วยให้ประสิทธิภาพการทำนายดีขึ้นโดยค่า RMSE อยู่ที่ ~63% |
| [8] | ใช้แบบจำลองทาง ML เพื่อสำรวจความสัมพันธ์ของพารามิเตอร์อินพุตกับพลังงานแสงอาทิตย์ โดยใช้อัลกอริทึม 2 วิธีคือ SVM และ GPR | - ยังขาดพารามิเตอร์ที่จะทำให้ผลของการทำนายดีขึ้น เช่น ขนาดของแผงโซลาร์เซลล์ ฝุ่นที่เกาะอยู่บนแผง เป็นต้น | โมเดลที่ให้ประสิทธิภาพการทำนายที่ดีที่สุดสำหรับงานวิจัยนี้คือ Matern 5/2 GPR ซึ่งค่า RMSE อยู่ที่ 7.967% |
| [9] | นำโมเดลทาง ML ไปทำนายร่วมกับข้อมูลทางสถิติเพื่อเพิ่มประสิทธิภาพการทำนายและลดต้นทุนการจัดตำแหน่ง | - ขาดการอธิบายโมเดล ML ที่นำมาใช้และเทคนิคทางสถิติที่ใช้ใน hybrid model | ในงานวิจัยนี้ไม่ได้ระบุถึงประสิทธิภาพของโมเดลที่ใช้ในการทดลอง |
| [10] | เป็นการทำนายค่าการแผ่รังสีของดวงอาทิตย์โดยใช้โมเดล machine learning ทั้ง univariate และ multi-variate model และสร้างสมการทางคณิตศาสตร์เพื่อคำนวณค่ากำลังการผลิตไฟฟ้าของเซลล์พลังงานแสงอาทิตย์ | - ความแม่นยำของการทำนายขึ้นอยู่กับสภาพอากาศในระหว่างการทดลอง และความถูกต้องของข้อมูลที่นำมาใช้ในการทำนาย | ผลการทำนายที่ได้โมเดล Random Forest ให้ประสิทธิภาพในการทำนายดีที่สุดเมื่อเทียบกับโมเดลอื่น โดยมีค่า Overall % error ในการทำนาย เฉลี่ยรวมอยู่ที่ 10.06% |

| งานวิจัย | สรุปภาพรวมงานวิจัย | ข้อจำกัด | ประสิทธิภาพโมเดล |
|----------|---|--|--|
| [11] | เป็นการทำนายโดยนำเทคนิค kriging มาใช้ในการจำลองเชิงพื้นที่และเวลา โดยนำ Naïve Bayes มาใช้คาดการณ์ความน่าจะเป็นของข้อมูล | - ขาดการคำนึงผลกระทบของสภาพอากาศในระหว่างการทดลอง เช่น การบังแสง และการมีเมฆปกคลุมซึ่งส่งผลต่อกำลังการผลิต และ ไม่ได้นำไปเปรียบเทียบกับโมเดลอื่น | จากผลการทดสอบพบว่าค่าของ MAE นั้นมีค่าต่ำกว่า 10% ซึ่งแสดงให้เห็นว่าโมเดลนี้มีความแม่นยำในการทำนาย |
| [12] | ใช้โครงข่ายประสาทเทียม (ANN) ประเมินข้อมูลพยากรณ์อากาศจากฐานข้อมูลช่วงระยะเวลา 1 ปี ของการแผ่รังสีของดวงอาทิตย์ อุณหภูมิบนเซลล์ และกำลังไฟฟ้าจากโรงไฟฟ้า 1 MW ที่อิตาลี จากนั้นจึงทำนาย | - วิธีการนี้ใช้แบบจำลอง ANN แบบ AFFNN ที่แตกต่างกันสามแบบสำหรับสภาพอากาศช่วงกลางวันที่แตกต่างกัน ซึ่งมีความซับซ้อนมากกว่าการใช้แบบจำลองเพียงแบบเดียว | จากผลการทดสอบพบว่าในวันที่มีแดด ค่า RMSE จะอยู่ที่ 3.3% ในวันที่มีเมฆบางส่วน ค่า RMSE จะอยู่ที่ 4.2% และในวันที่มีดครึ้ม ค่า RMSE จะอยู่ที่ 5.3% |
| [13] | ใช้ Convolutional Neural Network (CNN) ทำนายค่ากำลังการผลิตโดยใช้ข้อมูลรูปภาพท้องฟ้าและประวัติการผลิตไฟฟ้าด้วยเซลล์พลังงานแสงอาทิตย์เพื่อทำนายล่วงหน้า 15 นาที | - ในกระบวนการเรียนรู้ของ CNN จะใช้เวลานานในการเรียนรู้เสมอ - ความแม่นยำอาจลดลงเมื่อกรอบระยะเวลาเกินที่กำหนด | จากผลการทดสอบพบว่าในวันที่แดดมาก โมเดลนี้ให้ความสามารถในการทำนายอยู่ที่ 26.2% และในวันที่มีเมฆอยู่ที่ 16.1% |
| [14] | เป็นการนำโมเดล LSTM มาปรับปรุงความแม่นยำในการทำนายของโมเดล RNN เพื่อให้เกิดผลการทำนายที่ดีขึ้น โดยใช้ข้อมูลอากาศย้อนหลัง | - สภาพท้องฟ้าในแต่ละวันส่งผลต่อการทำนายอย่างมาก โดยเฉพาะวันเมฆมากส่งผลให้การทำนายคลาดเคลื่อนมาก | จากผลการทดสอบพบว่าโมเดล RNN ที่ถูกปรับปรุงด้วย LSTM นั้นให้ค่า R^2 มีค่ามากกว่า 0.9 สำหรับการทำนายรายชั่วโมง |

| งานวิจัย | สรุปภาพรวมงานวิจัย | ข้อจำกัด | ประสิทธิภาพโมเดล |
|-----------------|--|--|---|
| [15] | เป็นการนำโครงข่ายประสาทเทียม(NN) ของหน่วยความจำระยะสั้นระยะยาว (LSTM) และข้อมูลพยากรณ์อากาศแบบสังเคราะห์ที่ถูกสร้างขึ้นโดยการบูรณาการความรู้ทางสถิติของข้อมูลการแผ่รังสีของดวงอาทิตย์ในอดีตกับข้อมูลการพยากรณ์ลักษณะของท้องฟ้ามาใช้ในการทำนายการผลิตไฟฟ้าด้วยเซลล์แสงอาทิตย์ | <ul style="list-style-type: none"> - ความแม่นยำของการทำนายนั้นขึ้นอยู่กับข้อมูลที่นำมาใช้นั้น มีความถูกต้องมากน้อยหรือไม่ - การพยากรณ์อากาศแบบสังเคราะห์นั้นอิงข้อมูลในอดีตซึ่งสภาพอากาศจริงอาจมีความคลาดเคลื่อน | จากผลการทดสอบพบว่าประสิทธิภาพการทำนายดีขึ้น ~33% เมื่อเทียบกับการพยากรณ์รายชั่วโมง และ มากถึง 44.6% เมื่อเทียบกับการพยากรณ์รายวัน |
| (Wai) | เป็นการนำโมเดล DNN ไปปรับปรุงประสิทธิภาพการทำงานโดยนำโมเดล LSTM เข้ามาใช้ร่วมกันในการทำนายเพื่อให้ได้ผลออกมาที่แม่นยำขึ้น | ยังขาดปัจจัยบางตัวที่สามารถเพิ่ม accuracy ให้กับโมเดลได้ | จากผลการทดลองพบว่าโมเดลดังกล่าวสามารถเพิ่มประสิทธิภาพการทำนายในระยะเวลา 1 สัปดาห์ของทั้ง 6 แพลนท์ได้ค่าความถูกต้องอยู่ที่ 97% |
| (villegas-Mier) | ได้มีการเสนอวิธีการทำนายค่ารังสีจากดวงอาทิตย์โดยใช้การปรับอัลกอริทึม Random forest ให้เหมาะสม ซึ่งผลลัพธ์ได้แสดงถึงความแม่นยำเมื่อเทียบกับแบบทั่วไป | <ul style="list-style-type: none"> - การศึกษานี้ยังขาดการพิจารณาถึงผลกระทบของปัจจัยภายนอก - ไม่ได้ให้ข้อมูลการคำนวณที่จำเป็นอย่างละเอียด | จากผลการทดลองพบว่าโมเดล Random forest ให้ accuracy ในการทำนายอยู่ที่ 95% |
| (Wang) | เป็นการนำโมเดล LSTM มาปรับปรุงการทำนายของโมเดล RNN | - ต้องใช้ข้อมูลจำนวนมาก และในบางกรณีไม่สามารถคาดการณ์ค่าพลังงานได้ | จากผลการทดลอง โมเดล LSTM-RNN ให้ค่า accuracy 96.19% |

จากการทบทวนวรรณกรรมวิทยานิพนธ์ฉบับนี้จึงได้จัดทำขึ้นเพื่อเสนองานวิจัยที่ใช้สำหรับการทำนายค่าพลังงานไฟฟ้าที่ผลิตได้จากโรงไฟฟ้าพลังงานแสงอาทิตย์ที่ตั้งอยู่ในประเทศไทย โดยนำโมเดลการเรียนรู้ของเครื่องเข้ามาใช้ในการทำนาย เพื่อเป็นแนวทางในการจัดการทรัพยากรพลังงานไฟฟ้าให้ดียิ่งขึ้น ส่งผลให้ลดต้นทุนในการจัดการพลังงานไฟฟ้าที่จะจ่ายเข้าสู่โครงข่ายไฟฟ้าหลัก โดยจะนำข้อมูลพารามิเตอร์ที่เป็นปัจจัยในการผลิตไฟฟ้าด้วยพลังงานแสงอาทิตย์ที่บันทึกได้มาทำนาย ร่วมกับการนำข้อมูลการผลิตที่ได้ในอดีตมาประเมินหาประสิทธิภาพเพื่อเพิ่มความแม่นยำให้กับโมเดลที่ได้จัดทำขึ้น จากตัวอย่างงานวิจัยที่ศึกษามาได้มีการนำโมเดลการเรียนรู้ของเครื่องหลายชนิดมาใช้ในการทำนายและได้เปรียบเทียบประสิทธิภาพการทำนายของแต่ละโมเดลว่าโมเดลชนิดใดให้ประสิทธิภาพในการทำนายที่ดีที่สุด แต่ยังไม่พบผลงานวิจัยที่ใช้โมเดลการวิเคราะห์การถดถอยเชิงเส้น มาใช้ทำนายร่วมกับโมเดลแรนดอมฟอเรสต์ ดังนั้นวิทยานิพนธ์ฉบับนี้จึงได้เสนอวิธีการทำนายค่ากำลังการผลิตไฟฟ้าจากพลังงานแสงอาทิตย์โดยนำโมเดลข้างต้นไปเรียนรู้ข้อมูลเพื่อวิเคราะห์และทำนายค่ากำลังการผลิตโดยทำนายเปรียบเทียบกับโมเดลการเรียนรู้ของเครื่องอีก 2 ชนิด ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และอาร์มีนา (ARIMA) จากนั้นประเมินประสิทธิภาพการทำนายด้วยค่าของรากที่สองเฉลี่ยของค่าความผิดพลาด (Root Mean Square Error, RMSE) ค่าเฉลี่ยสัมบูรณ์ของค่าความผิดพลาด (Mean Absolute Error, MAE) และค่าของ R-squared

1.3 วัตถุประสงค์ของวิทยานิพนธ์

1. เพื่อสร้างโมเดลการเรียนรู้ของเครื่องที่จะนำมาใช้ทำนายค่ากำลังการผลิตไฟฟ้าพลังงานแสงอาทิตย์ได้แม่นยำยิ่งขึ้น
2. เพื่อพิจารณาถึงปัจจัยภายนอกที่ส่งผลกระทบต่อกำลังการผลิตและความแม่นยำในการทำนาย
3. เพื่อประมาณการณ์ผลผลิตที่ได้จากพลังงานแสงอาทิตย์ล่วงหน้า เพื่อเตรียมไฟฟ้าจากแหล่งอื่นจ่ายเข้าระบบโครงข่ายอย่างมีประสิทธิภาพที่สุด

1.4 ขอบเขตของวิทยานิพนธ์

1. พิจารณาข้อมูลย้อนหลังเฉพาะช่วงปีพุทธศักราช 2563 ถึง เดือนสิงหาคม 2566
2. ทดสอบการทำนายโดยนำข้อมูลเฉพาะโซน C ที่โรงไฟฟ้าพลังงานแสงอาทิตย์ของ การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) ขนาด 5 เมกะวัตต์ ตั้งอยู่ อ.ทับสะแก จ.ประจวบคีรีขันธ์

3. ใช้ภาษา Python ทำงานผ่านโปรแกรม Jupyter Notebook เพื่อสร้างโมเดลการเรียนรู้ สำหรับการทำนายค่าพลังงานไฟฟ้าที่ผลิตจากเซลล์แสงอาทิตย์

1.5 ขั้นตอนการดำเนินการ

1. กำหนดวัตถุประสงค์และขอบเขตของงานวิจัย
2. ศึกษางานวิจัยที่นำโมเดลการเรียนรู้ของเครื่องมาใช้ในการทำนายพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์
3. ศึกษาอัลกอริทึมการวิเคราะห์การถดถอยเชิงเส้นและแรนดอมฟอเรสต์สำหรับการถดถอย เพื่อกำหนดรูปแบบการเรียนรู้ข้อมูลด้วยโปรแกรมคำสั่งที่กำหนด
4. ศึกษาวิธีการปรับจูนค่าไฮเปอร์พารามิเตอร์ของโมเดลเพื่อพัฒนาประสิทธิภาพของโมเดลการทำนาย
5. ศึกษาวิธีการสร้างโมเดลการเรียนรู้ด้วยภาษา Python ผ่านโปรแกรม Jupyter Notebook
6. ศึกษาวิธีการผสานการเรียนรู้ของโมเดลการเรียนรู้ที่เสนอรวมไปถึงโมเดลสำหรับบустประสิทธิภาพการทำนาย
7. ออกแบบและเขียนโปรแกรมสำหรับงานทำนายพลังงานไฟฟ้าที่ผลิตจากเซลล์แสงอาทิตย์
8. ฝึกและทดสอบโมเดลการเรียนรู้ที่เสนอ และโมเดลการเรียนรู้ที่นำมาเปรียบเทียบผลการทำนาย
9. ประเมินผลและเปรียบเทียบการทำนายของโมเดลการเรียนรู้ของเครื่องแต่ละชนิด
10. วิเคราะห์ สรุป และรายงานผลการวิจัยของวิทยานิพนธ์
11. เรียบเรียงข้อมูลผลการทดลองและจัดทำรูปเล่มวิทยานิพนธ์
12. นำเสนอวิทยานิพนธ์กับกรรมการสอบต่อไป

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. โมเดลที่สร้างขึ้นจะช่วยพัฒนาประสิทธิภาพในการทำนายให้แม่นยำขึ้น
2. ได้ทราบถึงปัจจัยภายนอกที่เพิ่มเข้ามาในการทดลองนั้นส่งผลต่อการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์หรือไม่
3. ได้ข้อมูลการทำนายการผลิตไฟฟ้าล่วงหน้า เพื่อนำผลการทำนายไปใช้จัดการระบบจ่ายไฟฟ้าได้อย่างเหมาะสม

บทที่ 2

ความรู้พื้นฐานและทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องในงานวิจัยนี้จะช่วยในการเตรียมข้อมูลที่จะนำไปใช้ในการทำนายพลังงานไฟฟ้าที่ผลิตได้จากโรงไฟฟ้าพลังงานแสงอาทิตย์ผ่านโมเดลการเรียนรู้ของเครื่อง และเพื่อเป็นแนวทางในการวิเคราะห์ข้อมูลปัจจัยต่างๆ ที่ส่งผลต่อการผลิตพลังงานไฟฟ้าด้วยเซลล์แสงอาทิตย์

2.1 การเตรียมข้อมูลสำหรับการทำนายด้วยโมเดลการเรียนรู้ของเครื่อง

โดยปกติในโรงไฟฟ้าจะมีระบบฐานข้อมูล เพื่อใช้ในการตรวจสอบข้อมูลพลังงานไฟฟ้าที่เซลล์พลังงานแสงอาทิตย์สามารถผลิตได้ในช่วงเวลาต่างๆ เพื่อเป็นการลดปัญหาในการรับส่งข้อมูลนั้น โรงไฟฟ้าจึงจำเป็นต้องส่งข้อมูลไปที่ฐานเก็บข้อมูลสำรอง ก่อนส่งข้อมูลทั้งหมดเข้าสู่ฐานเก็บข้อมูลหลัก ซึ่งอาจทำให้เกิดความล่าช้าในการรับส่งข้อมูลมากเกินไปและเพื่อเป็นการหลีกเลี่ยงปัญหาการสูญหายของข้อมูลขณะส่งกลับมายังศูนย์รวบรวมข้อมูล ดังนั้นเราจึงต้องนำโครงข่ายประสาทเทียมเข้ามาใช้ร่วมในการแก้ไขปัญหาในส่วนนี้โดยระบบจะนำข้อมูลสภาพอากาศมาใช้ในการชดเชยข้อมูลส่วนที่ขาดหายไปจากการรับข้อมูลจากเซลล์พลังงานแสงอาทิตย์ จากนั้นนำข้อมูลที่อ่านค่าได้ ณ ปัจจุบันส่งไปยังโมเดลที่จัดเตรียมไว้ เพื่อคาดการณ์กำลังการผลิตไฟฟ้าพลังงานแสงอาทิตย์เพื่อสามารถตอบสนองตามเวลาจริงได้

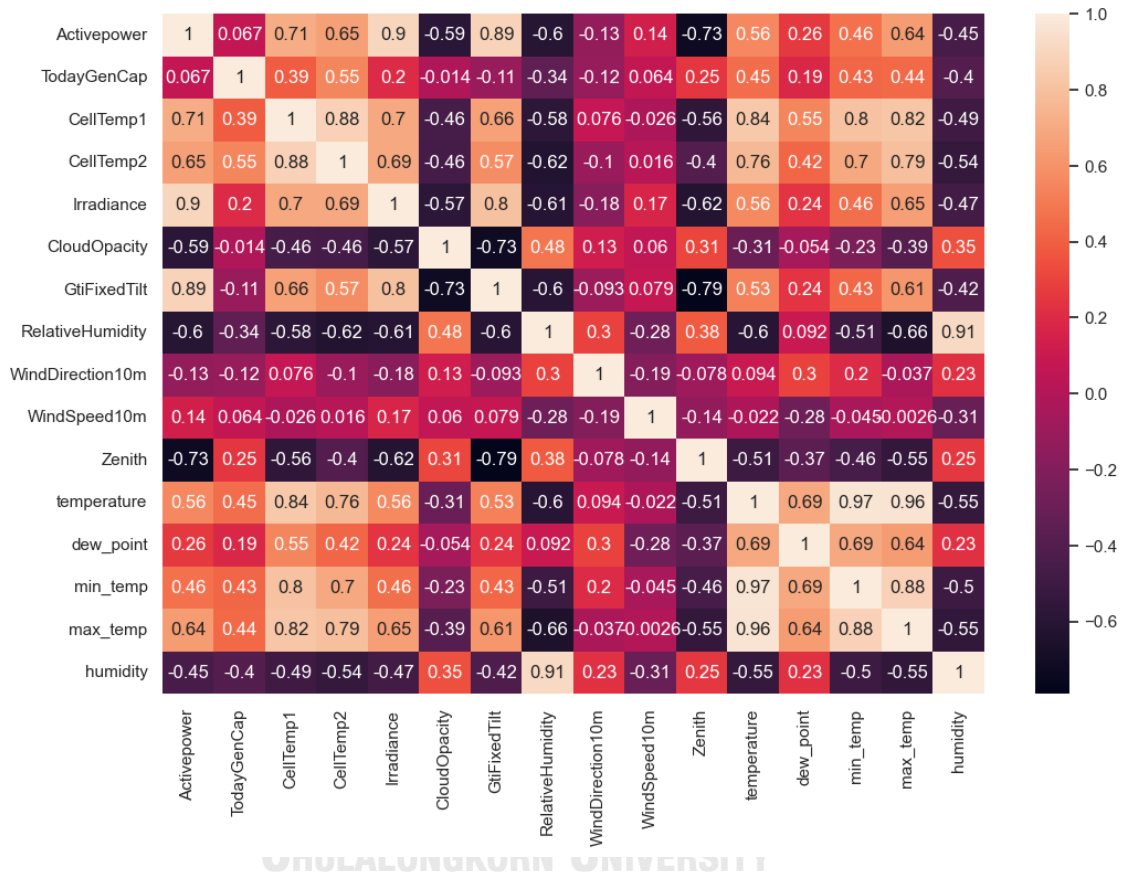
2.1.1 การวิเคราะห์ความสัมพันธ์ของข้อมูล

ในการแก้ปัญหาการขาดข้อมูลการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์แบบเรียลไทม์ จึงต้องหาปัจจัยที่มีความสัมพันธ์เกี่ยวข้องกับการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์ ในการศึกษานี้ได้เลือกการวิเคราะห์ค่าสัมประสิทธิ์ความสัมพันธ์ของเพียร์สัน (Pearson's Correlation Analysis) ดังสมการต่อไปนี้

$$R(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

โดยที่ \bar{x} คือค่าเฉลี่ยของกลุ่มตัวอย่างที่ 1, \bar{y} คือค่าเฉลี่ยของกลุ่มตัวอย่างที่ 2 ซึ่งค่าของ $R(x, y)$ มีค่าอยู่ระหว่าง 0 ถึง 1 หากค่าของ $R(x, y)$ มีค่าเข้าใกล้ 1 แสดงว่าข้อมูลของทั้งสองตัวอย่งนั้นมีความสัมพันธ์กันมาก แต่หากค่าเข้าใกล้ 0 แสดงว่ากลุ่มตัวอย่างทั้งสองนั้นไม่มีความสัมพันธ์ต่อกัน

หรือมีแต่น้อยมาก จากการนำตัวแปรทางสภาพอากาศมาวิเคราะห์ความสัมพันธ์เทียบกับค่าพลังงานที่ผลิตได้จากโรงไฟฟ้าแสดงดังรูปที่ 1 โดยในรูปจะแสดงค่าความสัมพันธ์ทางสถิติระหว่างข้อมูลตัวแปรทั้งหมด ในงานวิจัยนี้จะเลือกข้อมูลค่าความสัมพันธ์ระหว่างพลังงานไฟฟ้าที่ผลิตได้กับตัวแปรทางสภาพอากาศชนิดอื่นเพื่อวิเคราะห์ว่าตัวแปรใดส่งผลต่อค่าพลังงานไฟฟ้าที่จะผลิตได้จากเซลล์แสงอาทิตย์



รูปที่ 1 แผนภาพความร้อนแสดงความสัมพันธ์ระหว่างข้อมูลตัวแปร

2.1.2 การกำหนดมาตรฐานของข้อมูล (Standardization)

เนื่องจากข้อมูลที่น่ามาใช้ นั้นมีขนาดการเปรียบเทียบข้อมูลที่แตกต่างกัน ดังนั้นจึงมีความจำเป็นต้องกำหนดมาตรฐานของข้อมูลเพื่อเป็นการปรับปรุงความแม่นยำของโมเดลที่ใช้ในการทำนาย ในการแก้ไขปัญหานี้ได้มีการเสนอวิธีแปลงข้อมูลเดิมให้เป็นข้อมูลตัวใหม่ที่ปราศจากหน่วยและมิติของข้อมูล ซึ่งในการศึกษานี้ได้เสนอการกำหนดมาตรฐานข้อมูลสามารถแสดงได้ดังสมการต่อไปนี้

$$x_{new} = \frac{x_i - \mu}{\sigma} \quad (2)$$

โดย μ และ σ คือค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง x_i

หลังจากที่ได้ปรับมาตรฐานของข้อมูลให้ไม่มีหน่วยทางกายภาพแล้วนั้นจะต้องคืนค่าจริงที่สอดคล้องให้กลับไปยังมาตรฐานเดิมของข้อมูล ผ่านสมการดังต่อไปนี้

$$y = y_{pred} \times \sigma + \mu \quad (3)$$

โดย y คือค่าผลลัพธ์ของข้อมูลจริง ส่วน y_{pred} คือค่าผลลัพธ์ที่ได้จากการทำนายผ่านโมเดลการเรียนรู้ของเครื่อง

2.1.3 กระบวนการนอร์มัลไลเซชัน (Normalization)

การทำนอร์มัลไลเซชันเป็นการลดความซับซ้อนของชุดข้อมูลเนื่องจากตัวแปรแต่ละชนิดมีช่วงค่าของข้อมูลที่แตกต่างกัน กระบวนการนอร์มัลไลเซชันจะปรับช่วงของข้อมูลให้มีค่าอยู่ระหว่าง 0 กับ 1 เพื่อที่จะให้ตัวแปรทุกตัวอยู่ในช่วงเดียวกัน โดยสมการสำหรับการนอร์มัลไลเซชันสามารถทำได้ดังสมการด้านล่างนี้

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

2.2 โมเดลการเรียนรู้ของเครื่องสำหรับการทำนายพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์

2.2.1 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression)

การวิเคราะห์การถดถอยเชิงเส้น คือ การนำข้อมูลหรือตัวแปรตั้งแต่สองตัวขึ้นไปมาหาความสัมพันธ์ของข้อมูล โดยค่าของข้อมูลนั้นจะแสดงออกในรูปแบบแนวเส้นตรงหรือค่าใกล้เคียงเส้นตรง จากนั้นจึงนำผลที่ได้จากการทำ Linear Regression ไปวิเคราะห์ถึงผลลัพธ์ที่ได้จากชุดข้อมูล โดยในบทวิจัยนี้จะนำโมเดลการถดถอยเชิงเส้นแบบหลายตัวแปรมาใช้ในการจัดการชุดข้อมูล

ในการถดถอยเชิงเส้นแบบหลายตัวแปร เอาต์พุต p จะถือว่าเป็นฟังก์ชันเชิงเส้น ผลรวมน้ำหนักถ่วงของตัวแปรอินพุตหลาย ๆ ตัว o_1, o_2, \dots, o_d และ ค่าบวกรวมที่แปรผันตามเวลา ในบทวรรณกรรมทางสถิติกล่าวว่ารูปแบบนี้เรียกว่าฟังก์ชันการถดถอย แต่นักสถิติเรียกว่าฟังก์ชันหลายตัวแปรเมื่อมีหลายผลลัพธ์ (M. Vetri Selvi) ซึ่งแสดงแทนได้เป็น p_1 ถึง p_d โดยแบบจำลองเชิงเส้นหลายตัวแปรสามารถแสดงแทนได้ดังเมทริกซ์สมการด้านล่างนี้

การแสดงความสัมพันธ์เชิงเส้นระหว่างตัวแปรอินพุตและเอาต์พุต

$$p^t = g(o^t | \beta_0, \beta_1, \dots, \beta_d) + \varepsilon = \beta_0 + \beta_1 o_1^t + \beta_2 o_2^t + \dots + \beta_d o_d^t + \varepsilon \quad (4)$$

ในกรณีของการถดถอยเชิงเส้นแบบตัวแปรเดียว สามารถอนุมานได้ว่าค่า \mathcal{E} เป็นปกติที่มีค่าเฉลี่ยเป็นศูนย์และค่าความแปรปรวนคงที่ และในการเพิ่มความแม่นยำไปได้มากที่สุดคือการลดค่าผลรวมของค่าความคลาดเคลื่อนกำลังสองให้น้อยที่สุด ซึ่งโดยปกติจะหาอนุพันธ์ปกติเทียบกับพารามิเตอร์ $\beta_j, j = 0, \dots, d$ โดยจะได้สมการทั่วไปดังนี้

$$E(\beta_0, \beta_1, \dots, \beta_d | o) = \frac{1}{2} \sum_t (p^t - \beta_0 - \beta_1 o_1^t - \beta_2 o_2^t - \dots - \beta_d o_d^t)^2 \quad (5)$$

$$\sum_t p^t = N\beta_0 + \beta_1 \sum_t o_1^t + \beta_2 \sum_t o_2^t + \dots + \beta_d \sum_t o_d^t \quad (6)$$

ค่าของ $\beta_0, \beta_1, \dots, \beta_d$ ได้มาจากวิธีกำลังสองที่น้อยที่สุด โดยที่ β คำนวณโดยใช้

$$\beta = (O^T O)^{-1} O^T p \quad (7)$$

ค่าสัมประสิทธิ์การถดถอย $\beta_0, \beta_1, \dots, \beta_d$ จะถูกกำหนดโดยการใช้ข้อมูลชุดฝึกอบรมที่ผ่านมา จากนั้นการคาดการณ์ของตัวแปรในเมทริกซ์ “ P ” สำหรับข้อมูลชุดใหม่ของตัวแปรเมทริกซ์ “ O ” แทนได้เป็น $P = \beta O$ โดยที่ O คือชุดข้อมูลอินพุตใหม่ ซึ่งนำไปใช้คาดการณ์เอาต์พุต P จากสัญญาณของ β_j แสดงให้เห็นว่า O_j มีความสัมพันธ์เชิงบวกหรือลบกับผลลัพธ์ p_j อันดับของตัวแปรอินพุตถูกกำหนดด้วยค่าสัมบูรณ์ของ β_j และช่วยกำจัดอินพุตที่มีค่า β_j ใกล้ศูนย์หรือเท่ากับศูนย์ออกจากระบบ

2.2.2 แรนดอมฟอเรสต์สำหรับการถดถอย (Random Forest for Regression)

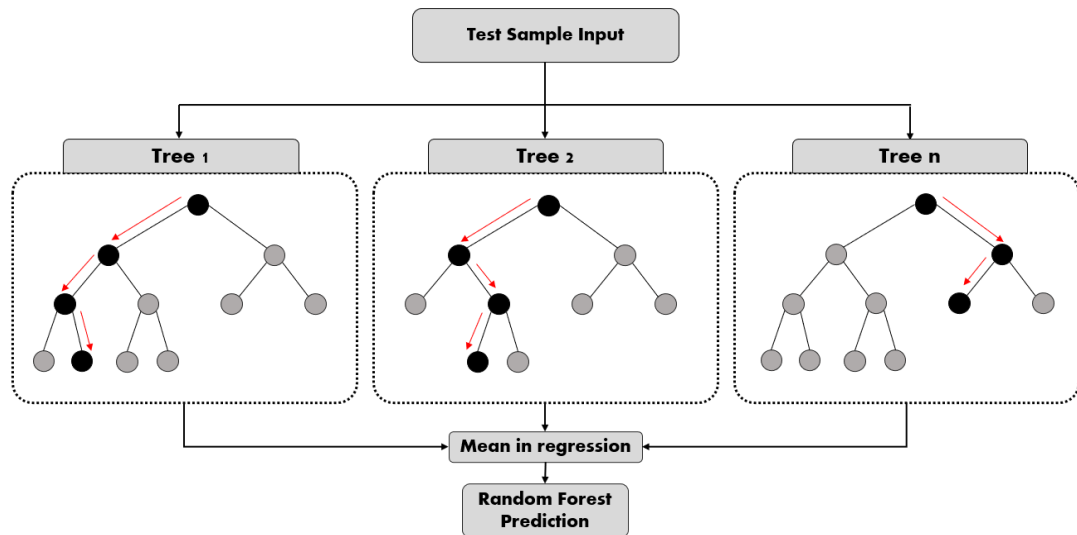
เป็นโมเดลการเรียนรู้ของเครื่องที่สร้างขึ้นมาจากการนำต้นไม้ตัดสินใจ (decision tree) มาสร้างโมเดล แต่มีความแตกต่างกันตรงที่แรนดอมฟอเรสต์สามารถเพิ่มจำนวน tree ได้มากขึ้นส่งผลให้มีประสิทธิภาพและความแม่นยำที่สูงขึ้น โมเดลแรนดอมฟอเรสต์มีหลักการคือจะมีการสร้างต้นไม้ตัดสินใจขึ้นและสร้างต้นไม้ขึ้นมาแทนที่การแตกกิ่ง (node) จากต้นเดิม เพื่อที่จะค้นหาคุณลักษณะที่ดีที่สุดจากชุดย่อยของคุณลักษณะแบบสุ่ม (villegas-Mier) โดยในงานวิจัยนี้จะเลือกใช้คุณลักษณะข้อผิดพลาดจากข้อมูลที่ไม่ถูกนำมาทดสอบในชุดของตัวอย่างหรือที่เรียกกันว่า Out of bag error (OOBE) โดยสามารถหาได้จากสมการด้านล่างนี้

$$OOBE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y})^2 \quad (8)$$

อีกคุณลักษณะที่สำคัญคือการวัดความสำคัญของตัวแปร (Variable Importance measure, VI) ซึ่งได้มาจากการเรียงสับเปลี่ยนคุณลักษณะและหาค่าเฉลี่ยของความแตกต่างโดยใช้ค่าของ OOBE ก่อนและหลังการสับเปลี่ยนต้นไม้ทั้งหมด โดยสามารถหาได้จากสมการด้านล่างนี้

$$VI(X^j) = \frac{1}{q} \sum_{l=1}^q (\overline{OOBE}_l - OOBE_l) \quad (9)$$

โดยที่ \overline{OOBE} คือค่าเฉลี่ยของ OOBE ที่ประมาณเอาไว้ และจากรูปที่ 2 จะแสดงหลักการทำงานของ โมเดลแรนดอมฟอเรสต์สำหรับการถดถอย



รูปที่ 2 แผนผังแสดงรูปแบบของโมเดลแรนดอมฟอเรสต์ชนิดถดถอย

2.2.3 การปรับจูนค่าไฮเปอร์พารามิเตอร์ (Hyperparameter Optimization)

ไฮเปอร์พารามิเตอร์คือการตั้งค่าที่ผู้ใช้สามารถกำหนดค่าต่างๆ เองก่อนที่จะนำชุดของข้อมูล นั้นไป เรียนรู้เพื่อเพิ่มประสิทธิภาพให้กับโมเดลการเรียนรู้ของเครื่อง เช่นในกรณีของอัลกอริทึมของ ต้นไม้ตัดสินใจ ที่จะใช้การประมาณค่าและจำนวนของกิ่งมาใช้ในการตัดสินใจ ในทางกลับกัน ค่าพารามิเตอร์ของโมเดล เช่น ค่าน้ำหนักของโครงข่ายประสาทเทียม จะสามารถเรียนรู้ได้จากการนำ ชุดข้อมูลไปเรียนรู้ จากนั้นจึงปรับจูนค่าพารามิเตอร์ที่เหมาะสมในภายหลังจากกระบวนการเรียนรู้ เสร็จสิ้น ซึ่งการปรับค่าพารามิเตอร์นั้นจะขึ้นอยู่กับผลการทดลองมากกว่าทฤษฎี เมื่อเวลาผ่านไปชุด ข้อมูลได้รับการทดสอบจนผลลัพธ์ออกมาดีที่สุด แต่ก็อาจนำไปสู่ปัญหาการโอเวอร์ฟิตติ้ง ซึ่งการ หลีกเลี่ยงการเกิดโอเวอร์ฟิตติ้งด้วยการตรวจสอบความถูกต้องของผลลัพธ์ และจะต้องเลือกอัลกอริ ทึมที่เหมาะสมสำหรับการปรับค่าพารามิเตอร์ด้วย

2.2.4 การค้นหาชุดของตัวปรับจูนข้อมูลของไฮเปอร์พารามิเตอร์

โมเดลแรนดอมฟอเรสต์จะมีค่าพารามิเตอร์หลักที่จะต้องปรับจูนอยู่ทั้งหมด 3 พารามิเตอร์ โดยมีพารามิเตอร์ที่ต้องปรับดังนี้

พารามิเตอร์ที่ 1 คือ ค่าของ $n_estimators$ หรือการปรับค่าของตัวประมาณ ยิ่งมีค่ามากยิ่งขึ้น แม้จะมีผลผลิตที่ลดลง ที่เห็นได้ชัดคือเมื่อต้นไม้มียิ่งจำนวนมากขึ้น จะทำให้เวลาที่ใช้นานขึ้น สิ่งที่น่าคาดหวัง

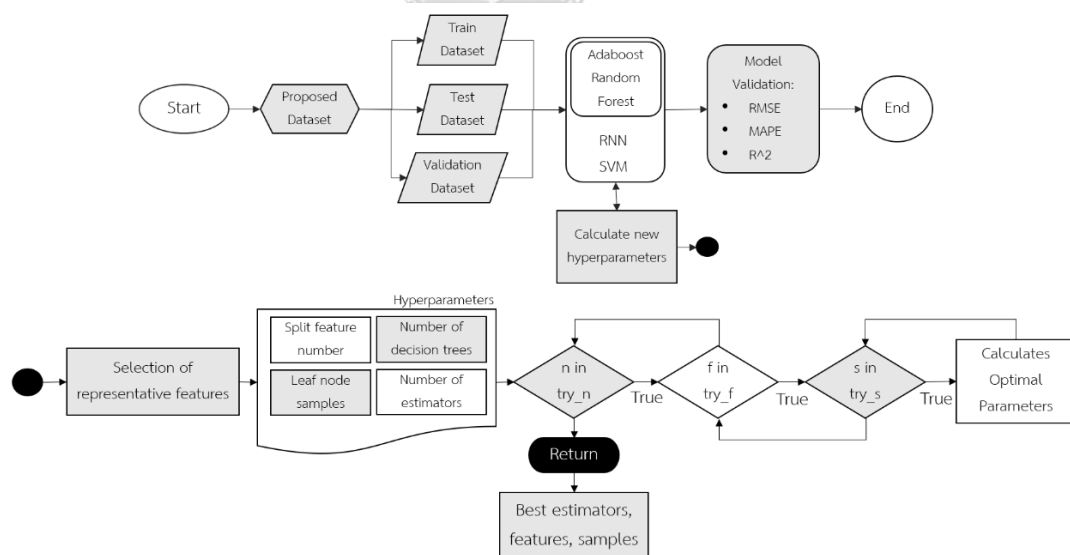
จากอัลกอริทึมนี้คือการค้นหาจำนวนต้นไม้ตัดสินใจที่จะทำให้ได้รับความแม่นยำที่สุด โดยที่ยังรักษาสมดุลของเวลาในการเรนข้อมูลที่เหมาะสม

พารามิเตอร์ที่ 2 คือ ค่าของ $max_features$ หรือค่าสูงสุดก่อนที่จะถูกตัดไปขึ้นโหนดใหม่ ซึ่งเมื่อนำต้นไม้มาพิจารณาแล้วหากพบว่าแตกต่างจากต้นอื่นๆ และความสัมพันธ์ของชุดข้อมูลเหล่านี้สัมพันธ์กันน้อยมาก นี่จึงเป็นสิ่งที่จะทำให้สามารถประเมินอิทธิพลของแต่ละคุณลักษณะแยกกันได้

พารามิเตอร์ที่ 3 คือ ค่าของ $maximum_depth$ หรือการกำหนดค่าความลึกของต้นไม้เพื่อหลีกเลี่ยงการเกิด *overfitting* โดยต้นไม้แต่ละต้นจะมีความลึกพอแต่จะไม่ลึกเกินไปจนเกิดการ *overfitting* โดยสามารถสร้างข้อจำกัดของแต่ละพารามิเตอร์ได้โดยการปรับพารามิเตอร์ (min_sample_split), (min_sample_leaf), ($min_weight_fraction_leaf$), หรือ (max_leaf_nodes)

สำหรับการค้นหาไฮเปอร์พารามิเตอร์ที่ดีที่สุด จะสร้างฟังก์ชัน RF Optimizer Algorithm โดยข้อมูลที่นำเข้าไปในฟังก์ชันคือ n = จำนวนของตัวประมาณ, f = จำนวนของคุณสมบัติ, และ s = จำนวนของตัวอย่าง ซึ่งฟังก์ชันนี้จะช่วยดำเนินการหาค่าพารามิเตอร์ที่ดีที่สุด

จากรูปที่ 3 แสดงรูปแบบการทำงานของโปรแกรมโดยได้แบ่งชุดข้อมูลออกเป็นชุดทดสอบและชุดตรวจสอบ โดยใช้อัลกอริทึมการเรียนรู้ที่รวมไปถึงการเพิ่มประสิทธิภาพของไฮเปอร์พารามิเตอร์สำหรับแรนดอมฟอเรสต์ อดาบูสท์ และการตรวจสอบตัวชี้วัด



รูปที่ 3 แผนผังเสนอการปรับปรุงการทำงานของอัลกอริทึมแรนดอมฟอเรสต์ชนิดถดถอย

2.2.5 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

ซัพพอร์ตเวกเตอร์แมชชีน เป็นวิธีการเรียนรู้ที่ได้รับการพัฒนาจากทฤษฎีการเรียนรู้ทางสถิติ ที่มีประสิทธิภาพสูงกว่าโมเดลอื่น โดยรากฐานของโมเดลนี้ได้มีการกำหนดกรอบข้อมูลไว้ เพื่อแก้ปัญหาการเรียนรู้ของกลุ่มตัวอย่างที่ถูกจำกัด แนวคิดของซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ในการทำนายการถดถอยได้อธิบายไว้ดังนี้

สำหรับชุดตัวอย่าง $P(x, y), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in R^n \times R$ สมมติว่าฟังก์ชันการถดถอยคือ $F = \{f|x = \omega^T \cdot x + b, \omega \in R^n\}$ โดยสามารถเสนอฟังก์ชันความเสี่ยงได้ดังนี้

$$R_{reg} = \frac{1}{2} \|\omega\|^2 + C \cdot R_{emp}[f] \quad (10)$$

โดย $\|\omega\|^2$ เป็นฟังก์ชันอธิบาย, $f(\cdot)$ เป็นส่วนของฟังก์ชันที่ซับซ้อน, C เป็นค่าคงที่ ที่กำหนดระหว่างค่าความเสี่ยงและความซับซ้อนของแบบจำลอง (Jie)

ในการสร้างไฮเปอร์เพลนที่เหมาะสมที่สุดในกรณีที่มีข้อมูลแยกกันไม่ออกเชิงเส้น แนวคิดหลักของซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยแบบไม่เชิงเส้น คือการนำเวกเตอร์อินพุต x ไปจับคู่กับ สเปซที่คุณลักษณะมิติสูง โดยใช้กระบวนการแบบไม่เชิงเส้น ในพื้นที่ที่สูงขึ้นมีความเป็นไปได้มากขึ้นที่ข้อมูลจะสามารถแยกออกจากกันได้อย่างเป็นเส้นตรง (Mashud Rana 2016) โดยสามารถอธิบายปัญหาดังกล่าวได้ดังนี้

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \quad (11)$$

ขึ้นอยู่กับ $y_i(\omega \cdot \phi(x_i) + b) \geq 1 - \xi_i, \xi \geq 0, i = 1, \dots, C > 0$

ข้อมูลภายใน $\phi(x_i)$ ในพื้นที่ที่มีมิติสูงสามารถแทนที่ได้ด้วยฟังก์ชันเคอร์เนล $K(x, x_j)$ การคำนวณที่จำเป็นสามารถดำเนินการได้โดยตรงในการป้อนข้อมูลอินพุต

โดยฟังก์ชันเคอร์เนลแบบฐานเรเดียลสามารถแสดงได้ดังนี้

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (12)$$

ฟังก์ชันเคอร์เนลแบบแกมมาพหุนามสามารถแสดงได้ดังนี้

$$K(x, x_i) = (1 + x \cdot x_i)^d \quad (13)$$

โดยที่ d คือดีกรีของเคอร์เนลพหุนาม, γ คือค่าคงที่ที่กำหนดความกว้างของเคอร์เนลแบบฐานเรเดียล สามารถสร้างเครื่องมือเรียนรู้ที่แตกต่างกันได้โดยใช้ฟังก์ชันของเคอร์เนล $K(x_i, x_j)$ ในการใช้งานจริงฟังก์ชันเคอร์เนลมีอิทธิพลต่อผลลัพธ์ที่เกิดขึ้น สิ่งสำคัญคือการเลือกฟังก์ชันเคอร์เนลที่เหมาะสม เพื่อเพิ่มประสิทธิภาพของฟังก์ชัน โดยส่วนมากฟังก์ชันเคอร์เนลที่นิยมใช้คือ ฟังก์ชันเคอร์เนลแบบฐานเรเดียล ฟังก์ชันเคอร์เนลแบบ พหุนาม และฟังก์ชันซิกมอยด์

2.2.6 อาริมา (Auto Regressive Integrated Moving Average)

ในแบบจำลอง ARIMA ค่าเอาต์พุตที่คาดการณ์จะประกอบไปด้วยค่าเอาต์พุตย้อนหลังและข้อผิดพลาดที่เหลืออยู่ โดยอนุกรมของเวลาต้องเป็นไปตามข้อกำหนดพื้นฐานของแบบจำลอง ARIMA คือจะต้องอยู่ร่วมกับที่ และลักษณะของอนุกรมลำดับเวลาสำหรับโมเดล ARIMA นั้นแบ่งเป็นตามฤดูกาลและนอกฤดูกาล (S. Atique) ซึ่งในบทความนี้จะนำชนิดตามฤดูกาลมาใช้เพียงชนิดเดียวโดยสามารถอธิบายได้ดังข้างล่างนี้

ฤดูกาลของอนุกรมลำดับเวลาซึ่งแสดงโดย S นั้นมีลักษณะเฉพาะด้วยรูปแบบในอนุกรมเวลา ซึ่งเป็นคาบเวลา และ S มีค่าเท่ากับคาบของเวลา แบบจำลอง SARIMA นั้นเป็นสมการทางคณิตศาสตร์ที่ใช้แสดงแทนผลที่ได้จากทั้งแบบตามฤดูกาลและนอกฤดูกาล แสดงได้ดังสมการข้างล่างนี้

$$ARIMA(p, d, q) * (P, D, Q)_s$$

โดยที่ p, d และ q แสดงถึงลำดับของการถดถอยอัตโนมัติที่ไม่ใช่ฤดูกาล ค่าความแตกต่าง และค่าเฉลี่ยการเคลื่อนที่ ตามลำดับ P, D, Q แสดงถึงลำดับของการถดถอยอัตโนมัติตามฤดูกาล ค่าความแตกต่าง และค่าเฉลี่ยการเคลื่อนที่ ตามลำดับ สำหรับค่าที่มีความละเอียดมากยิ่งขึ้นของ SARIMA สามารถแสดงแทนได้ดังสมการต่อไปนี้

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS})(1 - B)^d(1 - B^S)^D y_t = (1 + \theta_1 B + \dots + \theta_q B^q)(1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS}) \epsilon_t \quad (14)$$

โดยที่ ϕ และ Φ เป็นพารามิเตอร์สำหรับ Auto Regression, θ และ Θ เป็นพารามิเตอร์สำหรับค่าเฉลี่ยการเคลื่อนที่ และ B แทนตัวดำเนินการ Backshift โดยการทำงานของ B แสดงในสมการนี้

$$B^m y_t = y_{t-m} \quad (15)$$

หากได้รับชุดข้อมูลที่เป็นอนุกรมลำดับเวลาไม่คงที่จำเป็นต้องแปลงให้เป็นอนุกรมลำดับเวลาคงที่ โดยในส่วนนี้ใช้เพียงค่าความแตกต่างในการแปลง หลังจากแปลงชุดข้อมูลแล้วจึงจะสามารถประมาณค่าของ p, P, q, Q, d, D และ S ได้ โดยค่าสุดท้ายจะถูกวัดหลังจากทดลองด้วยโมเดลต่างๆ ในท้ายที่สุดจึงจะได้โมเดลที่เหมาะสมและตรวจสอบความถูกต้องด้วยเกณฑ์ของข้อมูล Akaike (AIC) โดยโมเดลที่ดีที่สุดจะต้องให้ค่า AIC ที่ต่ำที่สุด โดยสมการของ AIC สามารถแสดงได้ดังนี้

$$AIC = -2 \log(\text{maximum likelihood}) + 2k \quad (16)$$

โดย k คือจำนวนพารามิเตอร์ที่ปรับอย่างอิสระ

2.3 การประเมินความแม่นยำของข้อมูลการทำนายด้วยวิธีการทางสถิติ

ในขั้นตอนนี้ได้นำเสนอตัวชี้วัดสำหรับการประเมินแบบจำลองการคาดการณ์ โดยในงานวิจัยฉบับนี้ได้นำตัวชี้วัด 4 ชนิดมาใช้ในการประเมินความถูกต้องของข้อมูลที่ได้จากการทำนายด้วยโมเดลการเรียนรู้ของเครื่อง

โดยชนิดแรกคือค่าความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Mean Square Error, MSE) เป็นฟังก์ชันที่ดูง่ายที่สุดในการคำนวณ โดยนำค่าความแตกต่างระหว่างการคาดการณ์ด้วยแบบจำลองกับข้อมูลจริง แล้วนำมายกกำลังสองและใช้ค่าเฉลี่ยกับชุดข้อมูลทั้งหมด โดยค่าของ MSE มีค่าเป็นบวกเสมอ โดยสามารถคำนวณได้จากสมการ

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (17)$$

ค่าของรากที่สองของค่าความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) จะเป็นการคำนวณความเหมาะสมของข้อมูลเพื่อป้องกันการเกิดค่าความผิดพลาดที่สูงมาก โดยสมการของค่า RMSE เป็นดังนี้

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (18)$$

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error, MAE) เป็นตัวบ่งชี้ประสิทธิภาพของแบบจำลองการทำนาย ซึ่งสังเกตได้จากตัวแปรที่คาดการณ์ไว้ใกล้เคียงกับที่สังเกตได้มากน้อยเพียงใด โดยสามารถหาได้จากสมการ

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (19)$$

ค่าของ R-Squared หรือค่าสัมประสิทธิ์ของการวัดค่าทางสถิติว่าข้อมูลมีความใกล้เคียงกับเส้นถดถอยเพียงใด

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (20)$$

ข้อมูลตัวชี้วัดเหล่านี้เป็นเครื่องมือที่ใช้เปรียบเทียบประสิทธิภาพของโมเดล ยิ่งค่าเข้าใกล้หนึ่งแสดงว่าโมเดลนั้นเหมาะสมที่จะนำมาใช้ทำนายข้อมูลเหล่านี้

บทที่ 3

การออกแบบและพัฒนาโมเดลการเรียนรู้ของเครื่องสำหรับการทำนายพลังงานไฟฟ้าที่ผลิตได้จาก เซลล์แสงอาทิตย์

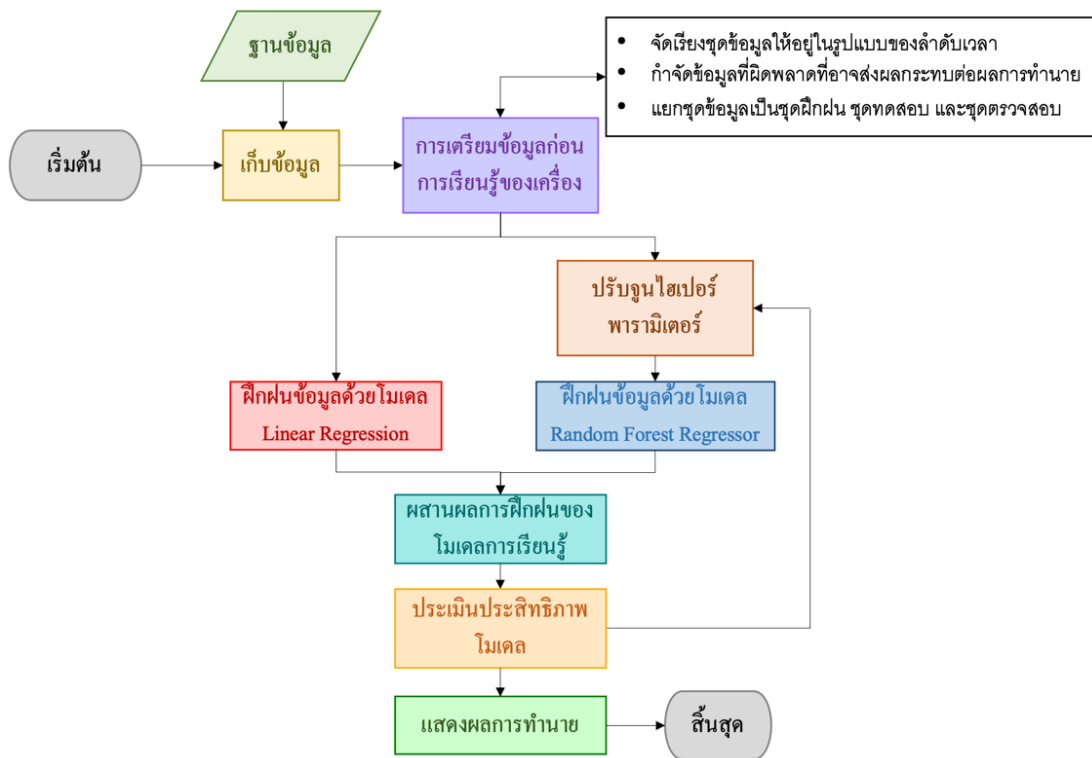
ในงานวิทยานิพนธ์ฉบับนี้ผู้วิจัยต้องการออกแบบการทดลองการทำนายค่าพลังงานไฟฟ้าที่ผลิตจากเซลล์แสงอาทิตย์โดยจะนำข้อมูลจากโรงไฟฟ้าพลังงานแสงอาทิตย์ของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทยที่ตั้งอยู่ อ.ทับสะแก จ.ประจวบคีรีขันธ์ ดังแสดงในรูปที่ 4 เป็นสถานที่ทดสอบโดยจะนำข้อมูลที่บันทึกไว้ในระบบของทีศูนย์ควบคุมโรงไฟฟ้าและนำผลสภาพอากาศย้อนหลังผ่านเว็บไซต์ Solcast และ OpenWeatherMap มาใช้ในการทดลอง



รูปที่ 4 ศูนย์ควบคุมโรงไฟฟ้าพลังงานแสงอาทิตย์

3.1 ภาพรวมขั้นตอนการทำงานของงานวิจัย

ในการแสดงภาพรวมการดำเนินการวิจัยสามารถแสดงรูปแบบการดำเนินการวิจัยได้ดังรูปที่ 5 โดยเริ่มจากการเก็บข้อมูลการผลิตย้อนหลังและบันทึกข้อมูลสภาพอากาศย้อนหลัง จากนั้นนำข้อมูลไปตรวจสอบความถูกต้องและทำนอร์มัลไลเซชันข้อมูลผ่านโปรแกรม Microsoft Excel และนำข้อมูลเข้าสู่โปรแกรม Jupyter เพื่อแบ่งชุดการทดสอบสำหรับการเรียนรู้ผ่านโมเดลการเรียนรู้ เมื่อกระบวนการเรียนรู้เสร็จสิ้น จึงประเมินผลการทำนายของแต่ละโมเดลพร้อมทั้งเปรียบเทียบและวิเคราะห์ผลการทำนาย และสรุปผลการทดลองต่อไป



รูปที่ 5 ภาพรวมแสดงการดำเนินการวิจัย

3.2 รายงานวิธีการดำเนินการวิจัย

ขั้นตอนที่ 1 เริ่มเก็บข้อมูลค่าพลังงานไฟฟ้าที่ผลิตจากเซลล์แสงอาทิตย์ และข้อมูลสภาพอากาศที่ส่งผลต่อกำลังการผลิตไฟฟ้าจากเซลล์พลังงานแสงอาทิตย์ โดยจะเก็บข้อมูลต่างๆ ที่โรงไฟฟ้าพลังงานแสงอาทิตย์ 5 เมกะวัตต์ ที่ อ. ทับสะแก จ.ประจวบคีรีขันธ์ โดยข้อมูลจะถูกบันทึกค่าเข้าสู่ระบบในทุกๆ 1 ชั่วโมง โดยจะเป็นข้อมูล ณ จุดเวลาในแต่ละชั่วโมง โดยโรงไฟฟ้าแห่งนี้แบ่งโซนควบคุมของเซลล์แสงอาทิตย์ออกเป็น 5 โซน ในแต่ละโซนจะมีการใช้งานเซลล์ที่ผลิตจากวัสดุต่างชนิดกัน ในงานวิจัยนี้จัดเก็บข้อมูลจากโซน C เนื่องจากเป็นพื้นที่ที่มีความผิดพลาดของข้อมูลน้อยที่สุด และมีความต่อเนื่องมากที่สุด ในโซนนี้ได้ทำการติดตั้งแผงโซลาร์เซลล์ชนิด micro crystalline silicon thin film (μ -Si) โดยมีภาพเซลล์แสงอาทิตย์ที่ติดตั้งอยู่ในโรงไฟฟ้าแสดงอยู่ในรูปที่ 6 ในส่วนของข้อมูลสภาพอากาศจะเรียกข้อมูลย้อนหลังจากฐานข้อมูลขององค์กรเอกชน (เว็บไซต์ SolCast และ OpenWeatherMap) เพื่อนำไปใช้เป็นอินพุตในการทำนายการผลิต เมื่อนำข้อมูลดิบมาจะต้องนำข้อมูลไปผ่านกระบวนการก่อนการเรียนรู้ (Pre-processing method) เพื่อลดความผิดพลาดใน

การทำนายที่อาจจะเกิดขึ้น ขั้นตอนย่อยต่อไปนี้จะแสดงโปรแกรมสำหรับกระบวนการก่อนการเรียนรู้ พร้อมทั้งคำอธิบายในแต่ละขั้นตอน



รูปที่ 6 ภาพแผงโซลาร์เซลล์บริเวณโซน C

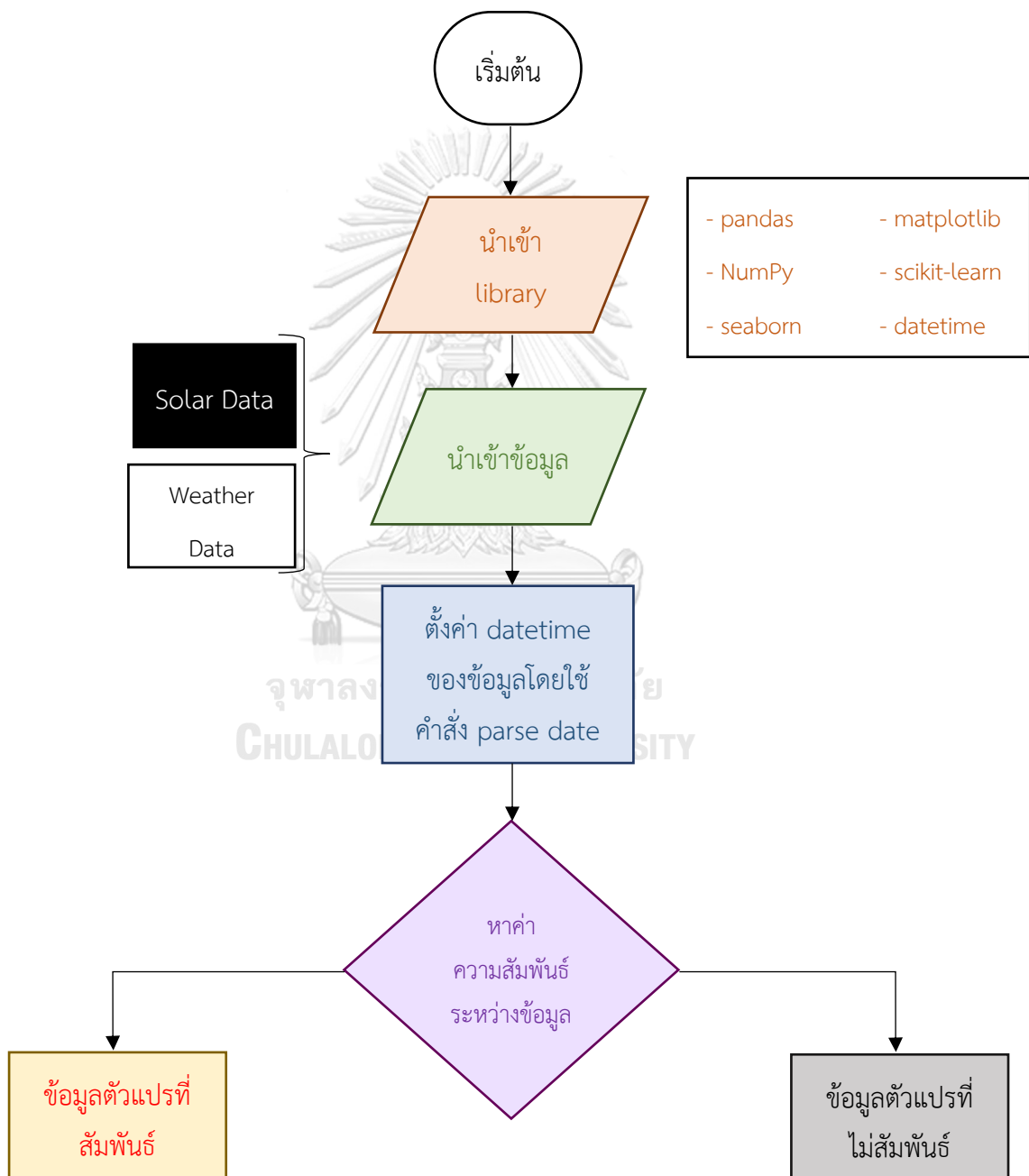
ขั้นตอนที่ 1.1 การจัดเก็บและตรวจสอบข้อมูล ขั้นตอนนี้ผู้วิจัยได้จัดทำตารางสำหรับการตรวจสอบความถูกต้องของข้อมูลจากโรงไฟฟ้าพลังงานแสงอาทิตย์ว่ามีข้อมูลรายชั่วโมงจุดใดที่ขาดหายไป และตรวจสอบว่ามีการบันทึกชั่วโมงใดบ้างที่เกิดความผิดพลาดในการบันทึก เมื่อทำการตรวจสอบเรียบร้อยแล้วจึงจะนำข้อมูลที่ผ่านการตรวจสอบและแก้ไขแล้วไปใช้ต่อไปในการทดลอง ผู้วิจัยได้จัดทำตารางการตรวจสอบใส่ไว้อยู่ที่ท้ายวิทยานิพนธ์ฉบับนี้ตามตารางที่ 4 ถึงตารางที่ 7

ขั้นตอนที่ 1.2 นำข้อมูลที่ผ่านการตรวจสอบแล้วเข้าสู่กระบวนการนอร์มัลไลเซชัน ผ่านสมการที่ (4) โดยในการทดลองนี้ได้ทำการนอร์มัลไลเซชันผ่านโปรแกรม Excel เนื่องจากโปรแกรมนี้สามารถเลือกข้อมูลที่ต้องการทำนอร์มัลไลเซชัน เนื่องจากตัวแปรที่ต้องการทำนายคือค่าพลังงานไฟฟ้าที่ผลิตได้ในแต่ละชั่วโมง (กำหนดให้เป็นตัวแปร y) ในขั้นตอนนี้จึงต้องการทำนอร์มัลไลเซชันเฉพาะข้อมูลตัวแปร x จากนั้นจึงนำข้อมูลที่ได้ไปแบ่งชุดข้อมูลเพื่อเตรียมพร้อมเข้าสู่ขั้นตอนการเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่อง โดยในรูปที่ 7 แสดงตัวอย่างชุดข้อมูลนอร์มัลไลซ์

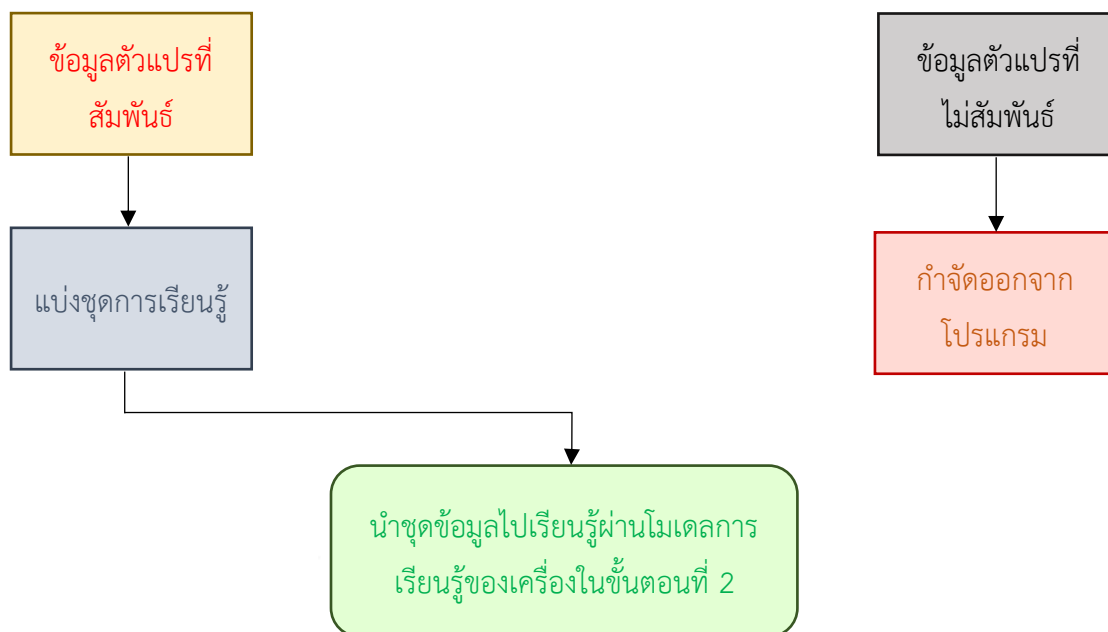
| Datetime | Activepower | TodayGenCap | CellTemp1 | CellTemp2 | Irradiance | AirTemp | Azimuth | CloudOpacity | DewpointTemp | Dhi | ... | wind_speed | wind_deg | rain_1h |
|---------------------|-------------|-------------|-----------|-----------|------------|---------|---------|--------------|--------------|--------|-----|------------|----------|---------|
| 2020-01-01 07:00:00 | 14.47 | 0.0002 | 0.2769 | 0.1941 | 0.0116 | 0.3406 | 0.1778 | 0.5062 | 0.7308 | 0.0669 | ... | 0.0002 | 0.2389 | 0.0157 |
| 2020-01-01 08:00:00 | 119.46 | 0.0065 | 0.3257 | 0.2532 | 0.0835 | 0.4348 | 0.1611 | 0.5082 | 0.7538 | 0.2588 | ... | 0.0003 | 0.2444 | 0.0214 |
| 2020-01-01 09:00:00 | 249.69 | 0.0228 | 0.3746 | 0.3291 | 0.1852 | 0.5145 | 0.1417 | 0.1165 | 0.7692 | 0.4085 | ... | 0.0004 | 0.1917 | 0.0157 |
| 2020-01-01 10:00:00 | 717.80 | 0.0994 | 0.4691 | 0.6076 | 0.1503 | 0.5000 | 0.1083 | 0.0093 | 0.7923 | 0.4577 | ... | 0.0005 | 0.1750 | 0.0107 |
| 2020-01-01 11:00:00 | 955.17 | 0.2119 | 0.4919 | 0.5359 | 0.2774 | 0.5580 | 0.0583 | 0.0247 | 0.7846 | 0.5387 | ... | 0.0004 | 0.1750 | 0.0000 |

รูปที่ 7 ตัวอย่างชุดข้อมูลที่ผ่านการนอร์มัลไลเซชัน

หลังจากกระบวนการก่อนการเรียนรู้ข้อมูลเสร็จสิ้น จึงนำเข้าสู่โปรแกรม Jupyter Notebook เพื่อเตรียมความพร้อมที่จะนำข้อมูลไปเรียนรู้ จากนั้นตั้งค่า *index* ของข้อมูลแต่ละชั่วโมงให้เป็นวันที่และเวลาเพื่อให้ง่ายต่อการแสดงผล สำหรับการแบ่งชุดข้อมูลในงานวิจัยนี้ได้ทำการแบ่งชุดข้อมูลออกเป็น 3 ชุด ได้แก่ ชุดเรียนรู้ ชุดตรวจสอบ และชุดทดสอบ โดยแบ่งสัดส่วนของชุดข้อมูลออกเป็นร้อยละ 60 20 และ 20 ตามลำดับ ซึ่งขั้นตอนการทดลองที่ 1 สามารถแสดงได้ดังแผนผัง flowchart ดังต่อไปนี้



รูปที่ 8 แผนภาพ flowchart แสดงการดำเนินการขั้นตอนที่ 1



รูปที่ 9 แผนภาพ flowchart แสดงการดำเนินการขั้นตอนที่ 1 (ต่อ)

จากแผนภาพ flowchart ได้แสดงถึงขั้นตอนการดำเนินการในขั้นตอนที่ 1 โดยในขั้นตอนนี้ได้มีการนำเข้าข้อมูลและไลบรารีของชุดคำสั่งที่ใช้ในการทดลอง การวิเคราะห์ความสัมพันธ์ระหว่างข้อมูล และการแบ่งชุดข้อมูล ในส่วนของการแบ่งชุดข้อมูลสำหรับการเรียนรู้ การตรวจสอบ และการทดสอบนั้น จากการทดลองวิจัยพบว่าทางเลือกเดือนจากแต่ละฤดูกาลจะส่งผลต่อการเรียนรู้และการทำนายมากที่สุด แต่ด้วยเนื่องจากชุดข้อมูลที่ยังมีอยู่อย่างจำกัด ผู้วิจัยจึงเลือกฤดูร้อนและฤดูฝนมาใช้ในการตรวจสอบและทดสอบ เนื่องจากสภาพอากาศประเทศไทยมีสองฤดูนี้เป็นหลัก ซึ่งเราจะเห็นได้ว่าในช่วงปีหลังมานี้แม้ในช่วงฤดูหนาวจะมีอากาศหนาวในช่วงสัปดาห์แต่เป็นเพียงช่วงระยะสั้น อีกทั้งเมื่อเปรียบเทียบกับประเทศอื่นพบว่าสภาพอากาศในประเทศไทยก็ยังสามารถจัดรวมอยู่ในฤดูร้อนได้ ผู้วิจัยจึงได้จัดทำชุดข้อมูลออกเป็น 3 ชุดโดยในแต่ละชุดข้อมูลจะมีการแบ่งเพียงสองฤดู คือ ฤดูร้อนและฤดูฝน โดยมีอัตราส่วนของข้อมูลร้อยละ 60 20 และ 20 ตามลำดับ ซึ่งโค้ดโปรแกรมสำหรับการทดลองขั้นตอนที่ 1 สามารถแสดงได้ดังรูปต่อไปนี้

```

In [1]: # import libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from datetime import datetime
import matplotlib.dates as mdates
from sklearn import preprocessing

In [2]: # Load .csv data includes solar data in 2020 & 2021 and weather data from 2020 to Jun 2023
custom_data_parser = lambda x: datetime.strptime(x, '%d-%m-%Y %H:%M:%S')
solar_data = pd.read_csv("SolarDataNormalized.csv", low_memory=False, index_col = 'Datetime',
                        parse_dates=['Datetime'], date_parser = custom_data_parser)
weather_dataSol = pd.read_csv("WeatherDataAllNormalized.csv", low_memory=False, index_col = 'Datetime',
                              parse_dates=['Datetime'], date_parser = custom_data_parser)
weather_dataOpen = pd.read_csv("OpenweathermapNormalized.csv", low_memory=False, index_col = 'Datetime',
                                parse_dates=['Datetime'], date_parser = custom_data_parser)

In [3]: # Check dataset shape
solar_data.shape, weather_dataSol.shape, weather_dataOpen.shape

Out[3]: ((14968, 7), (32376, 16), (31552, 14))

```

รูปที่ 10 การนำเข้าชุดคำสั่งและชุดข้อมูล

จากรูปที่ 10 ได้แสดงการนำเข้าคำสั่ง pandas, NumPy, seaborn, matplotlib, datetime, scikit-learn และนำเข้าชุดข้อมูลจากโรงไฟฟ้าและสภาพอากาศที่ผ่านการนอร์มัลไลเซชันมาแล้ว ในการนำเข้าได้มีการตั้งค่า index ของข้อมูลให้เป็นค่าวัน เดือน ปี และเวลา เนื่องจากข้อมูลที่ใช้อ้างอิงตามลำดับเวลา การทดลองจึงต้องการให้ความสำคัญกับลำดับเวลาของข้อมูล เพื่อง่ายต่อการนำข้อมูลไปใช้ทดลองและวิเคราะห์ผลการทดลอง หลังจากนำเข้าข้อมูลเข้าสู่โปรแกรมการทดลองได้มีการรวมข้อมูลทั้งหมดเป็นชุดข้อมูลเดี่ยวพร้อมทั้งสร้างคอลัมน์สำหรับวัน เดือน ปี แสดงโค้ดโปรแกรมได้ดังรูปที่ 11 และ 12 เพื่อนำข้อมูลเดือนไปใช้ในการแบ่งฤดูกาลของชุดข้อมูลในรูปที่ 13

จุฬาลงกรณ์มหาวิทยาลัย

2. Pre-processing Data Section

```

In [4]: # Change types of Date and Time using pd.to_datetime() functions
solar_data["Date"] = pd.to_datetime(solar_data["Date"] + ' ' + solar_data["Time"])
solar_data=solar_data.drop(["Time"], axis=1)

In [5]: weather_dataSol["Date"] = pd.to_datetime(weather_dataSol["Date"] + ' ' + weather_dataSol["Time"])
weather_dataSol=weather_dataSol.drop(["Time"], axis=1)
weather_dataOpen["Date"] = pd.to_datetime(weather_dataOpen["Date"] + ' ' + weather_dataOpen["Time"])
weather_dataOpen=weather_dataOpen.drop(["Time"], axis=1)

In [6]: weather_data = pd.merge(weather_dataSol, weather_dataOpen, on=['Datetime', 'Date'])

In [9]: # Merging datasets of solar informations and weather conditions
merged_data = pd.merge(solar_data, weather_data, on=['Datetime', 'Date']) # merge data from solar plant and meteorological data
merged_data.head(3)

```

รูปที่ 11 โค้ดโปรแกรมการรวมชุดข้อมูล


```
In [10]: merged_data['Day']=merged_data["Date"].dt.day
merged_data['Month']=merged_data["Date"].dt.month
merged_data['Year']=merged_data["Date"].dt.year
merged_data['Hour']=merged_data["Date"].dt.hour
merged_data=merged_data.drop(["Date"],axis=1)
merged_data.head(5)
```

รูปที่ 12 โค้ดโปรแกรมการรวมชุดข้อมูล

3. Splitting Data For Training, Validating, and Testing

```
In [15]: # Create train and test datasets
df_train = pd.concat([merged_data[merged_data.Month == 1], merged_data[merged_data.Month == 2],
                    merged_data[merged_data.Month == 3], merged_data[merged_data.Month == 6],
                    merged_data[merged_data.Month == 7], merged_data[merged_data.Month == 10],
                    merged_data[merged_data.Month == 11], merged_data[merged_data.Month == 12]])
df_val = pd.concat([merged_data[merged_data.Month == 4], merged_data[merged_data.Month == 8]])
df_test = pd.concat([merged_data[merged_data.Month == 5], merged_data[merged_data.Month == 9]])
len(df_train), len(df_val), len(df_test)
```

Out[15]: (9896, 2278, 2423)

รูปที่ 13 โค้ดโปรแกรมสำหรับการแบ่งชุดข้อมูล

หลังจากแบ่งชุดข้อมูลแล้วจึงนำข้อมูลไปวิเคราะห์ความสัมพันธ์ผ่านวิธีการทางสถิติและวิธีการของเพียร์สัน ในการทดสอบหาความสัมพันธ์ทางสถิติได้มีการใช้คำสั่ง `Dataframe.corr()` และแสดงผลผ่านแผนภาพ heatmap และการทดสอบหาความสัมพันธ์ของเพียร์สัน ได้ใช้คำสั่ง `pearsonr` โดยนำเข้าจากไลบรารีของ Scipy จากนั้นนำข้อมูลการวิเคราะห์ความสัมพันธ์มาใช้ในการพิจารณาตัวแปรที่ส่งผลต่อการทำนายจริง จากนั้นจึงได้กำจัดข้อมูลตัวแปรที่ไม่เกี่ยวข้องโดยรูปที่ 14 แสดงโค้ดโปรแกรมการกำจัดตัวแปรที่ไม่ส่งผลต่อการทำนาย

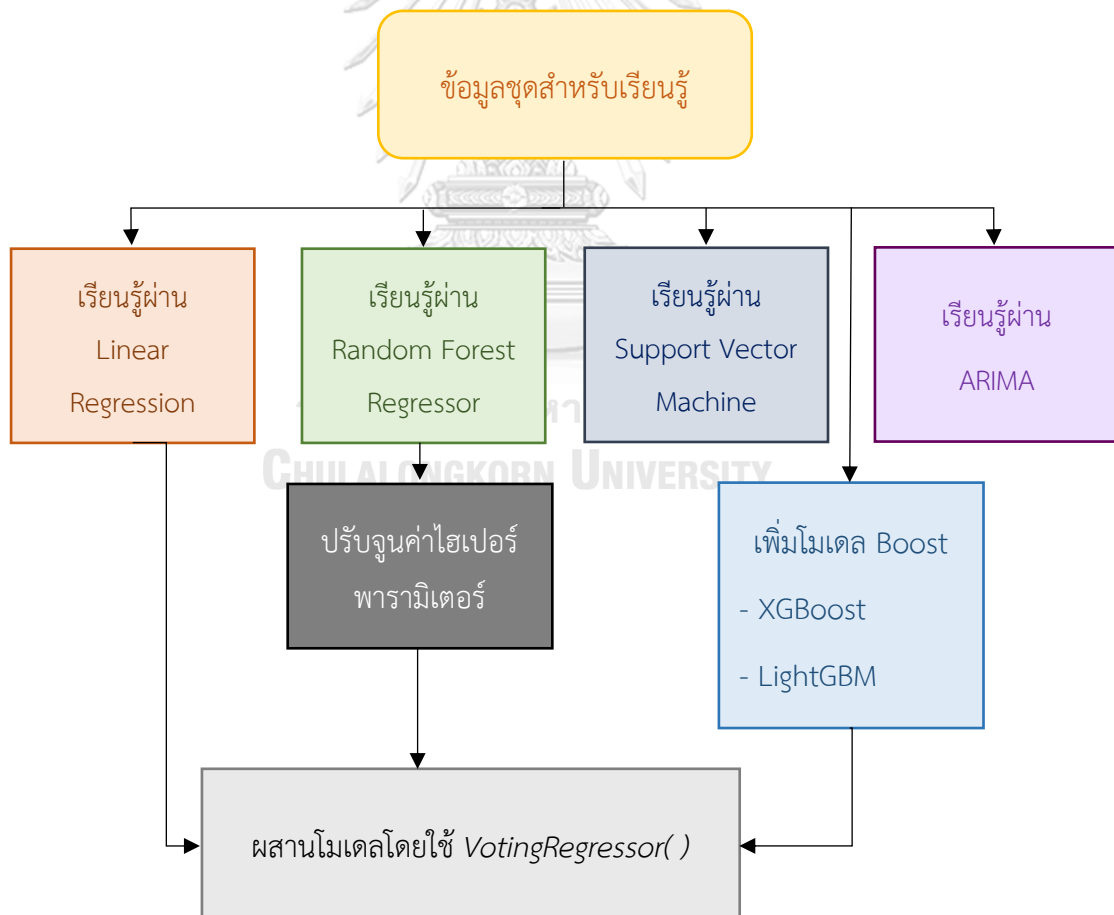
Clean variables that didn't correlated with Target variable (Power)

```
In [16]: X_train, y_train = df_train.drop(['Activepower', 'TodayGenCap', 'AirTemp', 'Azimuth',
                                       'DewpointTemp', 'Dhi', 'Dni', 'Ebh', 'Ghi', 'PrecipitableWater',
                                       'WindDirection10m', 'WindSpeed10m', 'feels_like', 'dew_point',
                                       'humidity', 'wind_speed', 'wind_deg', 'rain_1h', 'clouds_all', 'weather_main',
                                       'weather_description', 'Day', 'Month', 'Year', 'Hour'], axis=1), df_train.Activepower
X_val, y_val = df_val.drop(['Activepower', 'TodayGenCap', 'AirTemp', 'Azimuth',
                           'DewpointTemp', 'Dhi', 'Dni', 'Ebh', 'Ghi', 'PrecipitableWater',
                           'WindDirection10m', 'WindSpeed10m', 'feels_like', 'dew_point',
                           'humidity', 'wind_speed', 'wind_deg', 'rain_1h', 'clouds_all', 'weather_main',
                           'weather_description', 'Day', 'Month', 'Year', 'Hour'], axis=1), df_val.Activepower
X_test, y_test = df_test.drop(['Activepower', 'TodayGenCap', 'AirTemp', 'Azimuth',
                               'DewpointTemp', 'Dhi', 'Dni', 'Ebh', 'Ghi', 'PrecipitableWater',
                               'WindDirection10m', 'WindSpeed10m', 'feels_like', 'dew_point',
                               'humidity', 'wind_speed', 'wind_deg', 'rain_1h', 'clouds_all', 'weather_main',
                               'weather_description', 'Day', 'Month', 'Year', 'Hour'], axis=1), df_test.Activepower
X_train.shape, X_val.shape, X_test.shape, y_train.shape, y_val.shape, y_test.shape
```

Out[16]: ((9896, 10), (2278, 10), (2423, 10), (9896,), (2278,), (2423,))

รูปที่ 14 โค้ดโปรแกรมแสดงการกำจัดตัวแปร

ขั้นตอนที่ 2 เป็นขั้นตอนในการเตรียมโมเดลการเรียนรู้ของเครื่อง โดยนำโมเดลการวิเคราะห์การถดถอยเชิงเส้น และ โมเดลแรนดอมฟอเรสต์สำหรับการถดถอย มาประยุกต์ใช้ร่วมกัน เพื่อปรับปรุงประสิทธิภาพในการทำนายให้ดีขึ้น โดยในแต่ละโมเดลนั้นมีทั้งข้อเด่นและข้อด้อยที่ต่างกัน ดังนั้นการทำงานร่วมกันจึงช่วยดึงจุดเด่นของแต่ละโมเดลมาใช้ทำนายได้อย่างมีประสิทธิภาพ เหตุผลที่ได้เสนอการวิเคราะห์การถดถอยเชิงเส้นมาปรับปรุงการทำงานโมเดลแรนดอมฟอเรสต์นั้นคือ โมเดลการวิเคราะห์การถดถอยเชิงเส้นช่วยประเมินความสำคัญของตัวแปรอินพุต และ ปรับค่าน้ำหนักถ่วงของแต่ละตัวแปรให้เหมาะสม อีกทั้งยังช่วยให้การปรับพารามิเตอร์ในโมเดลแรนดอมฟอเรสต์ง่ายขึ้นด้วย รวมไปถึงจะช่วยจัดลำดับความสำคัญของตัวแปรที่ส่งผลต่อประสิทธิภาพการทำนายให้เป็นระเบียบและง่ายต่อการวิเคราะห์ถึงปัญหาระหว่างการทำวิจัย ในขั้นตอนนี้เป็นขั้นตอนที่สำคัญที่สุด เนื่องจากการเตรียมโมเดลการเรียนรู้ของเครื่องให้พร้อมสำหรับการเรียนรู้และทำนายผลการทดลอง ในแผนภาพ flowchart ต่อไปนี้จะแสดงการดำเนินการสร้างโมเดลการเรียนรู้ของเครื่องสำหรับการทำนายค่าพลังงานไฟฟ้าที่ผลิตได้จากเซลล์แสงอาทิตย์



รูปที่ 15 แผนภาพ flowchart สำหรับการดำเนินการขั้นตอนที่ 2

จากแผนภาพรูปที่ 15 แสดงวิธีการดำเนินการวิจัยในขั้นตอนที่ 2 โดยสามารถอธิบายวิธีการในแต่ละขั้นตอนได้ดังต่อไปนี้

ขั้นตอนที่ 2.1 นำชุดข้อมูลสำหรับการเรียนรู้มาเรียนรู้ผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้น โดยใช้ชุดคำสั่งของโมเดล *LinearRegression* ซึ่งเป็นชุดการเรียนรู้ที่จัดอยู่ในไลบรารีของ Scikit-Learn ประเภท *linear_model* สามารถแสดงโค้ดโปรแกรมการเรียนรู้ข้อมูลผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้นได้ดังแสดงในรูปที่ 16

Linear Regression

```
In [18]: %%time
# Let's fit by linear regression model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

CPU times: total: 15.6 ms
Wall time: 205 ms

```
Out[18]:
LinearRegression()
LinearRegression()
```

รูปที่ 16 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Linear Regression

ขั้นตอนที่ 2.2 นำชุดข้อมูลสำหรับการเรียนรู้มาเรียนรู้ผ่านโมเดลแรมดอมฟอเรสต์ชนิดถดถอย โดยใช้ชุดคำสั่ง *RandomForestRegressor* ซึ่งจัดอยู่ในไลบรารีของ Scikit-Learn ประเภท *ensemble* สามารถแสดงโค้ดโปรแกรมการเรียนรู้ผ่านโมเดลแรมดอมฟอเรสต์ชนิดถดถอยได้ดังรูปที่ 17

CHULALONGKORN UNIVERSITY

Random Forest

```
In [21]: # import model machine Learning
from sklearn.ensemble import RandomForestRegressor
```

```
In [22]: %%time
# Try to train RandomForestRegressor model
RandomForest = RandomForestRegressor()
RandomForest.fit(X_train, y_train)
```

CPU times: total: 20.2 s
Wall time: 21.6 s

```
Out[22]:
RandomForestRegressor()
RandomForestRegressor()
```

รูปที่ 17 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Random Forest Regressor

ขั้นตอนที่ 2.3 เป็นการค้นหาชุดของค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดสำหรับโมเดลแรนดอมฟอเรสต์ ในขั้นตอนนี้เริ่มจากการสร้างฟังก์ชันของตัวแปรที่ใช้ในการปรับจูน (RF Optimizer) เพื่อปรับจูนค้นหาพารามิเตอร์ผ่านคำสั่ง `RandomizedSearchCV` โดยรูปที่ 18 แสดงโค้ดโปรแกรมการค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดสำหรับการทำนายในการทดลองวิจัยนี้

Hyperparameter tuning with RandomizedSearchCV

```
In [25]: %%time
from sklearn.model_selection import RandomizedSearchCV

## Different RandomForestRegressor hyperparameters
rf_grid = {"n_estimators": np.arange(10,1000,10),
           "max_depth": [None, 3, 5, 10],
           "min_samples_split": np.arange(2, 20, 2),
           "min_samples_leaf": np.arange(1, 40, 2),
           "max_features": [0.5, 1]}

# Instantiate RandomizedSearchCV model
rs_model = RandomizedSearchCV(RandomForestRegressor(n_jobs=-1,
                                                    random_state=42),
                              param_distributions=rf_grid,
                              n_iter=5,
                              cv=5,
                              verbose=True)

# Fit the RandomizedSearchCV model
rs_model.fit(X_train, y_train)
```

รูปที่ 18 โค้ดโปรแกรมแสดงการค้นหาค่าไฮเปอร์พารามิเตอร์

จุฬาลงกรณ์มหาวิทยาลัย

หลังจากที่ได้ค้นหาพารามิเตอร์ที่ดีที่สุดแล้วจึงนำค่านั้นไปทดลองปรับจูนการทำนายของโมเดลแรนดอมฟอเรสต์เพื่อนำโมเดลหลังการปรับจูนไปผสมการทำนายร่วมกับโมเดลการวิเคราะห์การถดถอยเชิงเส้นต่อไป

ขั้นตอนที่ 2.4 การนำโมเดลบูสท์มาใช้ปรับปรุงประสิทธิภาพเพื่อเป็นการพัฒนาประสิทธิภาพในการทำนาย จึงได้เสนอให้นำโมเดล XGBoost และ LightGBM เข้ามาช่วยปรับปรุงประสิทธิภาพให้กับโมเดลการทำนาย ในรูปที่ 19 แสดงโค้ดโปรแกรมการเรียนรู้ที่ได้มีการนำสองโมเดลนี้มาใช้เพื่อนำไปปรับปรุงและพัฒนาให้มีความแม่นยำมากยิ่งขึ้น

```
In [35]: import xgboost as xgb
import lightgbm as lgb

In [36]: # Create training models of XGB
xgb_model = xgb.XGBRegressor(random_state=42)
# Train the model
xgb_model.fit(X_train, y_train)

In [38]: # Create training models of LightGBM
lgb_model = lgb.LGBMRegressor(random_state=42)
# Train the model
lgb_model.fit(X_train, y_train)
```

รูปที่ 19 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดลบูสต์

หลังจากปรับจูนค่าไฮเปอร์พารามิเตอร์และนำไปใช้ในการเรียนรู้ผ่านโมเดลแรนดอมฟอเรสต์ เพื่อปรับปรุงประสิทธิภาพในการเรียนรู้ และ นำไปเรียนรู้ผ่านโมเดลบูสต์ทั้งสองชนิด แล้วนั้น จึงได้นำ โมเดลการวิเคราะห์การถดถอยเชิงเส้น โมเดลแรนดอมฟอเรสต์ที่ปรับจูนค่าไฮเปอร์พารามิเตอร์ โมเดล XGBoost และ โมเดล LightGBM ทั้งหมดมาผสมผสานการทำนายหรือที่เรียกว่า Ensemble Method โดยใช้วิธีการผสมผ่านชุดคำสั่ง `VotingRegressor ()` ดังแสดงในรูปที่ 20

Let try to ensemble our proposed models includes:

```
LinearRegression , RandomForestRegressor , XGBoost , LightGBM
```

Prediction using validation data

```
In [40]: from sklearn.ensemble import VotingRegressor
# Create a dictionary of our models
estimators = [('model', model), ('RandomForest', ideal_model), ('xgb_model', xgb_model), ('lgb_model', lgb_model)]
# Create our voting classifier, inputting our models
ensemble = VotingRegressor(estimators)
```

รูปที่ 20 โค้ดโปรแกรมสำหรับการผสมโมเดล (Ensemble)

ในการผสมผสานการทำนายของโมเดลการเรียนรู้ของเครื่องในการทดลองนี้ ได้ปรับค่าน้ำหนักของแต่ละโมเดลอย่างละเท่าๆกัน หรือเรียกว่าการเฉลี่ยค่าน้ำหนักของโมเดล เนื่องจากทางผู้วิจัยได้ทดลองปรับค่าน้ำหนักของแต่ละโมเดล แต่พบว่าประสิทธิภาพในการทำนายลดลงเนื่องจากเป็นการปรับแบบ *manual* ในการทดลองวิจัยจึงได้ใช้วิธีการผสมผสานการทำนายแบบมาตรฐานเพื่อให้ผลการทำนายออกมามีประสิทธิภาพมากที่สุด

ขั้นตอนที่ 2.5 นำชุดข้อมูลการทดลองไปเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่องชนิดอื่นเพื่อเปรียบเทียบและวิเคราะห์ผลการทำนาย โดยในงานวิจัยฉบับนี้ได้นำโมเดลการเรียนรู้ของเครื่องมาใช้ในการเปรียบเทียบ 2 ชนิด คือโมเดลซัพพอร์ตเวกเตอร์แมชชีน (SVM) และ โมเดลอาร์มา (ARIMA) ซึ่งสามารถแสดงได้ดังรูปที่ 21 และ 22

Next, try to train using others machine learning models such as

SupportVectorMachine , AutoRegressiveIntegratedMovingAverage

```
In [60]: from sklearn import svm
regr = svm.SVR()
regr.fit(X_train, y_train)
```

```
Out[60]: SVR()
```

รูปที่ 21 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดล Support Vector Machine

จากรูปที่ 21 ได้มีการแสดงโค้ดโปรแกรมสำหรับโมเดลซัพพอร์ตเวกเตอร์แมชชีน ซึ่งในการทดลองนี้ได้นำโมเดลซัพพอร์ตเวกเตอร์แมชชีนชนิดถดถอยมาใช้ในการเรียนรู้ โดยมีค่าพารามิเตอร์เป็นค่าตั้งต้น ซึ่งมีค่าดังต่อไปนี้ kernel = rbf, degree = 3, gamma = 'scale', coef0 = 0.0, tol = 0.001, C=1.0, epsilon = 0.1, shrinking = True, cache_size = 200, verbose = False, max_iter = -1

Test with ARIMA model

```
In [67]: from pmdarima.arima import auto_arima
```

```
In [69]: arima_model = auto_arima(ARIMA_train1.Activepower,
start_p=0, d=0, start_q=0,
max_p=2, max_d=0, max_q=2,
start_P=0, D=1, start_Q=0,
max_P=1, max_D=1, max_Q=1, m=96,
seasonal=True,
error_action='warn', trace=True,
supress_warning=True, stepwise=True,
random_state=20, n_fits=1)
```

```
In [70]: dates = pd.date_range(start=ARIMA_test1['Date'].index[0],
end='2023-08-01 23:00:00', freq=pd.DateOffset(minutes=60))

prediction_test1 = pd.DataFrame(arima_model.predict(n_periods=len(dates)), index=dates)
prediction_test1.columns = ['Activepower']
```

รูปที่ 22 โค้ดโปรแกรมการเรียนรู้ผ่านโมเดลอาร์มา

จากรูปที่ 22 แสดงโค้ดโปรแกรมสำหรับการเรียนรู้ด้วยโมเดลอาร์มา (ARIMA) ในงานวิจัยนี้ได้เลือกฟังก์ชันโมเดล `auto_arima` มาใช้ ในส่วนของชุดข้อมูลที่นำมาเรียนรู้ เนื่องจากโมเดลนี้เป็นโมเดลการทำนายตามลำดับเวลา และเนื่องจากเป็นโมเดลที่ใช้ทรัพยากรของคอมพิวเตอร์ค่อนข้างสูง จึงไม่สามารถนำข้อมูลย้อนหลังทั้งหมดมาใช้ในการทำนายได้ ทางผู้วิจัยจึงได้เลือกข้อมูลย้อนหลัง 7 วัน เพื่อทำนายผลล่วงหน้า 1 วันจากนั้นจึงเก็บผลการทำนายที่ได้ไปเปรียบเทียบกับโมเดลการเรียนรู้ของเครื่องประเภทอื่นในผลงานวิจัยฉบับนี้

ขั้นตอนที่ 3 สร้างฟังก์ชันตัวชี้วัดสำหรับการประเมินประสิทธิภาพของโมเดลการเรียนรู้ของเครื่อง โดยฟังก์ชันนี้จะช่วยทดสอบค่า *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)* และ *R-squared Value (R² Score)* ซึ่งสามารถแสดงโค้ดโปรแกรมการสร้างฟังก์ชันตัวชี้วัดได้ดังแสดงในรูปที่ 23

Building an evaluation function

```
In [126]: from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
def rmse(y_test, y_preds):
    """
    Calculate root mean squared log error between predictions and true labels
    """
    return np.sqrt(mean_squared_error(y_test, y_preds))
def show_scores(model):
    train_preds = model.predict(X_train)
    val_preds = model.predict(X_val)
    test_preds = model.predict(X_test)
    scores = {"Training MAE": mean_absolute_error(y_train, train_preds),
             "Valid MAE": mean_absolute_error(y_val, val_preds),
             "Test MAE": mean_absolute_error(y_test, test_preds),
             "Training RMSE": rmse(y_train, train_preds),
             "Valid RMSE": rmse(y_val, val_preds),
             "Test RMSE": rmse(y_test, test_preds),
             "Training R^2": r2_score(y_train, train_preds),
             "Valid R^2": r2_score(y_val, val_preds),
             "Test R^2": r2_score(y_test, test_preds)}
    return scores
```

รูปที่ 23 โค้ดโปรแกรมการสร้างฟังก์ชันตัวชี้วัดสำหรับประเมินประสิทธิภาพ

บทที่ 4

การทดสอบและผลการวิจัย

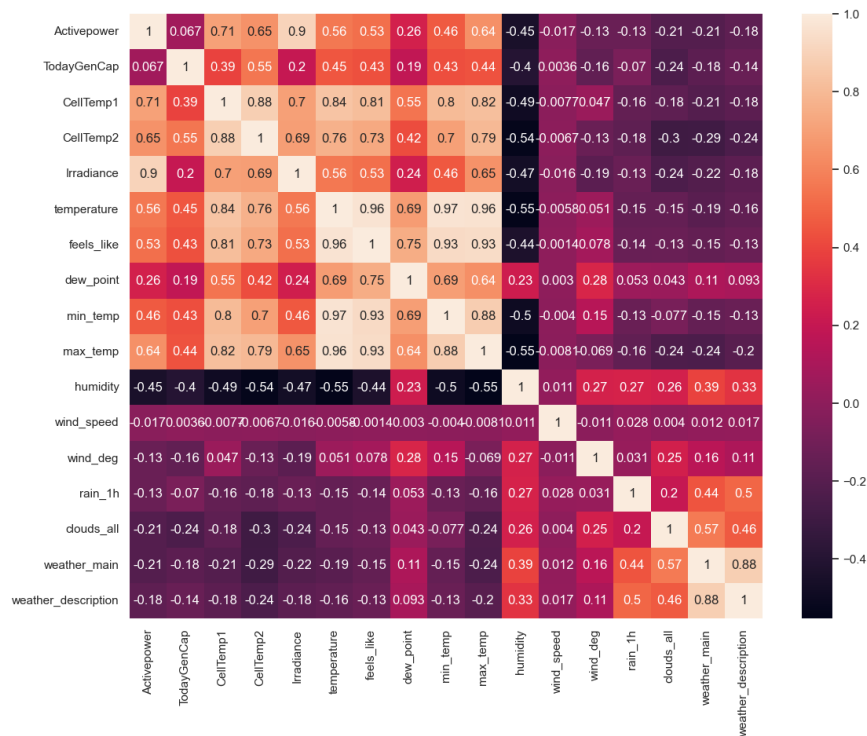
ในบทนี้จะแสดงผลการทดลองที่ได้จากการทำนายผ่านโมเดลการเรียนรู้ของเครื่องแต่ละชนิด ทั้งโมเดลการวิเคราะห์การถดถอยเชิงเส้น โมเดลแรนดอมฟอเรสต์ โมเดลซัพพอร์ตเวกเตอร์แมชชีน และโมเดล อาริมา รวมไปถึงโมเดลที่ได้จากการผสมการทำนายเพื่อปรับปรุงประสิทธิภาพให้ผลการทำนายดีขึ้น โดยจะแบ่งการทดลองออกเป็น 3 ส่วน คือ การทำกระบวนการก่อนการเรียนรู้ การทำนายผลผ่านโมเดลการเรียนรู้ของเครื่อง และการประเมินความสามารถในการทำนายของแต่ละโมเดล โดยผลการทดลองด้านล่างนี้จะแสดงข้อมูลที่ได้จากการทดลองในแต่ละขั้นตอน พร้อมทั้งอธิบายพารามิเตอร์ที่ใช้และโค้ดโปรแกรมที่เกี่ยวข้อง พร้อมทั้งแสดงแผนภาพการเปรียบเทียบผลของแต่ละโมเดลการเรียนรู้ของเครื่อง

4.1 การทดสอบหาความสัมพันธ์ระหว่างข้อมูล

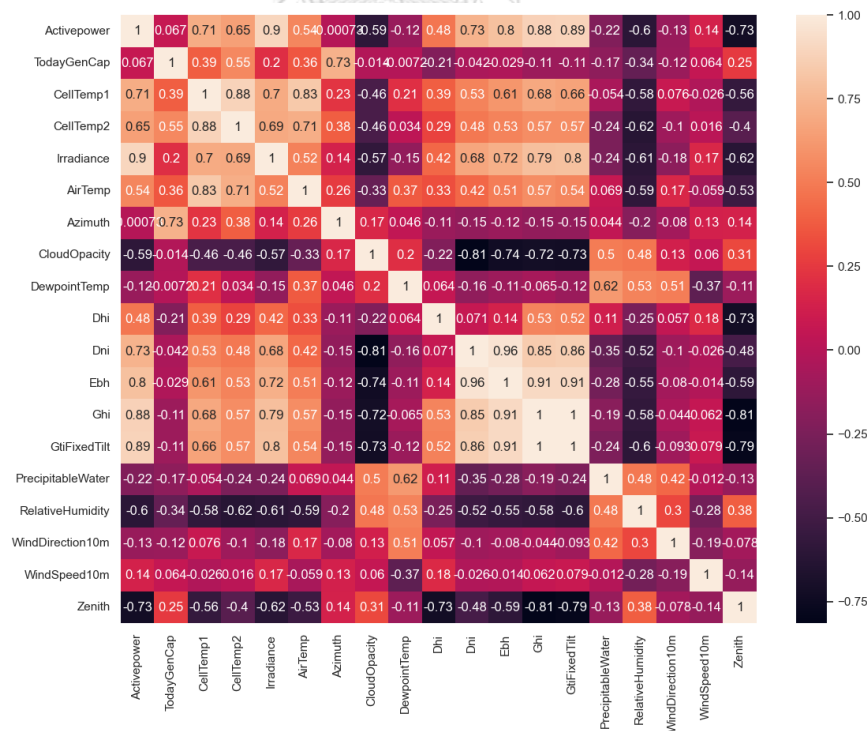
ในการทดสอบหาค่าความสัมพันธ์ระหว่างข้อมูลจากโรงไฟฟ้าพลังงานแสงอาทิตย์และข้อมูลสภาพอากาศที่ได้รับจากฐานข้อมูลขององค์กรเอกชน โดยรูปที่ 25 แสดงความสัมพันธ์ทางสถิติระหว่างข้อมูลโรงไฟฟ้าพลังงานแสงอาทิตย์เปรียบเทียบกับข้อมูลจาก www.solcast.com และ รูปที่ 26 แสดงความสัมพันธ์ทางสถิติระหว่างข้อมูลโรงไฟฟ้าพลังงานแสงอาทิตย์เปรียบเทียบกับข้อมูลจาก www.openweathermap.com ในขั้นตอนนี้ผู้วิจัยได้ทดสอบหาความสัมพันธ์ระหว่างข้อมูลพลังงานไฟฟ้าที่ผลิตได้เปรียบเทียบกับข้อมูลทางสภาพอากาศเพื่อหาความสัมพันธ์ทางสถิติและค่าสัมประสิทธิ์ความสัมพันธ์ของเพียร์สันแสดงวิธีการหาค่าสถิติของเพียร์สันในรูปที่ 24 จากนั้นจึงเลือกเฉพาะตัวแปรที่เกี่ยวข้องมาใช้ในการทดลอง หากมีตัวแปรที่ซ้ำกันจะเลือกเฉพาะตัวแปรที่แสดงค่าความสัมพันธ์ที่สูงกว่า รวมไปถึงตัวแปรที่แสดงความสัมพันธ์กันโดยตรงจะเลือกเฉพาะตัวแปรหลักที่ได้รับผลมาจากตัวแปรอื่นผ่านกระบวนการทางคณิตศาสตร์เพื่อให้ได้ค่ามา เมื่อได้ตัวแปรหรือพารามิเตอร์ที่สัมพันธ์กับค่าพลังงานไฟฟ้าที่ผลิตได้จากแสงอาทิตย์แล้วจึงสร้างชุดข้อมูลที่มีเฉพาะพารามิเตอร์ที่สนใจกับข้อมูลจากระบบของโรงไฟฟ้าพลังงานแสงอาทิตย์ จากนั้นจึงแบ่งชุดข้อมูล นำข้อมูลเข้าสู่กระบวนการเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่อง

```
from scipy.stats import pearsonr
print(pearsonr(merged_data['Activepower'], merged_data['TodayGenCap']))
print(pearsonr(merged_data['Activepower'], merged_data['CellTemp1']))
print(pearsonr(merged_data['Activepower'], merged_data['CellTemp2']))
print(pearsonr(merged_data['Activepower'], merged_data['Irradiance']))
print(pearsonr(merged_data['Activepower'], merged_data['AirTemp']))
print(pearsonr(merged_data['Activepower'], merged_data['Azimuth']))
print(pearsonr(merged_data['Activepower'], merged_data['CloudOpacity']))
print(pearsonr(merged_data['Activepower'], merged_data['DewpointTemp']))
print(pearsonr(merged_data['Activepower'], merged_data['Dhi']))
print(pearsonr(merged_data['Activepower'], merged_data['Dni']))
print(pearsonr(merged_data['Activepower'], merged_data['Ebh']))
PearsonResult(statistic=0.06707858106034092, pvalue=4.948806052797489e-16)
PearsonResult(statistic=0.7068698798454693, pvalue=0.0)
PearsonResult(statistic=0.6509327647722032, pvalue=0.0)
PearsonResult(statistic=0.8975340006890122, pvalue=0.0)
PearsonResult(statistic=0.5447039463178205, pvalue=0.0)
PearsonResult(statistic=0.0007256210104287301, pvalue=0.9301464231366323)
PearsonResult(statistic=0.5913005393338646, pvalue=0.0)
PearsonResult(statistic=0.12448020176928114, pvalue=1.6968515811880326e-51)
PearsonResult(statistic=0.476558201281158, pvalue=0.0)
PearsonResult(statistic=0.730656709467062, pvalue=0.0)
PearsonResult(statistic=0.795300635401561, pvalue=0.0)
PearsonResult(statistic=0.8774609262476487, pvalue=0.0)
```

รูปที่ 24 ตัวอย่างวิธีการหาค่าสัมประสิทธิ์ของเพียร์สัน



รูปที่ 25 ความสัมพันธ์ระหว่างข้อมูลจากโรงไฟฟ้าและสภาพอากาศจากเว็บไซต์ SolCast



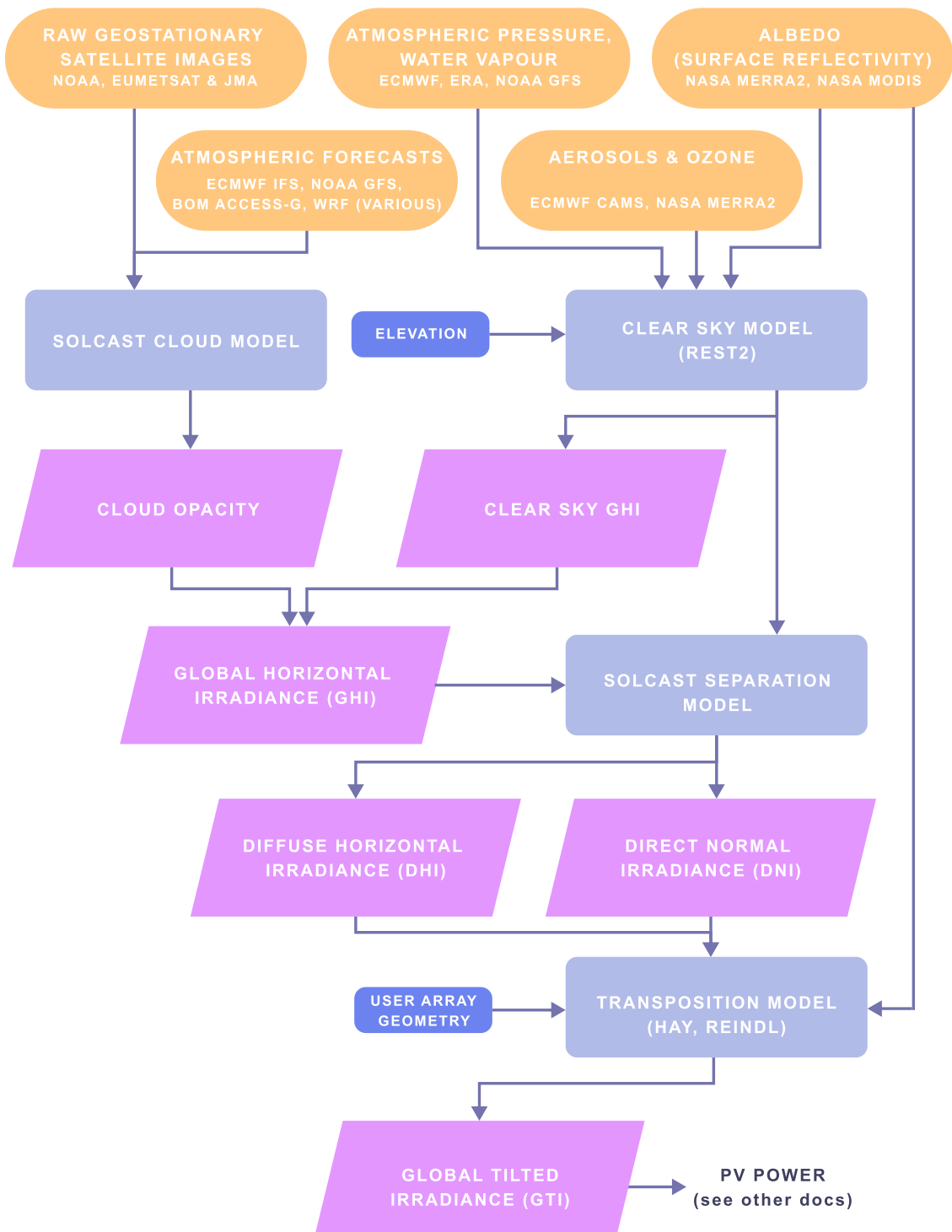
รูปที่ 26 ความสัมพันธ์ระหว่างข้อมูลจากโรงไฟฟ้าและสภาพอากาศจากเว็บไซต์ OpenWeatherMap

จากการทดสอบหาความสัมพันธ์ดังแสดงในรูปที่ 25 และ 26 พบว่ามีพารามิเตอร์ที่ส่งผลต่อการทำนายค่าพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์จำนวน 8 พารามิเตอร์ ดังต่อไปนี้

1. อุณหภูมิ (temperature, °C) คืออุณหภูมิสภาพอากาศโดยรอบโรงไฟฟ้าพลังงานแสงอาทิตย์ โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *OpenWeatherMap*
2. ปริมาณเมฆ (Cloud Opacity, %) คือปริมาณความหนาแน่นของเมฆต่อปริมาณพื้นที่ท้องฟ้าโดยรอบ แสดงถึงปริมาณที่เมฆยอมให้แสงสามารถผ่านได้ โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *SolCast*
3. อุณหภูมิต่ำสุด (Minimum Temperature, °C) คืออุณหภูมิต่ำที่สุดในชั่วโมงนั้นๆ โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *OpenWeatherMap*
4. อุณหภูมิสูงสุด (Maximum Temperature, °C) คืออุณหภูมิสูงสุดในชั่วโมงนั้นๆ โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *OpenWeatherMap*
5. ความเร็วลม (Wind Speed, m/s) คือปริมาณความเร็วลม ณ ชั่วโมงนั้นๆ โดยวัดที่บริเวณสูงกว่าระดับน้ำ ทะเลประมาณ 10 เมตร โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *OpenWeatherMap*
6. ความชื้นสัมพัทธ์ (Relative Humidity, %) คืออัตราส่วนของปริมาณความชื้นในอากาศเปรียบเทียบกับความชื้นอิ่มตัว ดังสมการที่ (21) โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *SolCast*

$$RH = \frac{\text{Actual Relative Humidity}}{\text{Saturation Relative Humidity}} \times 100\% \quad (21)$$

7. มุมของจุดสุดยอด (zenith Angle, degree) คือมุมระหว่างจุดสุดยอดกับขอบฟ้าโดยค่ากำหนดขณะที่พระอาทิตย์ขึ้นและตกมุมของจุดสุดยอดจะอยู่ที่ 0 องศา ในขณะที่เมื่อพระอาทิตย์ตั้งฉากอยู่บนหัวจะมีมุมของจุดสุดยอดอยู่ที่ 90 องศา ความสำคัญของพารามิเตอร์คือเนื่องจากในแต่ละชั่วโมงพระอาทิตย์จะทำมุมกับเซลล์พลังงานแสงอาทิตย์ที่แตกต่างกันซึ่งส่งผลต่อปริมาณพลังงานไฟฟ้าที่ระบบจะสามารถผลิตได้ โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *Solcast*
8. ค่าการแผ่รังสีของดวงอาทิตย์สำหรับเซลล์พลังงานแสงอาทิตย์แบบปรับมุมไม่ได้ (*Global Tilt Irradiance*) คือค่ารังสีของดวงอาทิตย์ที่เกิดจากการนำข้อมูลทางสภาพอากาศที่เกี่ยวข้องมาวิเคราะห์ข้อมูลผ่านโมเดลการประมวลผลที่ทางองค์กรเอกชนได้กำหนดไว้ ซึ่งกระบวนการในการวิเคราะห์หาค่าการแผ่รังสีของดวงอาทิตย์สามารถแสดงได้ดังรูปที่ 27 โดยพารามิเตอร์นี้นำข้อมูลมาจากเว็บไซต์ *SolCast*



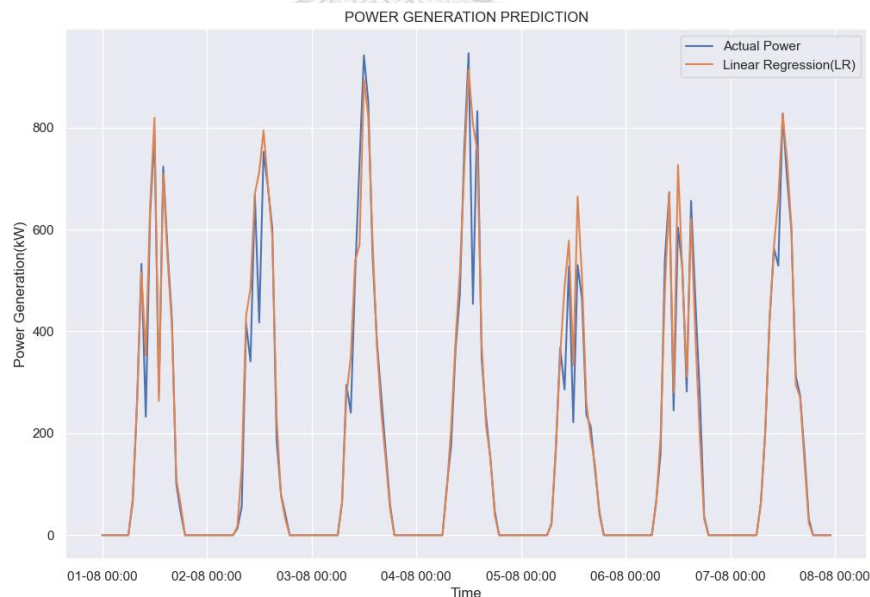
รูปที่ 27 แผนภาพแสดงโมเดลการวิเคราะห์หาค่าการแผ่รังสีของดวงอาทิตย์
 หมายเหตุ ภาพจาก : <https://solcast.com/irradiance-data-methodology>

เมื่อทราบความสัมพันธ์ระหว่างข้อมูลจึงได้สร้างชุดข้อมูลสำหรับการเรียนรู้โดยเหลือเฉพาะตัวแปรที่เกี่ยวข้อง จากนั้นจึงแบ่งชุดข้อมูลออกเป็น ชุดสำหรับเรียนรู้ 60% ชุดสำหรับตรวจสอบ 20% และชุดสำหรับทดสอบ 20% เมื่อแบ่งชุดข้อมูลออกเป็นสามชุดแล้วจึงนำชุดข้อมูลไปผ่านกระบวนการเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่องแต่ละชนิด เมื่อเรียนรู้ด้วยชุดข้อมูลสำหรับการเรียนรู้แล้วจึงทดสอบการทำนายด้วยชุดตรวจสอบและชุดทดสอบ เพื่อแสดงผลการทำนายและประเมินประสิทธิภาพในการทำนายของโมเดลการเรียนรู้ของเครื่องแต่ละชนิด

4.2 กระบวนการเรียนรู้ผ่านโมเดล

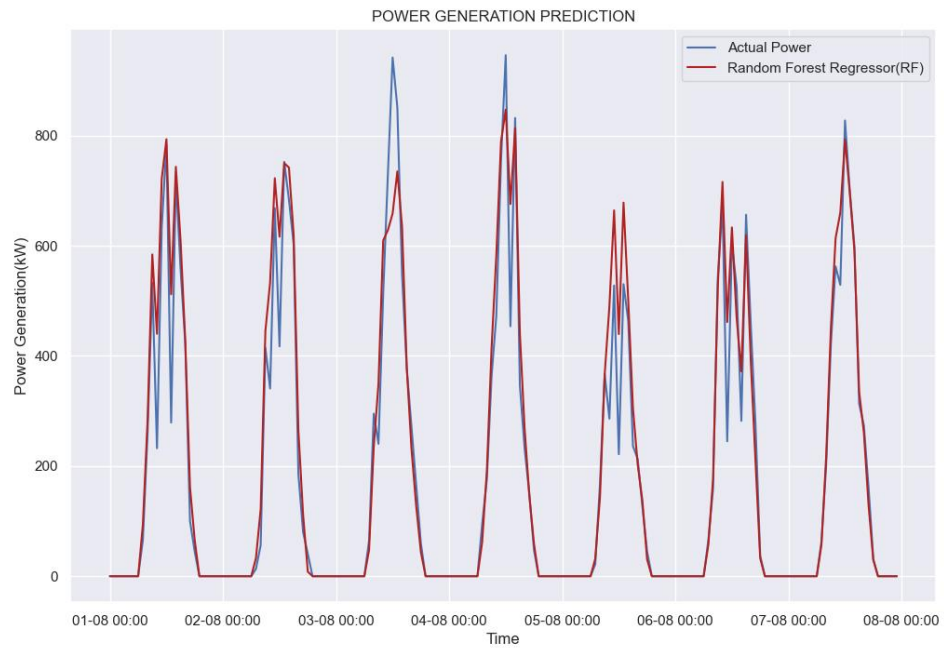
ในหัวข้อนี้จะแสดงผลการทำนายค่าพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์ผ่านโมเดลการเรียนรู้ของเครื่องแต่ละชนิด ผลการทำนายที่จะแสดงต่อไปนี้ได้นำข้อมูลการทำนายในช่วงระหว่างวันที่ 1 ถึง 7 สิงหาคม 2566 มาใช้ในการแสดงผลเปรียบเทียบการทำนายโดยข้อมูลการทำนายแสดงในตารางที่ 8 ซึ่งสามารถแสดงผลการเปรียบเทียบข้อมูลระหว่างแต่ละโมเดลกับค่าพลังงานที่สามารถผลิตได้จริงดังรูปที่ 28 ถึง 33

4.2.1 โมเดลการวิเคราะห์การถดถอยเชิงเส้น



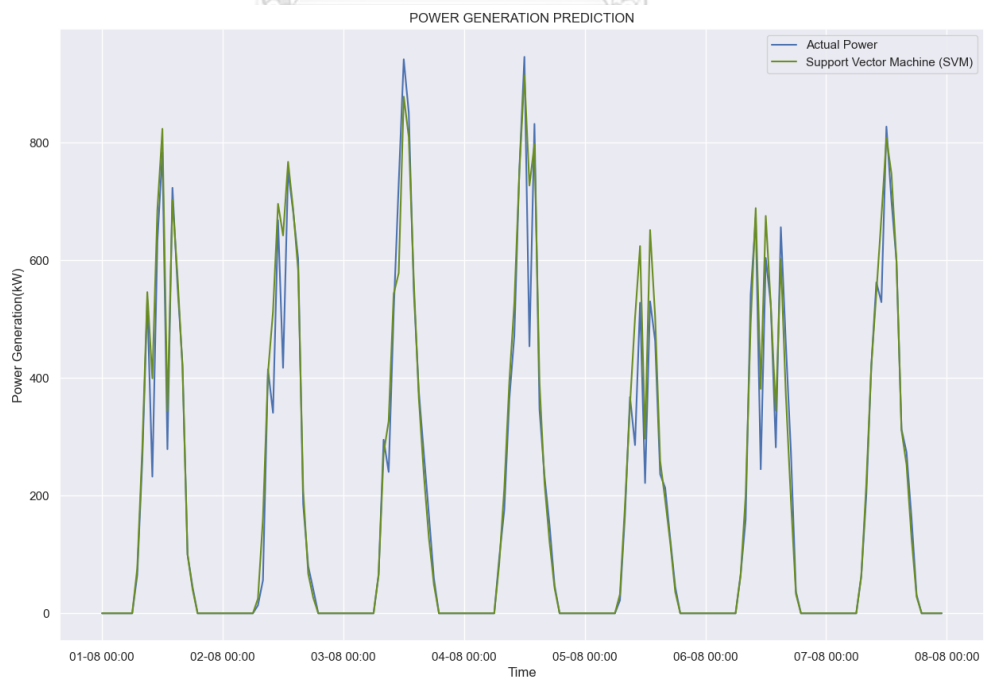
รูปที่ 28 ผลการทำนายผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้นเปรียบเทียบกับค่าที่ผลิตได้จริง

4.2.2 โมเดลแรนดอมฟอเรสต์



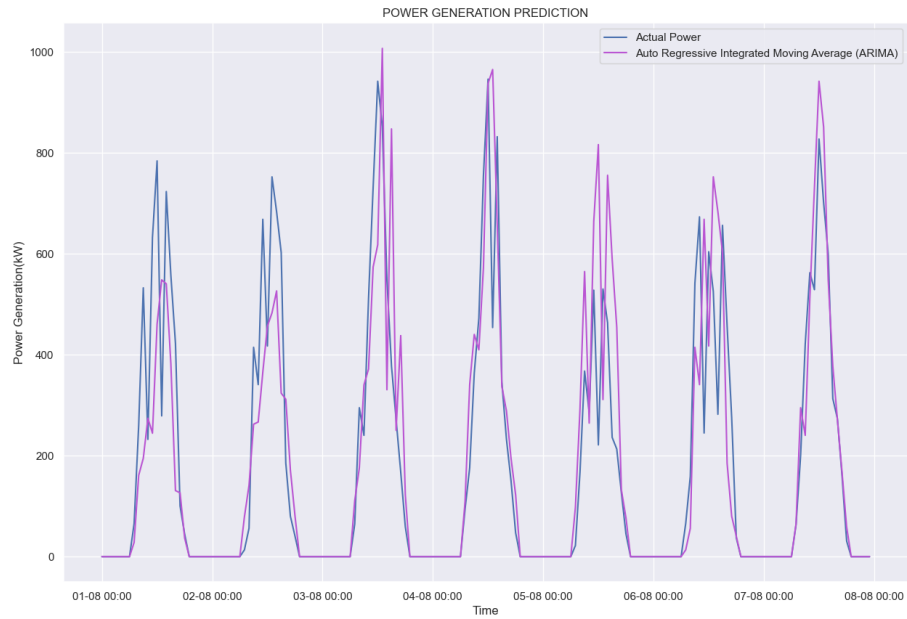
รูปที่ 29 ผลการทำนายผ่านโมเดลการวิเคราะห์การถดถอยเชิงเส้นเปรียบเทียบกับค่าที่ผลิตได้จริง

4.2.3 โมเดลซัพพอร์ตเวกเตอร์แมชชีน



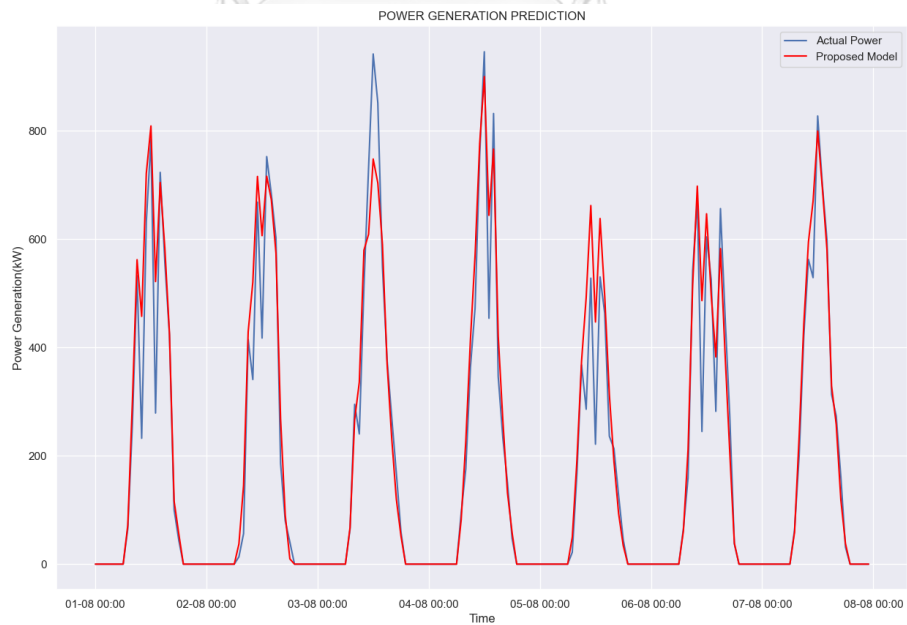
รูปที่ 30 ผลการทำนายผ่านโมเดลซัพพอร์ตเวกเตอร์แมชชีนเปรียบเทียบกับค่าที่ผลิตได้จริง

4.2.4 โมเดลอาร์มีนา

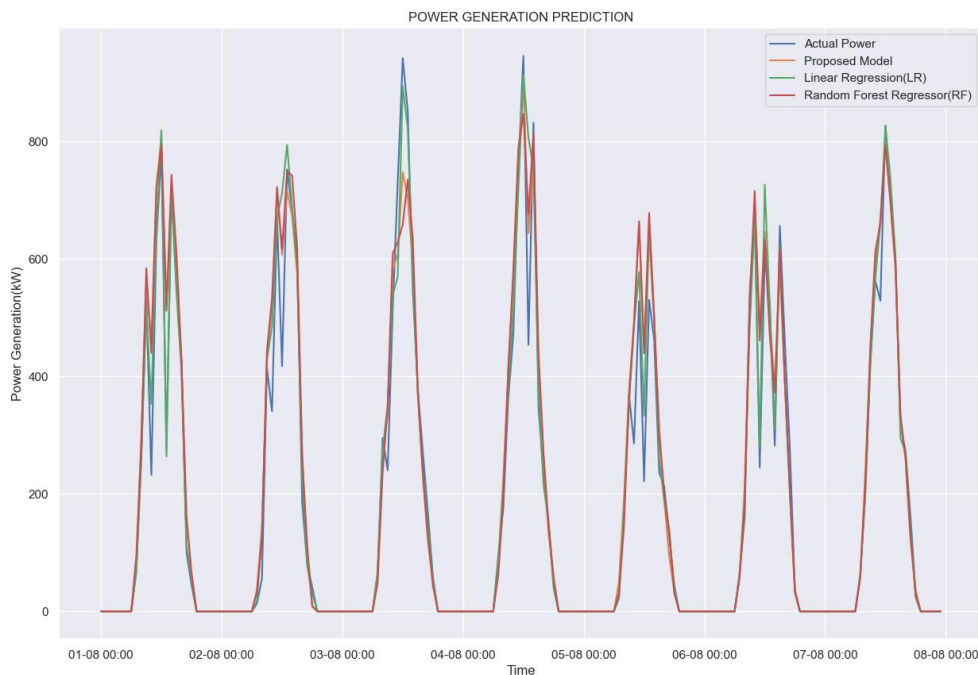


รูปที่ 31 ผลการทำนายผ่านโมเดลอาร์มีนาเปรียบเทียบกับค่าที่ผลิตได้จริง

4.2.5 โมเดลที่เสนอ (LinearRegression+RandomForest+XgBoost+LightGBM)



รูปที่ 32 ผลการทำนายผ่านโมเดลที่เสนอเปรียบเทียบกับค่าที่ผลิตได้จริง



รูปที่ 33 ผลการทำนายผ่านโมเดลที่เสนอเปรียบเทียบกับ LR และ RF

จากรูปที่ 28 ถึง 33 แสดงผลการทำนายผ่านโมเดลการเรียนรู้ของเครื่องชนิดต่างๆ แสดงให้เห็นถึงความแตกต่างระหว่างผลที่ได้จากการทำนายเปรียบเทียบกับค่าที่ผลิตได้จริง ในการแสดงผลได้กำหนดให้แกนตั้งแสดงค่าของค่าพลังงานไฟฟ้าที่ผลิตได้ในหน่วยกิโลวัตต์ และแกนนอนแสดงข้อมูลตามลำดับเวลา จากนั้นจึงนำผลการทดลองที่ได้ไปประเมินผ่านฟังก์ชันตัวประเมินในรูปที่ 24 และแสดงผลการประเมินในหัวข้อที่ 4.3

CHULALONGKORN UNIVERSITY

4.3 ผลการประเมินประสิทธิภาพโมเดลการเรียนรู้ของเครื่อง

ในหัวข้อนี้จะแสดงถึงผลการเปรียบเทียบประสิทธิภาพโมเดลการเรียนรู้ของเครื่องผ่านตัวประเมินผลที่ได้กล่าวไว้ในบทที่ 3 โดยตัวประเมินที่ใช้มีดังนี้

1. ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error, MAE)
2. ค่ารากที่สองของค่าความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root Mean Square Error, RMSE)
3. ค่าทางสถิติวัดความเข้าใจใส่เส้นถดถอย (R-Squared Score)

ตารางที่ 2 ค่าประเมินประสิทธิภาพโมเดลการเรียนรู้ของเครื่อง

| Model | Linear Regression | | | Random Forest | | |
|--------------------|--------------------|------------------|-------------------|------------------|------------------|------------------|
| Metrics | Train | Valid | Test | Train | Valid | Test |
| MAE | 66.40 (6.42%) | 60.07 (5.80%) | 64.46 (6.23%) | 17.61 (1.72%) | 47.37 (4.55%) | 30.79 (2.97%) |
| RMSE | 106.17 (10.26%) | 91.96 (8.88%) | 100.39 (9.69%) | 33.28 (3.25%) | 80.04 (7.73%) | 63.38 (6.12%) |
| R ² (%) | 88.77 | 91.45 | 88.44 | 98.89 | 93.52 | 93.95 |

| Model | Support Vector Machine | | | ARIMA | Proposed Model | | |
|--------------------|------------------------|--------------------|--------------------|--------------------|------------------|------------------|------------------|
| Metrics | Train | Valid | Test | Test | Train | Valid | Test |
| MAE | 74.03 (7.15%) | 73.51 (7.10%) | 74.63 (7.21%) | 77.06 (7.45%) | 39.73 (3.84%) | 46.35 (4.48%) | 29.56 (2.86%) |
| RMSE | 112.35 (10.85%) | 107.47 (10.38%) | 109.24 (10.55%) | 142.66 (13.78%) | 67.34 (6.51%) | 75.67 (7.31%) | 60.79 (5.87%) |
| R ² (%) | 87.42 | 88.32 | 86.31 | 69.34 | 95.48 | 94.21 | 94.43 |

จากตารางที่ 2 แสดงผลการประเมินของโมเดลการเรียนรู้ของเครื่องแต่ละชนิด โดยนำข้อมูลแต่ละชุดไปประเมินประสิทธิภาพการทำนายผ่านฟังก์ชันตัวประเมินที่สร้างไว้ในรูปที่ 24 โดยค่าที่แสดงในตารางการเปรียบเทียบตัวประเมินประสิทธิภาพโมเดลเกิดจากค่าเฉลี่ยของค่าความผิดพลาดของข้อมูลในแต่ละชุดการทดลอง จากนั้นนำผลการทดลองของโมเดลที่เสนอไปเปรียบเทียบกับโมเดลการเรียนรู้ในงานวิจัยอื่นในหัวข้อที่ 1.2 โดยสามารถแสดงผลการเปรียบเทียบดังแสดงในตารางที่ 3 ซึ่งในแต่ละบทวรรณกรรมจะนำเฉพาะโมเดลที่ให้ประสิทธิภาพที่ดีที่สุดมาใช้ในการเปรียบเทียบ

ตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่นในการทบทวนวรรณกรรม

| วรรณกรรม | โมเดลหลักที่ใช้ในงานวิจัย | การประเมินประสิทธิภาพโมเดล |
|----------|--|--|
| [4] | ANN | ในวันที่มีแดดและท้องฟ้าปลอดโปร่ง RMSE 12.5%, MAE 14.8% ในวันที่มีเมฆบางส่วน RMSE 24%, MAE 58.1% ในวันที่มีเมฆมาก RMSE 36.9%, MAE 345% |
| [5] | MME (RHNN, GTNN) (WRF) MME (GTSVM, GTNN) (ECMWF) MME (GTNN (WRF), GTNN (ECMWF)) MME | RMSE 11.6%, MAE 7.63% RMSE 10.9%, MAE 6.80% RMSE 10.8%, MAE 7.11% RMSE 10.5%, MAE 6.81% |
| [6] | ARIMA | RMSE 71.6 W/m ² |
| [7] | PFLRM with MLP | RMSE 63.39 W/m ² , MAD 35.61 W/m ² |
| [8] | SVM and GPR | โมเดลที่ดีที่สุด คือ Matern 5/2 GPR RMSE 7.967%, MAE 5.302% โมเดลที่แย่ที่สุด คือ Cubic SVM RMSE 21.72%, MAE 15.667% |
| [9] | LSTM with ensemble of ML | ไม่ได้ระบุค่าประสิทธิภาพในการ ทดลอง |
| [10] | ANN, SVM, MARS, CART, M5, และ Random Forest | โมเดลที่ให้ประสิทธิภาพที่ดีที่สุด คือ Random Forest Overall % error 10.06% |

| วรรณกรรม | โมเดลหลักที่ใช้ในงานวิจัย | การประเมินประสิทธิภาพโมเดล |
|----------|--|---|
| [11] | หาค่าความน่าจะเป็นของค่าพลังงานไฟฟ้าจากเซลล์แสงอาทิตย์โดยใช้โมเดล hybrid spatiotemporal ในการทำนาย | MAE 7.40% |
| [12] | ANN (AFFNN) | วันที่มีแดดมาก RMSE 3.3% วันที่มีเมฆบางส่วน RMSE 4.2% วันที่ฟ้าครึ้ม/เมฆมาก RMSE 5.3% |
| [13] | CNN ร่วมกับภาพท้องฟ้า | ค่าความสามารถในการทำนาย วันที่มีแดด/ฟ้าโปร่ง 26.22% วันที่มีเมฆ 16.11% |
| [14] | LSTM ร่วมกับ RNN | โรงไฟฟ้าที่ Atlanta, USA RMSE 45.84 W/m ² MAE 31.86 W/m ² R ² 97% โรงไฟฟ้าที่ New York, USA RMSE 41.37 W/m ² MAE 30.19 W/m ² R ² 97% โรงไฟฟ้าที่ Hawaii, USA RMSE 66.69 W/m ² MAE 46.04 W/m ² R ² 95% |
| [15] | LSTM NN with Synthetic WF | ประสิทธิภาพการทำนายเพิ่มขึ้น 33% เทียบการพยากรณ์รายชั่วโมง 46% เทียบการพยากรณ์รายวัน |

| วรรณกรรม | โมเดลหลักที่ใช้ในงานวิจัย | การประเมินประสิทธิภาพโมเดล |
|----------------|---------------------------|---|
| [16] | LSTM ร่วมกับ DNN | RMSE 3.62%, MAE 2.03%, Accuracy 97.88% |
| [17] | RF | RMSE 76.15 W/m ² MAE 34.41 W/m ² R ² 94% |
| [18] | LSTM ร่วมกับ RNN | RMSE 6.29%, MAE 2.78%, Accuracy 96.19% |
| Proposed Model | LR ร่วมกับ RF | RMSE 5.87%, MAE 2.86%, R ² 94.43% |

บทที่ 5

สรุปและวิเคราะห์ผลการทดลอง

ในงานวิทยานิพนธ์ฉบับนี้ได้นำข้อมูลจากโรงไฟฟ้าพลังงานแสงอาทิตย์ขนาด 5 เมกะวัตต์ ตั้งอยู่ที่ อ.ทับสะแก จ.ประจวบคีรีขันธ์ และข้อมูลทางสภาพอากาศ ณ ตำแหน่งที่โรงไฟฟ้าตั้งอยู่ (ละติจูด 11.475 และ ลองจิจูด 99.597) จากเว็บไซต์ SolCast และ OpenweatherMap มาใช้ในการทดลอง จากที่ได้หาค่าความสัมพันธ์ทางสถิติและค่าความสัมพันธ์ของเพียร์สันระหว่างตัวแปรเป้าหมายกับตัวแปรอินพุต แสดงให้เห็นว่ามีตัวแปร 8 ตัวที่สัมพันธ์กับการผลิตไฟฟ้าจากพลังงานแสงอาทิตย์ คือ อุณหภูมิ ปริมาณเมฆ อุณหภูมิต่ำสุด อุณหภูมิสูงสุด ความเร็วลม ความชื้นสัมพัทธ์ มุมเซนิต และค่าการแผ่รังสีของดวงอาทิตย์สำหรับเซลล์แสงอาทิตย์แบบอยู่กับที่ เมื่อได้ตัวแปรที่จะนำมาใช้ทำนายแล้วได้สร้างชุดของข้อมูลและแบ่งออกเป็นสามชุดในอัตราส่วน 60 20 และ 20 โดยการแบ่งชุดข้อมูลได้นำเอาฤดูกาลมาใช้ในการแบ่งโดยข้อมูลฤดูร้อนและฤดูหนาวถูกจัดไว้เป็นฤดูเดียวกัน เนื่องจากภูมิอากาศประเทศไทยในฤดูหนาวมีความแตกต่างจากฤดูร้อนไม่มากและด้วยปริมาณข้อมูลย้อนหลังที่มีจำกัด ในขณะที่ฤดูฝนเป็นช่วงที่สภาพอากาศมีความแปรปรวนมากที่สุดจึงได้จัดข้อมูลช่วงฤดูฝนเป็นอีกประเภทหนึ่ง ซึ่งในงานวิจัยนี้ได้นำข้อมูลเดือนเมษายนและสิงหาคมมาใช้เป็นชุดตรวจสอบ และข้อมูลเดือนพฤษภาคมและกันยายนมาใช้เป็นชุดทดสอบ จากนั้นนำข้อมูลแต่ละชุดไปเรียนรู้ผ่านโมเดลการเรียนรู้ของเครื่องแต่ละชนิด ได้แก่ โมเดลการวิเคราะห์การถดถอยเชิงเส้น โมเดลแรนดอมฟอเรสต์ โมเดลซัพพอร์ตเวกเตอร์แมชชีน โมเดลอาร์มา และโมเดลที่ได้จากการ ensemble ของโมเดลการวิเคราะห์การถดถอยเชิงเส้นและโมเดลแรนดอมฟอเรสต์เข้าด้วยกันพร้อมปรับจูนค่าไฮเปอร์พารามิเตอร์ผ่านฟังก์ชัน *RandomizedSearchCV* จากการปรับจูนจึงได้ค่าพารามิเตอร์ที่นำมาใช้ดังนี้ $n_estimators=880$, $min_sample_leaf=12$, $min_sample_split=19$ และ $max_features=0.5$ นอกจากการปรับจูนค่าไฮเปอร์พารามิเตอร์ในการทดลองยังได้นำโมเดล XGBoost และ LightGBM มาช่วยปรับปรุงประสิทธิภาพการทำนายของโมเดลการเรียนรู้ด้วย จากผลการทดลองพบว่าโมเดลการวิเคราะห์การถดถอยเชิงเส้นช่วยปรับปรุงการทำนายของโมเดลแรนดอมฟอเรสต์ในช่วงที่สภาพอากาศปกติ ในขณะที่เดียวกันเมื่อพบสภาพอากาศที่แปรปรวนผลการทำนายของโมเดลแรนดอมฟอเรสต์ก็สามารถปรับปรุงผลการทำนายในช่วงเวลานี้ได้ดีกว่า การนำทั้งสองโมเดลการเรียนรู้มาช่วยกันปรับปรุงผลการทำนายจึงเป็นไปตามวัตถุประสงค์ แต่ด้วยความแม่นยำของข้อมูลทางสภาพอากาศที่ไม่ได้เกิดจากการนำเครื่องมือวัดไปบันทึก ณ สถานที่จริง จึงส่งผลให้ผลการประเมินประสิทธิภาพการทำนายออกมาค่อนข้างไม่ต่างจากการทำนายด้วยโมเดลประเภทเดียว รวมไปถึงการปรับจูนค่าพารามิเตอร์เนื่องจากข้อมูลมีปริมาณมากจึงไม่ได้นำฟังก์ชัน

GridSearchCV มาใช้ในการค้นหาจึงทำให้ค่าพารามิเตอร์ที่ได้นำมาใช้ในการทดลองอาจไม่ใช่ค่าที่ดีที่สุด แต่เป็นเพียงค่าที่แสดงให้เห็นว่าพัฒนาประสิทธิภาพการทำนายให้ดีขึ้น จากการประเมินผลการทดลองพบว่าโมเดลที่เสนอมีค่า R^2 Score เพิ่มขึ้น 0.5% มากกว่าการทำนายด้วยโมเดลแรนดอมฟอเรสต์เพียงชนิดเดียว จากการเปรียบเทียบประสิทธิภาพกับโมเดลการเรียนรู้ของเครื่องชนิดอื่นในตารางที่ 3 แสดงให้เห็นว่าโมเดลที่เสนอให้ประสิทธิภาพในการทำนายได้ดีใกล้เคียงการนำโมเดล LSTM ที่ใช้ปรับปรุงประสิทธิภาพการทำนายโครงข่ายประสาทเทียมหลายชนิด ซึ่งแสดงให้เห็นว่าการนำโมเดลการวิเคราะห์การถดถอยเชิงเส้นร่วมกับโมเดลแรนดอมฟอเรสต์สามารถนำมาใช้ทำนายค่าพลังงานไฟฟ้าที่ผลิตจากเซลล์แสงอาทิตย์ล่วงหน้าได้ซึ่งโมเดลนี้ยังช่วยลดข้อจำกัดในการทำนายระยะสั้นหรือระยะยาวเหมือนโมเดลบางชนิดที่ยังมีข้อจำกัดนี้ เช่น โมเดลอาร์มา จะสามารถให้ผลการทำนายที่แม่นยำได้เฉพาะการทำนายระยะสั้น เนื่องจากโมเดลการเรียนรู้ของเครื่องที่เสนอในงานวิทยานิพนธ์ฉบับนี้ใช้ค่าตัวแปรการพยากรณ์อากาศมาใช้ในการทำนาย หากข้อมูลการพยากรณ์อากาศจากฐานข้อมูลการพยากรณ์ที่มีความแม่นยำ อาจสามารถนำข้อมูลมาใช้ทำนายค่าพลังงานไฟฟ้าที่ผลิตได้ล่วงหน้าประมาณ 5-7 วัน

สำหรับการปรับปรุงประสิทธิภาพในอนาคต ผู้วิจัยเสนอให้มีการปรับปรุงระบบการบันทึกข้อมูลของโรงไฟฟ้าพลังงานแสงอาทิตย์ให้มีเสถียรภาพมากยิ่งขึ้นเพื่อลดความผิดพลาดในการบันทึกและการขาดหายไปของข้อมูล รวมไปถึงการนำเครื่องมือวัดตัวแปรทางสภาพอากาศที่มีความสัมพันธ์กับค่าพลังงานไฟฟ้าที่ได้จากเซลล์พลังงานแสงอาทิตย์ โดยนำเครื่องมือวัดไปตั้งอยู่ในบริเวณเซลล์พลังงานแสงอาทิตย์โซนที่สนใจเพื่อให้ผลมีความแม่นยำมากที่สุด อีกทั้งในอนาคตยังสามารถพัฒนาวิธีการในการผสมการทำนายของโมเดลการเรียนรู้ของเครื่องโดยหาฟังก์ชันชนิดอื่นที่สามารถค้นหาค่าน้ำหนักที่จะทำให้ประสิทธิภาพในการทำนายของโมเดลสูงขึ้น รวมไปถึงการปรับให้โมเดลการทำนายมีความสามารถในการเลือกค่าที่ดีที่สุดของแต่ละโมเดลมาใช้โดยไม่เพียงแต่เฉลี่ยค่าตามน้ำหนักของแต่ละโมเดล สุดท้ายนี้ได้เสนอให้นำโมเดลที่สร้างขึ้นไปพัฒนาร่วมกับโมเดลการเรียนรู้เชิงลึก (Deep Learning) หรือโครงข่ายประสาทเทียม (Neural Network) เพื่อพัฒนาประสิทธิภาพการทำนายให้ดียิ่งขึ้น

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| 30 | | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 31 | | 31 | | 31 | | 31 | 31 | | 31 | | 31 |

หมายเหตุ (ข้อผิดพลาดของข้อมูล) :



= ข้อมูลปกติ



= มีข้อมูลบางช่วงเวลาที่ขาดหายไป



= ข้อมูลที่บันทึกขาดหายไปทั้งหมด

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| 30 | | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 31 | | 31 | | 31 | | 31 | 31 | | 31 | | 31 |

หมายเหตุ (ข้อผิดพลาดของข้อมูล) :



= ข้อมูลปกติ



= มีข้อมูลบางช่วงเวลาที่ขาดหายไป



= ข้อมูลที่บันทึกขาดหายไปทั้งหมด

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| 30 | | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 31 | | 31 | | 31 | | 31 | 31 | | 31 | | 31 |

หมายเหตุ (ข้อผิดพลาดของข้อมูล) :



= ข้อมูลปกติ



= มีข้อมูลบางช่วงเวลาที่ขาดหายไป



= ข้อมูลที่บันทึกขาดหายไปทั้งหมด

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| 30 | | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 31 | | 31 | | 31 | | 31 | 31 | | 31 | | 31 |

หมายเหตุ (ข้อผิดพลาดของข้อมูล) :



= ข้อมูลปกติ



= มีข้อมูลบางช่วงเวลาที่ขาดหายไป



= ข้อมูลที่บันทึกขาดหายไปทั้งหมด

ตารางที่ 8 ผลการทำนายระหว่างวันที่ 1 ถึง 7 สิงหาคม 2566 และค่าพลังงานที่ผลิตได้จริง

| วันที่ | เวลา | Actual | LR | RF | SVM | ARIMA | PM |
|----------|----------|--------|--------|--------|--------|--------|--------|
| 8/1/2023 | 7:00:00 | 65.3 | 71.42 | 93.65 | 76.42 | 28.13 | 71.75 |
| 8/1/2023 | 8:00:00 | 259.98 | 262.73 | 283.1 | 280.31 | 161.18 | 301.58 |
| 8/1/2023 | 9:00:00 | 532.63 | 516.38 | 584.36 | 546.14 | 194.92 | 562.13 |
| 8/1/2023 | 10:00:00 | 232.36 | 353.05 | 440.14 | 399.47 | 274.27 | 457.5 |
| 8/1/2023 | 11:00:00 | 632.5 | 647.78 | 721.63 | 685.74 | 244.6 | 721.76 |
| 8/1/2023 | 12:00:00 | 784.16 | 819.38 | 793.53 | 823.82 | 460.68 | 809.06 |
| 8/1/2023 | 13:00:00 | 278.96 | 263.68 | 511.86 | 343.88 | 548.41 | 521.76 |
| 8/1/2023 | 14:00:00 | 723.42 | 710.43 | 743.48 | 703.3 | 540.64 | 704.72 |
| 8/1/2023 | 15:00:00 | 557.23 | 542.43 | 609.75 | 574.12 | 382.21 | 583.51 |
| 8/1/2023 | 16:00:00 | 422.4 | 407.52 | 426.99 | 417.06 | 130.86 | 425.76 |
| 8/1/2023 | 17:00:00 | 100.34 | 109.05 | 161.9 | 99.72 | 126.5 | 116.65 |
| 8/1/2023 | 18:00:00 | 44.59 | 59.7 | 65.51 | 40.83 | 36.69 | 56.59 |
| 8/2/2023 | 7:00:00 | 13.43 | 17.22 | 33.02 | 24.81 | 78.93 | 37.26 |
| 8/2/2023 | 8:00:00 | 56.18 | 134.13 | 121.87 | 161.5 | 141.32 | 145.04 |
| 8/2/2023 | 9:00:00 | 414.8 | 430.11 | 445.2 | 408.58 | 262.46 | 429.14 |
| 8/2/2023 | 10:00:00 | 340.87 | 484.71 | 533.6 | 513.31 | 266.76 | 520.56 |
| 8/2/2023 | 11:00:00 | 668.39 | 670.5 | 722.69 | 696.23 | 365.5 | 715.87 |
| 8/2/2023 | 12:00:00 | 417.41 | 712.18 | 616.53 | 642.44 | 456.63 | 606.47 |
| 8/2/2023 | 13:00:00 | 752.51 | 794.53 | 749.87 | 767.63 | 483.05 | 716.25 |
| 8/2/2023 | 14:00:00 | 683.4 | 687.14 | 742.06 | 687.65 | 526.48 | 672.81 |
| 8/2/2023 | 15:00:00 | 602.58 | 586.11 | 619.61 | 582.77 | 324.01 | 574.42 |
| 8/2/2023 | 16:00:00 | 185.36 | 227.79 | 264.72 | 208.97 | 311.88 | 268.23 |
| 8/2/2023 | 17:00:00 | 80.1 | 77.14 | 112.01 | 67.23 | 171.65 | 90.41 |

| วันที่ | เวลา | Actual | LR | RF | SVM | ARIMA | PM |
|----------|----------|--------|--------|--------|--------|---------|--------|
| 8/2/2023 | 18:00:00 | 40.79 | 30.92 | 8.29 | 26.29 | 80.85 | 9.61 |
| 8/3/2023 | 7:00:00 | 64.81 | 65.31 | 47.51 | 65.89 | 111.16 | 67.54 |
| 8/3/2023 | 8:00:00 | 295.09 | 275.34 | 235.26 | 272.03 | 175.96 | 267.24 |
| 8/3/2023 | 9:00:00 | 240.36 | 349.91 | 350.17 | 326.86 | 340.11 | 336.18 |
| 8/3/2023 | 10:00:00 | 512.14 | 537.94 | 609.68 | 543.25 | 371.74 | 579.64 |
| 8/3/2023 | 11:00:00 | 731.26 | 569.83 | 628.42 | 578.41 | 573.87 | 609.05 |
| 8/3/2023 | 12:00:00 | 941.97 | 894.4 | 658.68 | 878.52 | 617.94 | 747.98 |
| 8/3/2023 | 13:00:00 | 851.4 | 822.06 | 735.17 | 809.99 | 1007.09 | 703.62 |
| 8/3/2023 | 14:00:00 | 548.03 | 574.9 | 635.34 | 565.02 | 330.74 | 589.91 |
| 8/3/2023 | 15:00:00 | 377.19 | 367.96 | 379.04 | 365.51 | 847.5 | 370.22 |
| 8/3/2023 | 16:00:00 | 271.38 | 240.72 | 231.34 | 235.86 | 250.27 | 232.09 |
| 8/3/2023 | 17:00:00 | 167.88 | 148.55 | 127.61 | 128.26 | 438.01 | 119.54 |
| 8/3/2023 | 18:00:00 | 59.28 | 52.06 | 43.58 | 48.76 | 122.1 | 51.89 |
| 8/4/2023 | 7:00:00 | 94.5 | 87.99 | 61.7 | 85.06 | 109.35 | 81.57 |
| 8/4/2023 | 8:00:00 | 175.93 | 210.6 | 192.29 | 212.51 | 340.45 | 227.28 |
| 8/4/2023 | 9:00:00 | 363.66 | 379.55 | 415.15 | 393.76 | 440.38 | 414.42 |
| 8/4/2023 | 10:00:00 | 472.36 | 515.94 | 582.2 | 528.44 | 409.83 | 570.16 |
| 8/4/2023 | 11:00:00 | 755.08 | 718.84 | 789.07 | 752.05 | 572.99 | 776.7 |
| 8/4/2023 | 12:00:00 | 946.09 | 913.99 | 847.3 | 914.87 | 937.55 | 900.32 |
| 8/4/2023 | 13:00:00 | 453.97 | 806.49 | 675.92 | 727.31 | 965.25 | 644.04 |
| 8/4/2023 | 14:00:00 | 832.07 | 754.22 | 813.76 | 797.85 | 656.44 | 766.45 |
| 8/4/2023 | 15:00:00 | 346.5 | 374.6 | 441.06 | 392.74 | 337.55 | 421.55 |
| 8/4/2023 | 16:00:00 | 232.86 | 214.81 | 269.19 | 221.27 | 288.97 | 265.61 |
| 8/4/2023 | 17:00:00 | 150.25 | 152.22 | 147.61 | 121.31 | 194.59 | 131.47 |

| วันที่ | เวลา | Actual | LR | RF | SVM | ARIMA | PM |
|----------|----------|--------|--------|--------|--------|--------|--------|
| 8/4/2023 | 18:00:00 | 47.21 | 39.03 | 58.71 | 42.86 | 122.41 | 57.93 |
| 8/5/2023 | 7:00:00 | 21.94 | 23.92 | 32.04 | 32.43 | 97.57 | 49.93 |
| 8/5/2023 | 8:00:00 | 167.35 | 173.06 | 143.26 | 183.6 | 292.25 | 191.98 |
| 8/5/2023 | 9:00:00 | 367.73 | 353.12 | 366.26 | 353.81 | 564.9 | 374.92 |
| 8/5/2023 | 10:00:00 | 286 | 486.49 | 490.43 | 502.38 | 264.63 | 492.53 |
| 8/5/2023 | 11:00:00 | 527.97 | 578.15 | 664.32 | 624.39 | 664.77 | 662.16 |
| 8/5/2023 | 12:00:00 | 221.45 | 332.97 | 439.61 | 297.07 | 816.43 | 447.1 |
| 8/5/2023 | 13:00:00 | 530.32 | 664.34 | 678.37 | 651.67 | 311.23 | 638.17 |
| 8/5/2023 | 14:00:00 | 463.46 | 503.82 | 500.79 | 504.33 | 755.69 | 495.24 |
| 8/5/2023 | 15:00:00 | 236.26 | 259.69 | 304.93 | 259.19 | 589.5 | 313.63 |
| 8/5/2023 | 16:00:00 | 213.24 | 189.53 | 203.12 | 192.18 | 454.67 | 189.12 |
| 8/5/2023 | 17:00:00 | 126.98 | 137.12 | 137.38 | 120.39 | 132.61 | 93.57 |
| 8/5/2023 | 18:00:00 | 45.02 | 39.66 | 30.52 | 36.25 | 76.86 | 34.28 |
| 8/6/2023 | 7:00:00 | 65.76 | 62.9 | 57.16 | 64.13 | 13.43 | 61.49 |
| 8/6/2023 | 8:00:00 | 159.17 | 191.34 | 177.28 | 199.69 | 56.18 | 211.22 |
| 8/6/2023 | 9:00:00 | 541.94 | 481.56 | 525.03 | 492.05 | 414.8 | 515.71 |
| 8/6/2023 | 10:00:00 | 673.26 | 673.1 | 715.96 | 688.91 | 340.87 | 697.94 |
| 8/6/2023 | 11:00:00 | 244.92 | 280.2 | 461.49 | 381.61 | 668.39 | 486.56 |
| 8/6/2023 | 12:00:00 | 604.14 | 726.73 | 633.37 | 675.68 | 417.41 | 646.86 |
| 8/6/2023 | 13:00:00 | 525.52 | 520.18 | 469.22 | 523.93 | 752.51 | 497.36 |
| 8/6/2023 | 14:00:00 | 282.11 | 310.67 | 371.99 | 344.46 | 683.4 | 382.52 |
| 8/6/2023 | 15:00:00 | 656.46 | 621.66 | 618.86 | 602.41 | 602.58 | 582.65 |
| 8/6/2023 | 16:00:00 | 457.79 | 390.64 | 391.19 | 370.87 | 185.36 | 387.26 |
| 8/6/2023 | 17:00:00 | 275.67 | 195.89 | 209.39 | 193.03 | 80.1 | 208.2 |

| วันที่ | เวลา | Actual | LR | RF | SVM | ARIMA | PM |
|----------|----------|--------|--------|--------|--------|--------|--------|
| 8/6/2023 | 18:00:00 | 37.19 | 32.02 | 32.64 | 32.89 | 40.79 | 38.65 |
| 8/7/2023 | 7:00:00 | 63.28 | 62.7 | 57.12 | 63.19 | 64.81 | 58.82 |
| 8/7/2023 | 8:00:00 | 200.99 | 200.25 | 217.18 | 220.26 | 295.09 | 236.94 |
| 8/7/2023 | 9:00:00 | 421.22 | 424.3 | 450.57 | 421.58 | 240.36 | 443.68 |
| 8/7/2023 | 10:00:00 | 562.63 | 567.16 | 613.74 | 545.2 | 512.14 | 594.93 |
| 8/7/2023 | 11:00:00 | 529.03 | 661.09 | 660.32 | 673.4 | 731.26 | 669.97 |
| 8/7/2023 | 12:00:00 | 827.54 | 827.16 | 793.33 | 808.46 | 941.97 | 799.63 |
| 8/7/2023 | 13:00:00 | 701.47 | 736.88 | 697.39 | 747.39 | 851.4 | 695.73 |
| 8/7/2023 | 14:00:00 | 599.04 | 605.44 | 590.64 | 598.41 | 548.03 | 580.97 |
| 8/7/2023 | 15:00:00 | 312.99 | 294.72 | 334.67 | 310.94 | 377.19 | 328.56 |
| 8/7/2023 | 16:00:00 | 273.69 | 271.18 | 261.11 | 252.55 | 271.38 | 258.05 |
| 8/7/2023 | 17:00:00 | 160.9 | 143.01 | 128.77 | 127.09 | 167.88 | 118.65 |
| 8/7/2023 | 18:00:00 | 31.18 | 23.8 | 29.39 | 28.19 | 59.28 | 37.81 |

บรรณานุกรม

. "Renewable Energy Market Update-May 2022." from

<https://www.iea.org/reports/renewable-energy-market-update-may-2022>.

Jie, S., Wei-Jen, L., Yongqian, L., Yongping, Y., Peng, W. "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines." IEEE Transactions on industry applications **48**.

Leva, S. D., A.; Grimaccia, F.; Mussetta, M.; Ogliari, E. "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power." Math Comput. Simul. **2017** **131**: 88-100.

Lew, D. P., N.; Miller, D.; Jordan, G.; Clark, K.; Freeman, L. How do High Levels of Wind and Solar Impact the Grid. The Western Wind and Solar Integration Study 2010, National Renewable Energy Laboratory: Golden, CO, USA.

M. S. Hossain and H. Mahmood "Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast. ." IEEE Access **2020** **8**: 172524-172533.

M. Vetri Selvi, S. M. Investigation of Performance of Electric Load Power Forecasting in Multiple Time Horizons With New Architecture Realized in Multivariate Linear Regression and Feed-Forward Neural Network Techniques, IEEE Transactions On Industry Applications.

Mashud Rana, I. K., Vassilios G. Agelidis (2016). "Univariate and multivariate methods for very short-term solar photovoltaic power forecasting." Energy Conversion and

Management **121**: 380-390.

Mellit, A. P., A.M.; Lugh, V. "Short-term forecasting of power production in a large-scale photovoltaic plant." Sol. Energy **2014** **105**: 401-413.

Nam, S. H., J. "Probabilistic Forecasting Model of Solar Power Outputs Based on the Naive Bayes Classifier and Kriging Models." Energies **11**, **2018** **2982**: 11.

Pierro, M. B., F.; Felice, M.D.; Maggioni, E., Moser, D.; Perotto, A.; Spada, F.; Cornaro, C. "Multi-model ensemble for day ahead prediction of photovoltaic power generation." Solar Energy **2016** **134**: 132-146.

Porter, K. F., S.; Buvkley, M.; Rogers, J.; Hodge, B.M. Review of Variable Generation Integration Charges, National Renewable Energy Laboratory: Golden, CO, USA.

Reikard, G. H., S.E.; Jensen, T. "Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models." Renewable Energy **2017** **112**: 474-485.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

S. Atique, S. N., V. Roy, S. Bayne and J. Macfie Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques. 2020 IEEE Green Technologies Conference(GreenTech). Oklahoma City, OK, USA, 2020: 175-180.

Srivastava, R. Prediction of Electricity Generation using Solar Radiation Forecasting Data. International Conference on Electrical and Electronic Engineering (ICE3-2020). MMMUT GORKHPUR, INDIA.

Sun, Y. V., V.; Brandt, A.R. Convolutional neural network for short-term solar panel output prediction. IEEE 7th World Conference Photovoltaic Energy Conversion (WCPEC) (Joint Conf. 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC). Waikoloa, HI, USA: 2357-2361.

Vennila, C. T., A.; Sudha, T.S.; Sreenivasulu, U.; Reddy, N.P.R.; Jamal, K.; Lakshmaiah, D.; Jahadeesh, P.; Belay, A. "Forecasting Solar Energy Production Using Machine Learning." Hidawi International Journal of Photoenergy **2022**(7797488): 7.

villegas-Mier, C. G., Rodriguez-Resendiz, J., Álvarez-Alvarado ,J.M., Jiménez-Hernández, H., Odry, Á. "Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours." Micromachines **13**, 2022 **9**(1406).

Wai, R. J., Pin X.L. "Design of Intelligent Solar PV Power Generation Forecasting Mechanism Combined with Weather Information under Lack of Real-Time Power Generation Data." Energies **2022** **15**(10)(3838).

Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., Shi, M. "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework." Energy Conversion and Management **2020** **212**(112766).

Wang, G. S., Y.; Shu, L. "One-Day Ahead Daily Power Forecasting of Photovoltaic Systems Based on Partial Functional Linear Regression Models." Renewable Energy **2016** **96**: 469-478.

Yu, Y. C., J.; Zhu, J. "An LSTM short-term solar irradiance forecasting under complicated weather conditions. ." IEEE Access **2019** **7**: 145651-145666.

Zazoum, B. Solar Photovoltaic Power Prediction Using Different Machine Learning Methods. 8th International Conference on Power and Energy Systems Engineering (CPESE 2021): 19-25.





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

| | |
|-------------------|--|
| ชื่อ-สกุล | Phoomiphat Hongthanakornhirun |
| วัน เดือน ปี เกิด | 07 July 1999 |
| สถานที่เกิด | Bangkok, Thailand |
| ที่อยู่ปัจจุบัน | 222/231 The coast Tower A Sukhumvit Rd. Bang na Tai Sub-district Bangna District Bangkok 10260 |



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY