

การพยากรณ์การใช้พลังงานไฟฟ้าของอาคารสำนักงาน
ด้วยแบบจำลองการขับเคลื่อนด้วยข้อมูล

FORECASTING ELECTRICITY CONSUMPTION OF AN OFFICE
BUILDING USING DATA-DRIVEN MODELING

ภาณุเดช คุณชม¹

ดร.สมหญิง งามพรประเสริฐ²

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อศึกษาการสร้างรูปแบบการพยากรณ์การใช้พลังงานไฟฟ้าของอาคารสำนักงานด้วยการทำเหมืองข้อมูล โดยทำการรวบรวมข้อมูลจากการใช้พลังงานไฟฟ้า(kW)และข้อมูลค่าหน่วยในหนึ่งชั่วโมง(h)จำนวนทั้งหมด 2176 ชั่วโมง มาประยุกต์ใช้เทคนิคอัลกอริทึมในการทำเหมืองข้อความเพื่อวิเคราะห์ปัญหาทางด้านการจัดการพลังงานไฟฟ้า ผ่านทางโปรแกรม RapidMiner Studio โดยแบ่งข้อมูลออกเป็น 3 ชุดตามมิเตอร์ไฟฟ้าภายในอาคารสำนักงาน เพื่อนำมาสร้างรูปแบบการพยากรณ์ที่เหมาะสมกับตัวชุดข้อมูล โดยใช้ตัวแบบการพยากรณ์ Naïve Bayes, Generalized linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest, Gradient Booster Trees และ Support Vector Machine เพื่อหาตัวแบบที่เหมาะสมกับชุดข้อมูลมิเตอร์ไฟฟ้า โดยผลของการศึกษานี้พบว่า มิเตอร์ไฟฟ้าตัวที่1และมิเตอร์ไฟฟ้าตัวที่2 ตัวแบบการพยากรณ์ที่เหมาะสม ได้แก่ Gradient Booster Trees ส่วนมิเตอร์ไฟฟ้าตัวที่3 ตัวแบบการพยากรณ์ที่เหมาะสม ได้แก่ Random Forest ซึ่งตัวแบบของแต่ละมิเตอร์ไฟฟ้ามีค่าAccuracy และ Precision ที่มากที่สุด, มีค่า Classification error น้อยที่สุด และ MAPE เทียบกับข้อมูลจริงน้อยที่สุด

คำสำคัญ : พลังงานไฟฟ้า, ตัวแบบการพยากรณ์, เหมืองข้อมูล

¹ นักศึกษาหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาการจัดการทางวิศวกรรม มหาวิทยาลัยธุรกิจบัณฑิต

² ที่ปรึกษาการศึกษาวิทยานิพนธ์

ABSTRACT

This research aims to study the construction of a forecasting model for office building electricity consumption through data mining. It collects data from electric power consumption (kW) and the unit value in one hour (h) for 2176 hours to apply text mining algorithm techniques to analyze problems in power management. The RapidMiner Studio program divides the data into three sets according to the electricity meter in the office building. To create a forecasting model that is appropriate for the data set. This study used Naïve Bayes forecasting models, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forests, Gradient-Boosted Trees, and Support Vector machines to find models that fit the power meter data set. This study found suitable forecast models for electricity meters 1 and 2, including Gradient Booster Trees. As for the third electricity meter, the most suitable forecasting model is Random Forest, in which the model of each electricity meter has the highest accuracy and precision, the least classification error, and the least MAPE compared to actual data.

Keywords: electrical energy, forecasting models, data mining

1. บทนำ

ปัจจุบันการใช้พลังงานไฟฟ้ากำลังเป็นประเด็นสำคัญอย่างมากของประเทศไทย เป็นหนึ่งในปัจจัยสำคัญสำหรับการดำรงชีวิต โดยความต้องการของพลังงานไฟฟ้านั้น ในการผลิตไฟฟ้า ถือเป็นต้นทุนสำคัญของอาคารพาณิชย์ต่างๆ และครัวเรือน ซึ่งมีการคาดการณ์ถึงแนวโน้ม การใช้พลังงานไฟฟ้าที่สูงขึ้น และลดลงผันผวนไม่แน่นอนในแต่ละเดือนเป็นหนึ่งในปัจจัยสำคัญในการดำรงชีวิต โดยความต้องการพลังงานไฟฟ้าเป็นต้นทุนสำคัญต่อธุรกิจเป็นอย่างมาก โดยเฉพาะการใช้ไฟฟ้าภายในอาคารสำนักงานขนาดกลาง จะมีระบบและการใช้งานอุปกรณ์ที่ใช้พลังงานไฟฟ้าเข้ามาเกี่ยวข้อง ในปัจจุบันอาคารสำนักงานขนาดกลาง มีแนวโน้มในการใช้พลังงานที่มีความไม่แน่นอน ก่อให้เกิดปัญหาทางด้านการจัดการพลังงาน และใช้จ่ายที่สูงจากการจัดการพลังงานที่ไม่แน่นอนขึ้นตามมา จึงทำให้อาคารสำนักงานขนาดกลางต่างๆ เริ่มที่จะมีการนำโครงการที่เกี่ยวกับการควบคุมและการจัดการความต้องการพลังงานไฟฟ้าได้อย่างมีประสิทธิภาพ โดยในปัจจุบันก็มีเทคโนโลยี และแนวทางที่จะมาช่วยในการจัดการพลังงานไฟฟ้าอยู่มากมาย ซึ่งหนึ่งในวิธีการจัดการพลังงานไฟฟ้าภายในอาคารสำนักงานขนาดกลางนั้นคือ การเก็บข้อมูล ข้อมูลที่กล่าวมานั้นหมายถึงข้อมูลแสดงความต้องการไฟฟ้าในแต่ละช่วงเวลา ซึ่งจะแสดงพฤติกรรมการใช้ไฟฟ้าภายในอาคาร เพื่อที่จะนำข้อมูลดังกล่าวมาวางแผนในการจัดการพลังงานไฟฟ้าภายในตัวอาคารสำนักงาน เพิ่มความมั่นคงให้กับระบบไฟฟ้า สามารถลดต้นทุนที่ต้องจ่ายไปกับการใช้พลังงานไฟฟ้าได้ ทั้งนี้ผู้ศึกษา

สารนิพนธ์มีแนวคิดการจัดการพลังงานไฟฟ้าจากชุดข้อมูลที่เก็บข้อมูลความต้องการพลังงานไฟฟ้า โดยสังเกตเห็นความเปลี่ยนแปลงของการใช้พลังงานไฟฟ้าในแต่ละช่วงเวลา จึง หรือที่เรียกว่า ข้อมูลขนาดใหญ่ (Big Data) ข้อมูลที่ได้จากการเก็บข้อมูลตามที่ได้กล่าวมา จะทำการนำข้อมูลมาทำเหมืองข้อมูล (Data Mining) เพื่อวิเคราะห์ข้อมูล และหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ เช่น การจำแนกข้อมูล หรือการเชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน เป็นต้น หลังจากนั้นจะทำการพยากรณ์ความต้องการไฟฟ้าภายในอาคาร สำนักงานขนาดกลาง ค่าพยากรณ์ปริมาณความต้องการไฟฟ้าที่ได้มีความสำคัญในการวางแผนการจัดการพลังงานไฟฟ้า เพราะจะสามารถนำข้อมูลไปใช้ดูแนวโน้มความเสี่ยงการใช้ไฟฟ้าที่ไม่แน่นอน จะทำให้ทราบถึงปริมาณความต้องการพลังงานไฟฟ้าล่วงหน้าได้ และสามารถวางแผนสำหรับการตัดสินใจในอนาคต

2. วัตถุประสงค์ของการศึกษาวิจัย

1. เพื่อศึกษาและวิเคราะห์แนวทางการสร้างตัวแบบที่เหมาะสมในการพยากรณ์ชุดข้อมูลการใช้พลังงานไฟฟ้าภายในอาคารสำนักงานขนาดกลาง
2. เพื่อศึกษาวิธีการสร้างเหมืองข้อมูลและการวัดประสิทธิภาพของชุดข้อมูลที่พยากรณ์

3. ขอบเขตของงานวิจัย

1. การใช้พลังงานรายชั่วโมง ตั้งแต่วันที่ 1 กันยายน 2565 จนถึง 30 พฤศจิกายน 2565 ตลอดระยะเวลา 2,184 ชั่วโมง ของอาคารสำนักงานขนาดกลางแห่งหนึ่ง โดยวิเคราะห์ข้อมูลที่ได้จากระบบ บริหาร จัดการพลังงาน และความแตกต่างของการใช้พลังงานของแต่ละชั่วโมง
2. การศึกษาการวิเคราะห์ข้อมูลการใช้พลังงานไฟฟ้าภายในอาคารสำนักงานขนาดกลาง ถูกเก็บข้อมูล โดยการใช้โปรแกรม SCADA (Supervisory Control and Data Acquisition) โดยข้อมูลที่นำมาใช้นั้นจะเป็นค่าความต้องการไฟฟ้ามีหน่วยเป็นกิโลวัตต์ (kW) ทุกชั่วโมง ซึ่งเป็นโปรแกรมสำหรับติดต่อสื่อสารกับอุปกรณ์ ควบคุมต่าง ๆ เพื่อรับข้อมูลประมวลผล ซึ่งเก็บข้อมูลการใช้ไฟฟ้าทั้ง 24 ชั่วโมง แต่สำหรับกลุ่มผู้ใช้อื่น ๆ เช่น บ้านพักอาศัย หรือ อพาร์ทเมนต์ อาจจะมีผลกระทบของการใช้พลังงานไฟฟ้าที่แตกต่างกัน ขึ้นอยู่กับขนาด ของตัวอาคารหรือ พฤติกรรมการใช้พลังงานไฟฟ้า
3. การศึกษาเครื่องมือที่ใช้ในการหาตัวแบบการพยากรณ์ที่เหมาะสม คือโปรแกรม RapidMiner ซึ่งจะเป็นการใช้ตัวแบบทั้งหมด 8 ตัวแบบ ได้แก่ Naive Bayes, Generalized linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Tree, Random Forest , Gradient Booster Trees, Support Vector Machine

4. ประโยชน์ที่คาดว่าจะได้รับ

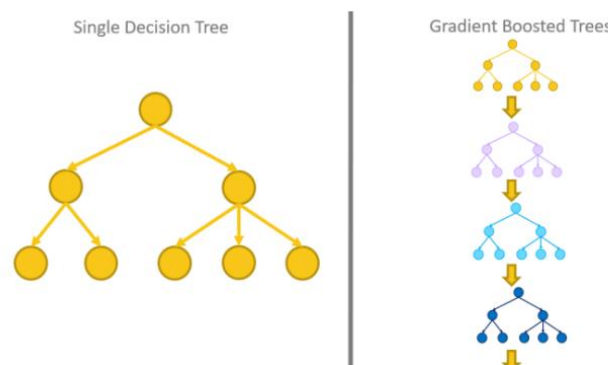
1. เพื่อที่จะนำมาประยุกต์ใช้กับ ชุดข้อมูลการใช้พลังงานไฟฟ้าภายในอาคารสำนักงานขนาดกลาง เพื่อสนับสนุนแผนการจัดการพลังงานไฟฟ้าของอาคารสำนักงานขนาดกลาง ที่มีการเก็บชุดข้อมูลการใช้พลังงานไฟฟ้าภายในอาคารสำนักงานรูปแบบเดียวกันได้
2. เพื่อเป็นแนวทางเกี่ยวกับการจัดการพลังงานไฟฟ้าภายในอาคารสำนักงานขนาดกลาง ด้วยวิธีการพยากรณ์ข้อมูลความต้องการไฟฟ้าในแต่ละช่วงเวลา เพื่อไปใช้ในการวางแผนการจัดการพลังงานไฟฟ้าในอนาคตต่อไปได้

5. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

5.1 ทฤษฎี

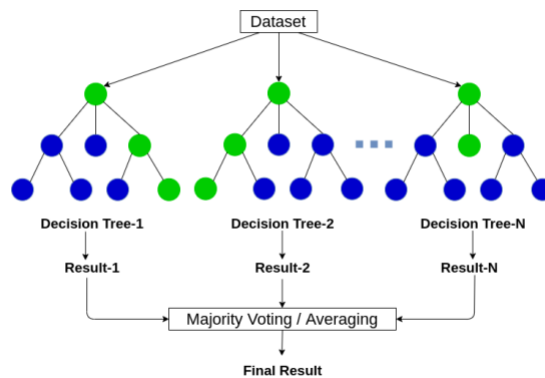
5.1.1 เหมืองข้อมูล (Data Mining) คือการนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) วิธีการทางสถิติ (Statistical Methods) วิธีการปัญญาประดิษฐ์ (Artificial Intelligence) หรือวิธีอื่น ๆ มาทำการวิเคราะห์และสกัดความรู้จากข้อมูลที่จัดเก็บไว้ในฐานข้อมูลหรือจัดเก็บไว้ในรูปอื่น ๆ โดยจุดประสงค์ของการทำเหมืองข้อมูล คือ การวิเคราะห์แนวโน้ม ความสัมพันธ์กฎ หรือรูปแบบของข้อมูล ซึ่งเป็นความรู้ที่ถูกซ่อนอยู่ภายใต้ข้อมูลขนาดใหญ่ และนำสารสนเทศที่ได้มาช่วยในการวางแผนการตัดสินใจ การบริหารหรือแก้ปัญหาในด้านต่าง ๆ ซึ่งถือได้ว่าเป็นเครื่องมือที่ช่วยเพิ่มคุณค่าให้กับข้อมูลที่มีอยู่ เหมืองข้อมูลจะสามารถแก้ปัญหาได้ บางปัญหาตามเทคนิควิธีการที่เลือกใช้ได้เท่านั้น โดยประโยชน์หลักของเหมืองข้อมูล คือ การค้นหาความรู้ที่ซ่อนอยู่ในฐานข้อมูล เพื่อให้ได้ซึ่งความรู้มาช่วยประกอบในการตัดสินใจ

5.1.2 Gradient Boosted Trees เป็นแบบจำลองที่มีพื้นฐานมาจาก Decision tree ซึ่งเป็นการปรับปรุงประสิทธิภาพของ แบบจำลองใหม่ค่าสูงขึ้น โดยการผสมสร้าง Decision tree หลายร้อยแบบจำลอง ทำซ้ำ และประเมินผลแต่ละแบบจำลองจนกว่า จะได้ Decision tree ที่สมบูรณ์



ภาพที่ 1 แสดงกระบวนการทำงานของ Gradient Boosted Trees

5.1.3 Random Forest คือ การสร้างแบบจำลองจาก Decision Tree หลายแบบจำลองย่อย โดยให้แต่ละแบบจำลองได้รับชุดข้อมูลย่อยที่ไม่เหมือนกันซึ่งข้อมูลนี้เป็นส่วนหนึ่งของชุดข้อมูลทั้งหมด แล้วทำการพยากรณ์ ข้อมูลในแต่ละ Decision Tree และคำนวณผลด้วยการ vote ที่ถูกเลือกโดย Decision Tree มากที่สุด



ภาพที่ 2 แสดงกระบวนการทำงานของ Random Forest

5.1.4 การวัดประสิทธิภาพของโมเดลแบบจำลองการพัฒนาระบบนั้น การประเมินผลการทำงานเป็นส่วนสำคัญ เพื่อให้ผู้พัฒนาได้ทราบประสิทธิภาพการทำงานของระบบว่ามีจุดเด่นจุดด้อยในด้านใดบ้าง เพื่อนำไปปรับปรุงระบบงานให้มีประสิทธิภาพมากยิ่งขึ้น ในงานด้านคั่นคั่นของสารสนเทศมีการวัดในด้านประสิทธิภาพของผลลัพธ์ที่ได้จากการสืบค้น ได้ด้วยค่าดังนี้

- การวัดค่าความถูกต้อง (Accuracy) คือ การคำนวณจำนวนคำตอบที่ถูกต้องทั้งหมดเทียบกับจำนวนคำตอบทั้งหมดที่นำไปให้แบบจำลองทำนายว่าผลการทำนายถูกต้องเท่าใด
- ค่าความแม่นยำ (Precision) คือ การคำนวณเทียบกับคำตอบทั้งหมดที่ผิด ว่าผลการทำนายถูกต้องเท่าใด

5.1.5 การวัดความผิดพลาดเป็นเปอร์เซ็นต์ (Mean Absolute Percentage Error :MAPE) คือ การวัดความผิดพลาดเป็นเปอร์เซ็นต์ซึ่งจะสามารถวิเคราะห์ความหมายได้ง่าย อีกทั้งรวมผลพยากรณ์ทั้งเกินจริงและต่ำกว่าค่าจริง

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{e_t}{Y_t} \right|}{n} \times 100 = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \times 100$$

โดยที่ Y_t คือ ค่าสังเกตของอนุกรมเวลา ณ เวลา t
 $\hat{Y}_t$ คือ ค่าพยากรณ์ของอนุกรมเวลา ณ เวลา t

e_t คือ ค่าความผิดพลาดของพยากรณ์ ณ เวลา t

n คือ ค่าจำนวนข้อมูลที่สนใจ

5.2 งานวิจัยที่เกี่ยวข้อง

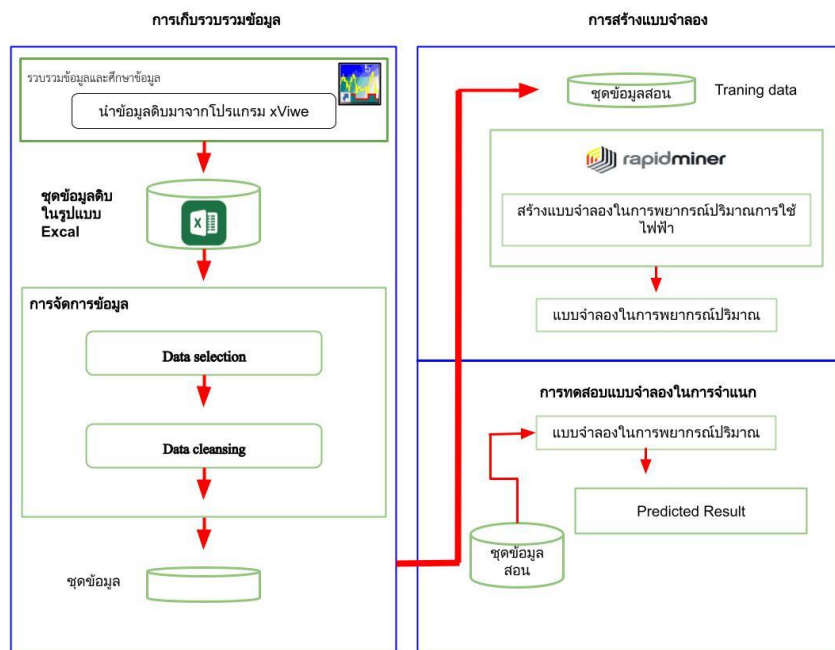
อัครฐ บรรจงศิลป์ (2550) ศึกษาการหารูปแบบการพยากรณ์ (Forecasting model) ที่เหมาะสมกับโรงงานอุตสาหกรรมแต่ละประเภท งานวิจัยที่ศึกษานั้น มีวัตถุประสงค์ในการหาตัวแบบที่เหมาะสมในการพยากรณ์ และ ข้อมูลที่ใช้ในการพยากรณ์เป็นประเภทปริมาณความต้องการไฟฟ้าคล้ายๆกัน จึงศึกษาการเตรียมและการจัดการข้อมูลจากงานวิจัยที่ศึกษานี้ จุดที่แตกต่างนั้นคือ งานวิจัยนี้ใช้ตัวแบบ Trend Analysis, Time Series Decomposition, Moving Average 3 เดือน, Moving Average 4 เดือน, Single Exponential Smoothing, Double Exponential Smoothing, Winter's Method และ งานวิจัยได้มีการวัดค่าผิดพลาดจากการพยากรณ์ที่เหมือนกัน

ธนวัฒน์ วรรณประภางานวิจัยที่ศึกษานั้น มีเครื่องมือที่ใช้ในการทำเหมืองข้อมูล โดยใช้โปรแกรม Rapid minerเหมือนกัน ผู้วิจัยได้รับองค์ความรู้ เกี่ยวกับวิธีการทำเหมืองข้อมูล และขั้นตอนการใช้โปรแกรม Rapid miner และได้ศึกษาเกี่ยวกับการแบ่งชุดข้อมูลสำหรับการสอนและทดสอบด้วยการแบ่งอัตราส่วน ด้วยวิธีการ Backward Selection และ Stepwise Selection ความแตกต่างกันคือ ชุดข้อมูลที่นำมาพยากรณ์นั้นเป็นข้อมูลโอกาสในการสำเร็จการศึกษา

วริศรา เจริญกิจสุพัฒน์ และ เอกสิทธิ์ พัทธวงศ์ศักดิ์ ได้ทำงานวิจัยเรื่องการพยากรณ์การเกิดภัยแล้งในประเทศไทยด้วยวิธีการเรียนรู้ของเครื่องเพื่อพยากรณ์การเกิดภัยแล้งของแต่ละตำบลในประเทศไทย และ หาแบบจำลองที่ดีที่สุดที่ใช้ในการพยากรณ์การเกิดภัยแล้ง งานวิจัยที่ศึกษานั้น มีเครื่องมือที่ใช้ในการทำเหมืองข้อมูล โดยใช้โปรแกรม Rapid minerเหมือนกัน ผู้วิจัยได้รับองค์ความรู้ เกี่ยวกับตัวแบบในการพยากรณ์ ซึ่งตัวแบบที่ศึกษามามีดังนี้ Neural Network, Gradient Boosted Trees และ Random Forest ความแตกต่างกันของงานวิจัยคือ ชุดข้อมูลที่นำมาพยากรณ์นั้นเป็นข้อมูลที่เรียนรู้จากข้อมูลปริมาณน้ำฝน เพื่อพยากรณ์อัตราการเกิดภัยแล้ง

6. วิธีดำเนินงานวิจัย

งานวิจัยนี้ได้นำเทคนิคเหมืองข้อมูลมาใช้ในการแบ่งแยกข้อมูล โดยการนำข้อมูลที่ได้จากการเก็บข้อมูลจะโปรแกรม xViwe ซึ่งเป็นโปรแกรม SCADA Software โดยที่เราได้นำข้อมูล มาทำเหมืองข้อมูลเพื่อวิเคราะห์ การใช้พลังงานไฟฟ้าในแต่ละช่วงเวลา ของมิเตอร์ภายในอาคารกรณีศึกษาโดยที่เราจะนำข้อมูลที่ได้จากมิเตอร์ไฟฟ้าทั้งหมดจำนวน 8 ตัว โดยที่แต่ละตัวจะมีการเก็บข้อมูลทุกๆ ชั่วโมงเป็นเวลา 24 ชั่วโมง หลังจากนั้นเราจะนำข้อมูลปริมาณการใช้พลังงานไฟฟ้าซึ่งมีหน่วยเป็น(kW) นำมาจัดการข้อมูลเพื่อเข้าโปรแกรม Rapid miner และให้โปรแกรม Rapid miner หาตัวแบบที่เหมาะสมในการพยากรณ์ปริมาณการใช้ไฟฟ้าของแต่ละมิเตอร์ไฟฟ้า



ภาพที่ 3 แผนผังขั้นตอนในการดำเนินงาน

6.1 รวบรวมข้อมูลและศึกษาข้อมูล ที่ได้รับการเก็บข้อมูลจากโปรแกรม xViwe โดยที่โปรแกรม xViwe นั้นมีการเก็บข้อมูลจะมีเตอร์ไฟฟ้าทุกช่วงเวลา โดยข้อมูล จะนำออกมาในรูปแบบ excel โดยที่จะจัดเลือกอาคารตัวอย่างกรณีศึกษา โดยที่ข้อมูลจะมีให้เลือกจะให้เป็นรีพอร์ทรายเดือน รายสัปดาห์ รายวัน รายชั่วโมง รายนาที

ภาพที่ 4 ตัวอย่าง หน้า UI ของ ตัวโปรแกรมที่เก็บข้อมูล

6.2 วิเคราะห์และเลือกข้อมูล Data selection ทำชุดข้อมูลในรูปแบบของไฟล์ Excel เมื่อเราได้ข้อมูลเป็นไฟล์ Excel ออกมาแล้วดังภาพตัวอย่าง เราก็จะนำข้อมูลมาวิเคราะห์ปริมาณที่จะนำมาพยากรณ์

ชุดข้อมูลจะมีประเภทที่เก็บข้อมูลมาได้คือจะมีจำนวนข้อมูล ทั้งหมด มีประเภทข้อมูลที่เก็บมาได้ นั้นมีทั้งหมด 23 ประเภทได้แก่ Date Time(ช่วงเวลารายชั่วโมง), kW (ค่าการใช้ไฟฟ้าในหนึ่งชั่วโมง), kWh(ค่าการใช้พลังงานทั้งหมด), Voltage L1 L2 L3 (ค่าแรงดันไฟฟ้า ของแต่ละเฟส), Current L1 L2 L3 (ค่ากระแสไฟฟ้า ของแต่ละเฟส), kW L1 L2 L3 (ค่ากำลังไฟฟ้า ในแต่ละเฟส), kVA L1 L2 L3(หน่วยของกำลังไฟฟ้าที่ปรากฏ ในแต่ละเฟส), kvar L1 L2 L3(กำลังไฟฟ้าแฝง ในแต่ละเฟส), PF L1 L2 L3 (ค่าพาวเวอร์แฟคเตอร์ ในแต่ละเฟส), Hz (ค่าความถี่ไฟฟ้าต่อวินาที)และMax Demand (ค่าความต้องการไฟฟ้าสูงสุด)

6.3 กลั่นกรองข้อมูล (Data cleansing)หลังจากได้ทำการกลั่นกรองข้อมูลเพื่อให้ได้ข้อมูลที่มีความสมบูรณ์สำหรับการประมวลผลต่อไป ทั้งนี้ข้อมูลที่นำมาวิเคราะห์ไม่ว่าจะเป็นตัวเลข ข้อมูลเป็นช่วงหรือข้อมูลที่เป็นข้อความ จะต้องทำการแปลงหรือแทนค่าให้อยู่ในรูปแบบมาตรฐานเดียวกัน เพื่อให้ง่ายต่อการวิเคราะห์ การแทนค่าข้อมูลที่ถูกต้องจะส่งให้การประมวลผลของคอมพิวเตอร์ทำได้อย่างรวดเร็ว มีความถูกต้องแม่นยำในการวิเคราะห์ข้อมูล และลดความผิดพลาดในการพยากรณ์ตัดข้อมูลบางส่วนที่ไม่มีค่าเป็นออกไป และ คงเหลือข้อมูลที่เราจะใช้ โดยที่ข้อมูลที่เราจะใช้นั้นมีดังนี้

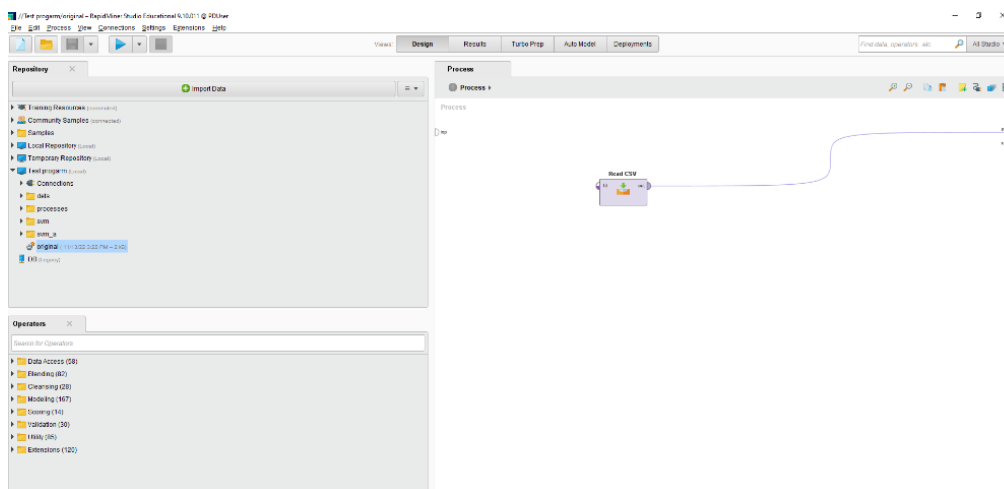
ตารางที่ 1 แสดงตามข้อมูลที่นำมาวิเคราะห์กลุ่มที่จะนำมาพยากรณ์

ตัวแปรที่นำมา	ความหมาย
Date Time	เป็นแถวที่บอกข้อมูลของวันที่และเวลาตลอดทั้งเดือน
kW	เป็นแถวที่บอกค่าความต้องการพลังงานไฟฟ้า ของแต่ละมิเตอร์ ในแต่ละชั่วโมง หน่วยเป็น กิโลวัตต์ (KW)
kWh	ค่าการใช้พลังงานทั้งหมด มีหน่วยเป็นกิโลวัตต์ต่อชั่วโมง
Max Demand	ค่าความต้องการไฟฟ้าสูงสุด

อาคารสำนักงานขนาดกลางที่เป็นกรณีศึกษานั้น ระบบไฟฟ้าทั้งอาคารจะประกอบไปด้วยมิเตอร์ไฟฟ้าจำนวนทั้งหมด 3 มิเตอร์ไฟฟ้า การจัดสรรข้อมูลที่เตรียมทำแบบพยากรณ์ จะทำการแบ่งแยกชุดข้อมูลค่าความต้องการการใช้พลังงานไฟฟ้า เป็นตัวแปร MDB1 MBD2 และ MDB3 ตามลำดับ

6.4 สร้างตัวแบบทดสอบเปรียบเทียบประสิทธิภาพและทำการพยากรณ์ข้อมูลการสร้างตัวแบบ (Model) ใช้เทคนิควิธีหรืออัลกอริทึมสำหรับการทดลอง โดยที่การที่เราจะเลือกอัลกอริทึมไหนจะใช้ฟังก์ชันหนึ่งของโปรแกรม Rapid miner เพื่อทำการเรียนรู้ Model และสร้างผลการคำนวณเพื่อหาอัลกอริทึมที่

เหมาะสม โดยมีวิธีการ ขั้นตอนแรกแปลงไฟล์ Data เป็นในรูปแบบ CSV โดยที่นำไฟล์มาใส่ใน operators ที่ชื่อว่า Read csv



ภาพที่ 5 แสดงการสร้าง operators Read CSV

ผู้วิจัยได้ใช้ฟังก์ชันของโปรแกรม Rapid miner ในคัดการเลือกอัลกอริทึม สำหรับการพยากรณ์ ข้อมูลโดยโปรแกรมจะทำการเรียนรู้ชุดข้อมูลที่ไว้ใช้สำหรับการพยากรณ์และหาความสัมพันธ์กับตัวกลุ่ม หัวข้อภายในชุดข้อมูลนั้นๆ โดยจะใช้การวิเคราะห์ข้อมูลแบบถดถอย (Regression Analysis) แบ่งชุดข้อมูลที่ไว้ในการเรียนรู้กับส่วนที่ไว้ใช้พยากรณ์ในอัตราส่วน 60 : 40 โดยจะใช้ชุดข้อมูล 60% นั้นไว้สำหรับการเรียนรู้ชุดข้อมูลย้อนหลัง ส่วน ข้อมูลอีก 40% นั้นไว้สำหรับการพยากรณ์ เนื่องจากเราต้องการชุดข้อมูลในเดือนถัดไป ซึ่งจำนวนข้อมูลทั้งหมดคือ 2176 การแบ่งชุดการเรียนรู้ข้อมูล 60% เพื่อความแม่นยำจำนวนหนึ่ง นั้น 40% จะเป็นข้อมูลที่ไว้ใช้พยากรณ์จำนวนแถวจะเท่ากับจำนวนชั่วโมงในจำนวนเดือนถัดไปพอดี

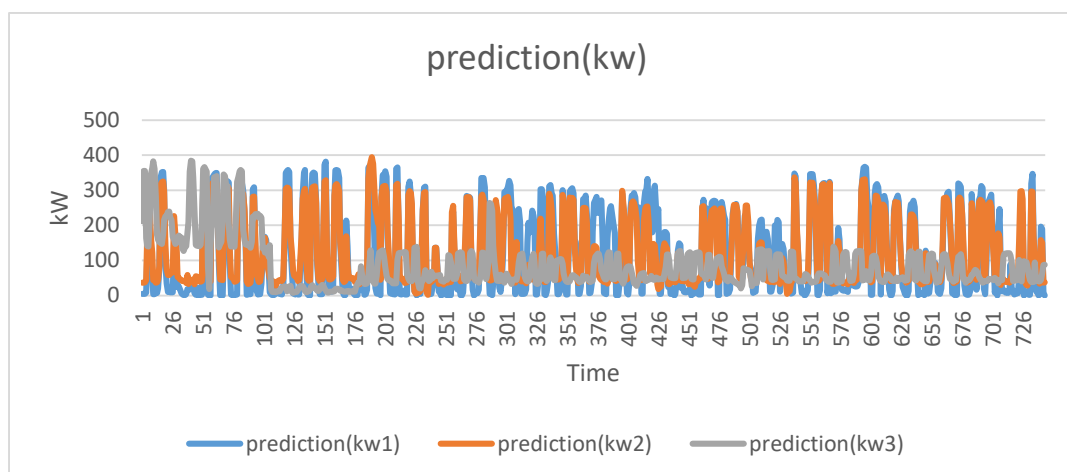
Row No.	Date Time	kW1	kW2	kW3
1	Sep 1, 2022, ...	6.379	43.129	135.339
2	Sep 1, 2022, ...	6.230	43.960	137.590
3	Sep 1, 2022, ...	6.200	42.910	135.910
4	Sep 1, 2022, ...	6.300	41.310	132.780
5	Sep 1, 2022, ...	1.270	42.960	133.130
6	Sep 1, 2022, ...	10.090	42.560	129.490
7	Sep 1, 2022, ...	56.760	74.120	190.430
8	Sep 1, 2022, ...	105.470	153.140	234.610
9	Sep 1, 2022, ...	295.670	291.100	379.600
10	Sep 1, 2022, ...	315.310	297.140	355.420
11	Sep 1, 2022, ...	343.200	316.070	356.550
12	Sep 1, 2022, ...	331.740	334.670	376.050
13	Sep 1, 2022, ...	329.500	336.350	366.320
14	Sep 1, 2022, ...	341.480	336.990	365.490
15	Sep 1, 2022, ...	215.950	299.630	346.360
16	Sep 1, 2022, ...	220.900	293.670	335.460
17	Sep 1, 2022, ...	205.830	242.580	325.190
18	Sep 1, 2022, ...	115.830	193.000	290.330
19	Sep 1, 2022, ...	43.900	195.190	225.150
20	Sep 1, 2022, ...	16.750	69.730	159.220
21	Sep 1, 2022, ...	8.840	48.260	146.050
22	Sep 1, 2022, ...	1.650	47.260	145.600
23	Sep 1, 2022, ...	6.370	39.900	141.490
24	Sep 1, 2022, ...	6.630	39.520	141.410
25	Sep 2, 2022, ...	6.440	42.160	142.930

ภาพที่ 6 ตัวอย่างแสดงฟังก์ชัน Auto model

จากการทดสอบการทำงานตัวแบบในการพยากรณ์หลายๆตัวพร้อมกัน ผู้ทำการวิจัยจะเลือกตัวแบบที่ได้จากโปรแกรมที่ทำการเลือกความเหมาะสมกับตัวชุดข้อมูลซึ่งมีทั้งหมด 8 ตัวแบบ ได้แก่ Naive Bayes, Generalized linear Model Fast Large Margin , Deep Learning , Decision Tree , Random Forest , Gradient Booster Trees และ Support Vector Machine ขึ้นต่อไปจะมาหาตัวแบบที่เหมาะสมกับชุดข้อมูลของเราโดยที่เราจะมาดูจากผลลัพธ์ในการจัดกลุ่มการพยากรณ์ตัวแบบ โดยจะดูจากค่าต่างๆที่ทางโปรแกรมได้แสดงผลประสิทธิภาพ ที่มีความสำคัญต่อการตัดสินใจเลือกตัวแบบในการพยากรณ์มีดังนี้

- ค่าAccuracy คือค่าความแม่นยำของตัวแบบ ซึ่งค่า Accuracy ยิ่งมากความแม่นยำในการพยากรณ์ยิ่งสูง
- ค่า classification error คือ ค่าการจัดกลุ่มผิดพลาด ซึ่งค่า classification error ยิ่งมีค่าน้อยความสมบูรณ์ของการพยากรณ์ก็ยิ่งมีประสิทธิภาพมากขึ้น
- ค่าความแม่นยำ (Precision) คือ การคำนวณเทียบกับคำตอบทั้งหมดที่ผิด ว่าผลการทำนายถูกต้องเท่าใด

จะพิจารณาประสิทธิภาพจากตัวข้อมูลที่พยากรณ์จากค่าทั้งสองค่านี้โดยที่ ค่า Accuracy จะต้องมีย่านค่ามากที่สุด และค่า classification error น้อยที่สุด และค่าค่าความแม่นยำ (Precision) มากที่สุด หลังจากที่เราได้ผลการพยากรณ์ โดยที่ผลการพยากรณ์นั้นจะได้ปริมาณการใช้ไฟฟ้าระยะเวลา 1 เดือนซึ่งข้อมูลพยากรณ์จะเป็นข้อมูลในเดือนถัดไป หลังจากนั้นให้ทำการนำข้อมูลออกมาเพื่อใช้งาน โดยทางโปรแกรมจะมีฟังก์ชันการ export ชุดข้อมูลพยากรณ์ ออกมาเป็นในรูปแบบ ไฟล์ Excel จะได้ข้อมูลดังภาพแผนภูมิปรมาณด้านล่าง



ภาพที่ 7 แสดงแผนภูมิปรมาณผลลัพธ์ชุดข้อมูลการพยากรณ์

จากชุดข้อมูลที่ได้จากการพยากรณ์ของชุดข้อมูลมิเตอร์ไฟฟ้าเปรียบเทียบ ข้อมูลพยากรณ์ กับ ตัวชุดข้อมูลจริง ต่อช่วงระยะเวลารายชั่วโมง จำนวนทั้งหมด 869 ชั่วโมง) ชุดข้อมูลที่ได้จากการพยากรณ์ จะต้องมีการวัดค่าความผิดพลาดเพื่อที่จะตรวจสอบผลการพยากรณ์เกินจริง (Over forecast) และค่าพยากรณ์ที่ต่ำกว่าความเป็นจริง (Under forecast) และเพื่อง่ายต่อการวิเคราะห์ได้ง่ายจึงใช้วิธีการวัดความผิดพลาดเป็น เปอร์เซ็นต์ มาพิจารณา โดยที่ถ้ามีค่า MAPE น้อยจะมีความถูกต้องมากที่สุด

7. ผลการศึกษา

ผลของการเปรียบเทียบแบบจำลองการทำนายปริมาณการใช้พลังที่เหมาะสมกับชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 1 จากการรันตัวแบบทั้ง 8 ตัวแบบที่ทางโปรแกรมได้ทำการคัดเลือกความเหมาะสมให้กับชุดข้อมูล มิเตอร์ไฟฟ้าตัวที่ 1 โดยตัวแบบที่ทำการแสดงผลนั้นพบว่าผลลัพธ์ค่าความสัมพันธ์ต่างๆ การตัดสินใจเลือกตัวแบบนั้นวัดค่าประสิทธิภาพหลักๆ จากค่ามีค่า Accuracy และ Precision มากที่สุดและมีค่า Classification error น้อยที่สุด และเปรียบเทียบ ข้อมูลพยากรณ์ กับ ตัวชุดข้อมูลจริง เพื่อง่ายต่อการ วิเคราะห์ได้ง่ายจึงใช้วิธีการวัดความผิดพลาดเป็นเปอร์เซ็นต์ มาพิจารณา โดยที่ถ้ามีค่า MAPE น้อยจะมีความถูกต้องมากที่สุด ผลงานวิจัยออกมาดังตารางด้านล่างนี้

ตารางที่ 2 แสดงผลลัพธ์ในการเปรียบเทียบค่าประสิทธิภาพต่างๆทั้งหมดของชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 1

model	Accuracy	Classification error	Precision	MAPE	total Time
Naive Bayes	69.30%	30.70%	69.30%	40.70%	10 s
Generalized linear	69.30%	30.70%	69.80%	50.70%	12 s
Fast Large Margin	69.30%	30.70%	69.30%	40.70%	12 s
Deep Learning	76.40%	23.60%	74.60%	29.60%	14 s
Decision Tree Model	91.00%	9.00%	89.30%	11.00%	8 s
Random Forest	90%	10%	89.40%	10%	1 min 17 s
Gradient Booster Trees	91.60%	8.40%	91.10%	7.40%	34 s
Support Vector Machine	69.30%	30.70%	69.30%	40.70%	33 s

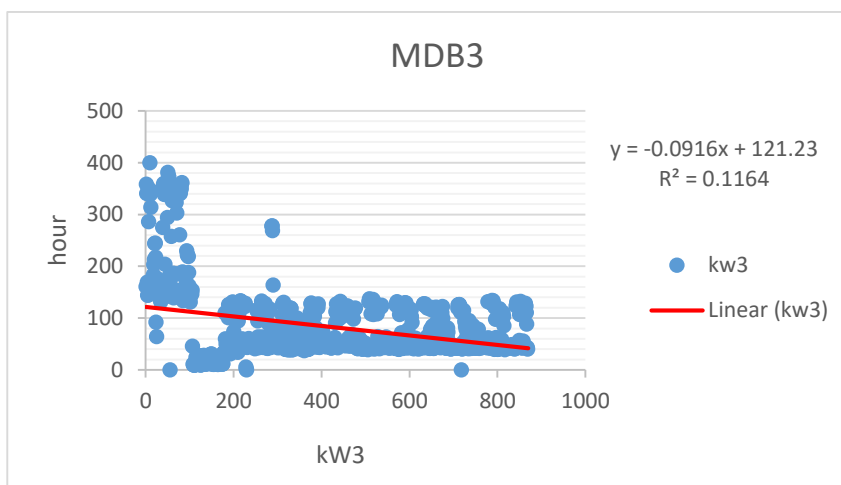
ตารางที่ 3 แสดงผลลัพธ์ในการเปรียบเทียบค่าประสิทธิภาพต่างๆทั้งหมดของชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 2

model	Accuracy	Classification error	Precision	MAPE	total Time
Naive Bayes	73.60%	26.40%	73.60%	46.40%	11 s
Generalized linear	63.30%	36.70%	70.60%	36.70%	12 s
Fast Large Margin	65.10%	34.90%	71.70%	35.90%	11 s
Deep Learning	79.30%	20.70%	78.40%	30.70%	14 s
Decision Tree Model	93.40%	6.60%	97.80%	6.60%	8 s
Random Forest	92.5%	7.50%	92.80%	7.50%	1 min 40 s
Gradient Booster Trees	95.80%	4.20%	97.20%	5.20%	33 s
Support Vector Machine	73.60%	26.40%	73.60%	46.40%	30 s

ตารางที่ 4 แสดงผลลัพธ์ในการเปรียบเทียบค่าประสิทธิภาพต่างๆทั้งหมดของชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 3

model	Accuracy	Classification error	Precision	MAPE	total Time
Naive Bayes	95.60%	4.80%	95.20%	5.80%	8 s
Generalized linear	95.30%	4.70%	96.20%	5.70%	7 s
Fast Large Margin	93.90%	6.10%	97.20%	6.10%	9 s
Deep Learning	97.30%	2.70%	98.30%	2.70%	12 s
Decision Tree Model	96.10%	3.90%	96.70%	3.90%	7 s
Random Forest	98.9%	1.1%	99.0%	1.15%	44 s
Gradient Booster Trees	96.90%	2.70%	98.80%	2.70%	35 s
Support Vector Machine	95.20%	4.80%	95.20%	4.80%	16 s

ตัวแบบทั้งหมดมีค่าประสิทธิภาพที่มีความสมบูรณ์ใกล้เคียงกัน เนื่องจากชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 3 ที่มีค่าปริมาณของชุดข้อมูลเกาะกลุ่มกันดังภาพ จึงทำให้การพยากรณ์เชิงเส้นมีความแม่นยำมาก และเปอร์เซ็นต์ความผิดพลาดต่ำ



ภาพที่ 8 แสดงชุดข้อมูลมิเตอร์ไฟฟ้าตัวที่ 3 ที่นำมาพยากรณ์

- รูปแบบการพยากรณ์ที่เหมาะสมกับการพยากรณ์ข้อมูลการใช้ไฟฟ้าของมิเตอร์ตัวที่ 1 คือ ตัวแบบ Gradient Booster Trees เพราะมีค่าประสิทธิภาพจากตัวแบบมีความเหมาะสมมากที่สุด โดย ค่า Accuracy มีค่าเท่ากับ 91.60% ,ค่า Classification error มีค่าเท่ากับ 8.4 % ค่าความแม่นยำ (Precision) มีค่าเท่ากับ 91.10% และมีค่า ความคลาดเคลื่อนจากการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ (MAPE) น้อยที่สุด มีค่าเท่ากับ 7.40 %

- รูปแบบการพยากรณ์ที่เหมาะสมกับการพยากรณ์ข้อมูลการใช้ไฟฟ้าของมิเตอร์ตัวที่ 2 คือ ตัวแบบ Gradient Booster Trees เพราะมีค่าประสิทธิภาพจากตัวแบบมีความเหมาะสมมากที่สุด โดย ค่า Accuracy มีค่าเท่ากับ 95.80% ,ค่า Classification error มีค่าเท่ากับ 4.20% ค่าความแม่นยำ (Precision) มีค่าเท่ากับ 97.20% และมีค่า ความคลาดเคลื่อนจากการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ (MAPE) น้อยที่สุด มีค่าเท่ากับ 5.20%

- รูปแบบการพยากรณ์ที่เหมาะสมกับการพยากรณ์ข้อมูลการใช้ไฟฟ้าของมิเตอร์ตัวที่ 3 คือ ตัวแบบ Random Forest เพราะมีค่าประสิทธิภาพจากตัวแบบมีความเหมาะสมมากที่สุด โดย ค่า Accuracy มีค่าเท่ากับ 98.9% ,ค่า Classification error มีค่าเท่ากับ 1.1% ค่าความแม่นยำ (Precision) มีค่าเท่ากับ 99.0% และมีค่า ความคลาดเคลื่อนจากการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ (MAPE) น้อยที่สุด มีค่าเท่ากับ 1.15%

8. สรุปผลการวิจัย

งานวิจัยนี้จะเป็นการนำข้อมูลมาพยากรณ์ปริมาณการใช้ไฟฟ้าในเดือนถัดไป เพื่อแสดงให้เห็นทราบถึงปริมาณการใช้พลังงานไฟฟ้าและเห็นพฤติกรรมการใช้ไฟฟ้าในแต่ละชั่วโมงในเดือนถัดไปได้ โดยใช้ข้อมูลความต้องการไฟฟ้าในแต่ละชั่วโมง ที่เก็บข้อมูลมาจากระบบ SCADA Software ผ่านโปรแกรม xView ซึ่งข้อมูลความต้องการทางไฟฟ้าภายในสำนักงานประกอบด้วย ข้อมูลของมิเตอร์ไฟฟ้า จำนวนทั้งหมด 3 ตัว และจากนั้นนำข้อมูลที่ได้มาพยากรณ์ปริมาณ ด้วยใช้วิธีการทำเหมืองข้อมูลโดยใช้เครื่องมือโปรแกรมที่ชื่อว่า RapidMiner โดยจะแบ่งชุดข้อมูลที่จะนำมาใช้พยากรณ์แบ่งออกเป็น 3 ชุด ตามมิเตอร์ไฟฟ้าดังข้อมูลที่กล่าวไว้ข้างต้น โดยจะใช้ชื่อชุดข้อมูล MDB1, MDB2 และ MDB3 ตามลำดับ โดยตัวแปรที่จะนำมาใช้ในการเรียนรู้ข้อมูลมี 4 ตัวแปร ได้แก่ เวลา(Date Time), ความต้องการพลังงานไฟฟ้า (kw), ค่าการใช้พลังงานไฟฟ้าทั้งหมด (kWh) และ ค่าความต้องการไฟฟ้าสูงสุด (Max Demand) เพื่อที่จะสร้างตัวแบบการพยากรณ์ที่มีความเหมาะสมกับชุดข้อมูลมากที่สุด โดยคัดเลือกจากตัวแบบทั้ง 8 ตัวแบบที่ทางโปรแกรมได้ประมวลผลออกมา ได้แก่ Naive Bayes, Generalized linear Model Fast Large Margin , Deep Learning , Decision Tree , Random Forest , Gradient Booster Trees และ Support Vector Machine และหาความผิดพลาดจากการพยากรณ์

จากการศึกษาพบว่า การพยากรณ์ปริมาณการใช้พลังงานไฟฟ้าภายในอาคารสำนักงาน โดยการศึกษาครั้งนี้ได้เลือกตัวแบบการพยากรณ์เป็นจำนวน 8 ตัวแบบ โดยในการเลือกตัวแบบพยากรณ์ที่เหมาะสมนั้นจะวัดจากค่าประสิทธิภาพที่สำคัญทั้งหมด 3 ตัว คือ Accuracy, Classification error, Precision และความผิดพลาดเป็นเปอร์เซ็นต์(Mean Absolute Error :MAPE) ผลลัพธ์ของการหาตัวแบบพยากรณ์ที่เหมาะสมกับชุดข้อมูลของแต่ละมิเตอร์ไฟฟ้าตัวแบบที่เหมาะสมสำหรับการพยากรณ์ที่เหมาะสมกับชุดข้อมูลแต่ละมิเตอร์ไฟฟ้าแต่ละตัวนั้นได้ข้อมูลดังนี้ มิเตอร์ไฟฟ้าตัวที่ 1 และ มิเตอร์ไฟฟ้าตัวที่ 2 ตัวแบบการพยากรณ์ที่เหมาะสม ได้แก่ Gradient Booster Trees ส่วนมิเตอร์ไฟฟ้าตัวที่ 3 ตัวแบบการพยากรณ์ที่เหมาะสม ได้แก่ Random Forest ซึ่งการที่มิเตอร์ไฟฟ้าตัวที่ 3 ใช้ตัวแบบที่แตกต่างกันเพราะ ข้อมูลของมิเตอร์ไฟฟ้าตัวที่ 3 มีความเกาะกลุ่มมากกว่า มิเตอร์ไฟฟ้าตัวที่ 1 และ 2 ผลลัพธ์ของการเลือกตัวแบบการพยากรณ์เลยแตกต่างกันดังตารางด้านล่างนี้

ตารางที่ 5 แสดงการเปรียบเทียบค่า Accuracy, Precision, classification Error และ MAPE

performants	MDB1	MDB2	MDB3
Model	Gradient Booster Trees	Gradient Booster Trees	Random Forest
Accuracy	91.60%	95.80%	98.90%
Classification error	8.40%	4.20%	1.10%
Precision	91.10%	97.20%	99.00%
MAPE	7.4%	5.20%	1.15%

บรรณานุกรม

- นภัสสร ตั้งวรวิตร. (2562). การประยุกต์ใช้เทคนิคอัลกอริทึมในการทำเหมืองข้อความเพื่อ วิเคราะห์ทัศนคติของผู้ใช้ที่มีต่อการเลือกซื้อโคมกลางแจ้ง บน ทวิตเตอร์. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- วิรัชตรา เจริญกิจสุพัฒน์ และ เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2562). งานวิจัยเรื่องการพยากรณ์การเกิดภัยแล้งในประเทศไทยด้วยวิธีการเรียนรู้ของเครื่อง A Large-Scale Drought Forecasting with Machine Learning
- วุฒิพงษ์ นิลจันทร์ และ นิติ เอี่ยมชื่น. (2562). การคาดการณ์ปริมาณการใช้ไฟฟ้าภาคเหนือตอนบนของประเทศไทยหน่วยวิจัยเพื่อการพัฒนานวัตกรรมเชิงพื้นที่. สาขาวิชาภูมิสารสนเทศศาสตร์ คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยพะเยา.
- ชนะวัฒน์ วรรณประภา. (2564). เทคนิคการทำเหมืองข้อมูลมาใช้ในการคัดกรองบุคคลที่มีแนวโน้มสัมฤทธิ์ผลสำหรับการเรียนระดับปริญญาตรี. มหาวิทยาลัยธุรกิจบัณฑิตย์.
- คงฤทธิ โกมาสถิต. (2555). วิธีการพยากรณ์อุปสงค์พลังงานไฟฟ้า (Electrical Energy Consumption) ของประเทศไทยในระยะยาวด้วยวิธีโครงข่ายประสาทเทียม (Artificial Neural Network). คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- อัครัช บรรจงศิลป์. (2550). การพยากรณ์การใช้พลังงานไฟฟ้าในอุตสาหกรรมขนาดใหญ่. มหาวิทยาลัยศิลปากร.
- ธนิตเชษฐ คำศรีสุข. (2562). การวิเคราะห์ข้อมูลการใช้พลังงานไฟฟ้าภายในบ้านสำหรับบ้าน อัจฉริยะ ด้วยวิธีการเหมืองข้อมูล. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี.