การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก

วุฒิชัย กำจรกิตติคุณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม วิทยาลัยนวัตกรรมด้านเทคโนโลยี
และวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
พ.ศ. 2561

Electric Usage Classification using Naïve Bayes and Neural Networks

Wuttichai Kamchonkitticoon

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering

Department of Computer and Telecommunication Engineering

College of Innovative Technology And Engineering,

Dhurakij Pundit University

2018

หัวข้อวิทยานิพนธ์	การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และนิวรัลเน็ทเวิร์ก
ชื่อผู้เขียน	วุฒิชัย กำจรกิตติคุณ
อาจารย์ที่ปรึกษา	ดร. ชัยพร เขมะภาตะพันธ์
สาขาวิชา	วิศวกรรมคอมพิวเตอร์และโทรคมนาคม
ปีการศึกษา	2560

บทคัดย่อ

งานวิจัยนี้นำเสนอการจำแนกประเภทผู้ใช้ไฟฟ้าโดยอาศัยข้อมูลการใช้ไฟฟ้าที่ได้จากระบบอ่านหน่วยอัตโนมัติ (AMR) เพื่อแก้ปัญหาฐานข้อมูลประเภทกิจการที่ไม่เป็นปัจจุบัน โดยการสังเคราะห์ข้อมูลการใช้ไฟฟ้าเพื่อหารูปแบบประจำ (Attribute) ตามประเภทกิจการที่กำหนด ในการศึกษานี้ใช้ซอฟท์แวร์แรพิดไมเนอร์ สตูดิโอ (Rapid Miner Studio) เป็นเครื่องมือในการทดสอบ ซึ่งลักษณะประจำดังกล่าวถูกนำมาทดสอบด้วยอัลกอริทึมนาอีฟเบย์ (Naïve Bayes) และโครงข่ายประสาทเทียม (Neural Networks) สำหรับจำแนกผู้ใช้ไฟฟ้าออกตามประเภทกิจการที่ได้กำหนดไว้ จากผลการทดสอบพบว่า วิธีการที่นำเสนอนี้สามารถจำแนกข้อมูลตัวอย่างได้อย่างถูกต้อง โดยอัลกอริทึมนาอีฟเบย์และโครงข่ายประสาทเทียมให้ความแม่นยำที่ร้อยละ 81.25 และ 87.5 ตามลำดับ ส่งผลให้หน่วยงานผู้ให้บริการพลังงานไฟฟ้าอย่างการไฟฟ้าส่วนภูมิภาค (กฟภ.) และการไฟฟ้านครหลวง (กฟน.) สามารถจัดสรรรูปแบบการให้บริการไฟฟ้าที่เหมาะสมกับกิจการแต่ละประเภท ทั้งยังสามารถวิเคราะห์พฤติกรรมการบริโภคพลังงานของผู้ใช้ไฟฟ้าที่ไม่สอดคล้องกับประเภทกิจการที่ระบุไว้ตอนเริ่มต้นซึ่งนำไปสู่การตรวจจับการลักลอบใช้ไฟฟ้าได้อีกด้วย

Thesis Title	Electric Usage Classification using Naive Bayes and Neural Network
Author	Wuttichai Kamchonkitticoon
Thesis Advisor	Ph.D. Chaiyaporn Khemapatapan
Department	Computer and Telecommunication Engineering
Academic Year	2017

ABSTRACT

This research presents electric usage classification based on AMR data for solving the problem about outdated customer class. This study proposes how to find the attributes for identifying users according to the customer class. The attributes are used by Naïve Bayes and Neural Networks algorithms. Rapid Miner studio is used as an analysis tool. It can be noted from the results that Naïve Bayes and Neural network have classification accuracies at 81.25 and 87.5 percentages respectively. From the result, the electric providers such as Provincial Electricity Authority (PEA) and Metropolitan Electricity Authority (MEA) are able to allocate electricity service models to suiting each type of business. It can be also used to analyze the energy consumption behavior of consumers that are not in line with the type of business which leads to the detection of tampering.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยความกรุณาเป็นอย่างยิ่งจาก ดร.ชัยพร เขมะภาตะพันธ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่คอยให้คำแนะนำ ติดตามผล และชี้ให้เห็นมุมมองในมิติใหม่ๆ ให้แก่ผู้วิจัย ขอขอบพระคุณอาจารย์ ดร.ประศาสน์ จันทราทิพย์ ผศ.ดร.ณรงค์เดช กีรติพรานนท์ และ ดร. ธนัญ จารุวิทยโกวิท กรรมการสอบวิทยานิพนธ์ ซึ่งสละเวลามาเป็นกรรมการสอบวิทยานิพนธ์ และได้ให้ข้อคิดเห็นที่เป็นประโยชน์ต่องานวิจัย นอกจากนี้ผู้วิจัยขอขอบพระคุณคณาจารย์ผู้ช่วยที่ปรึกษาทุก ๆ ท่าน ที่ได้ถ่ายทอดความรู้แก่ผู้วิจัยตลอดระยะเวลาการศึกษา

ผู้วิจัยขอขอบพระคุณ กองพัฒนาระบบมิเตอร์ การไฟฟ้าส่วนภูมิภาคที่ให้การสนับสนุนข้อมูลสำหรับดำเนินการวิจัยในครั้งนี้

ผู้วิจัยขอขอบพระคุณ เจ้าหน้าที่ที่เกี่ยวข้องทุกท่าน ในคณะวิศวกรรมศาสตร์ ที่คอยให้ความช่วยเหลือ ตลอดจนแนะนำกระบวนการในการทำงานให้แก่ผู้วิจัยด้วยดีเสมอมา

ผู้วิจัยขอขอบคุณเพื่อนๆ ร่วมรุ่นทุกคน ที่ช่วยเหลือและให้กำลังใจกันเสมอมาตลอดระยะเวลาการศึกษา

ท้ายสุดนี้ ขอกราบขอบพระคุณ คุณพ่อ คุณแม่และครอบครัว ที่คอยเป็นกำลังใจและให้การสนับสนุนผู้วิจัยในทุกๆ ด้านเสมอมาจวบจนสำเร็จการศึกษา

วุฒิชัย กำจรกิตติคุณ

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ฆ

บทคัดย่อภาษาอังกฤษ..... ง

กิตติกรรมประกาศ..... จ

สารบัญตาราง..... ซ

สารบัญภาพ..... ฌ

บทที่

1. บทนำ..... 1

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของการศึกษา..... 3

1.3 ขอบเขตของการศึกษา..... 3

1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 3

1.5 การทดสอบระบบ..... 4

1.6 วัสดุอุปกรณ์..... 4

1.7 แผนการดำเนินงาน..... 4

1.8 ความรู้ใหม่ที่ได้..... 5

1.9 การตอบรับการตีพิมพ์..... 5

1.10 โครงสร้างของรายงานส่วนที่เหลือ..... 5

2. แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง..... 6

2.1 ระบบอ่านหน่วยไฟฟ้าอัตโนมัติ (Automatic Meter Reading : AMR)..... 6

2.2 ลักษณะการคิดค่าไฟฟ้าและองค์ประกอบทางไฟฟ้า..... 12

2.3 การจำแนกประเภทข้อมูล (Classification) และอัลกอริทึม (Algorithm)..... 14

2.4 งานวิจัยที่เกี่ยวข้อง..... 28

3. การดำเนินงาน..... 32

3.1 แนวทางการวิจัยและพัฒนา..... 32

3.2 แผนการดำเนินงาน..... 32

3.3 เครื่องมือที่ใช้ในการวิจัย..... 33

สารบัญ (ต่อ)

บทที่	หน้า
3.4 การศึกษาข้อมูลการใช้ไฟฟ้าของผู้ใช้ไฟแต่ละประเภท	33
3.5 การออกแบบและกำหนดลักษณะประจำ (Attribute)	41
4. ผลการทดลอง	52
4.1 การทดลองและวัดประสิทธิภาพ	52
4.2 ผลการทดสอบโมเดลที่สร้างด้วยนาอีฟเบย์	52
4.3 ผลการทดสอบโมเดลที่สร้างด้วยนิวรัลเน็ทเวิร์ค	55
5. สรุปผลการวิจัยและข้อเสนอแนะ	58
5.1 สรุปผลการทดลอง	58
5.2 ปัญหาและข้อเสนอแนะ	59
บรรณานุกรม	60
ภาคผนวก	63
ประวัติผู้เขียน	71

สารบัญตาราง

ตารางที่ **หน้า**

1.1 แผนการดำเนินงาน..... 4

2.1 แสดงชุดข้อมูลของสภาพอากาศ……..... 19

2.2 แสดงผลการสร้างนาอีฟเบย์โมเดล..... 20

3.1 กำหนดสัดส่วนของเวลาในวันทำการ..... 45

3.2 Load Shape Impact..... 46

3.3 Attributes ที่ใช้ในการทดสอบ..... 50

4.1 ความเที่ยงตรงของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยนาอีฟเบย์..... 53

4.2 ความคลาดเคลื่อนของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยนาอีฟเบย์..... 53

4.3 ความเที่ยงตรงของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนาอีฟเบย์..... 54

4.4 ความคลาดเคลื่อนของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนาอีฟเบย์.. 54

4.5 ความเที่ยงตรงของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยด้วยนิวรัลเน็ทเวิร์ค…..... 55

4.6 ความคลาดเคลื่อนของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยด้วยนิวรัลเน็ทเวิร์ค..... 55

4.7 ความเที่ยงตรงของผลการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยด้วยนิวรัลเน็ทเวิร์ค..... 56

4.8 ความคลาดเคลื่อนของผลการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยด้วยนิวรัลเน็ทเวิร์ค..... 56

5.1 สรุปผลการทดลอง..... 58

สารบัญภาพ

ภาพที่ **หน้า**

1.1 ระบบการจัดการข้อมูลมิเตอร์ AMR ของการไฟฟ้าส่วนภูมิภาค..... 2

2.1 ระบบการสื่อสารของโครงการพัฒนาการอ่านหน่วยด้วยระบบอัตโนมัติ..... 8

2.2 ข้อมูลที่ได้จากระบบ AMR..... 11

2.3 ลักษณะความต้องการไฟฟ้าของประเทศไทย..... 13

2.4 Naive Bayes Flow Chart……..... 18

2.5 โครงข่ายการส่งข้อมูลแบบไม่ย้อนกลับ (Feedforward Network)..... 23

2.6 โครงข่ายการส่งข้อมูลแบบย้อนกลับ (Feedback Network)..... 24

2.7 เพอร์เซปตรอนแบบชั้นเดียว (Single-layer Perceptron : SLP)..... 25

2.8 เพอร์เซปตรอนแบบหลายชั้น (Multi-layer Perceptron : MPL)..... 26

2.9 ลักษณะการทำ Cross-validation..... 28

2.10 แผนผังการทำ Classification ของงานวิจัยที่นำเสนอ..... 29

2.11 แผนผังอัลกอริทึมของงานวิจัยที่นำเสนอ..... 30

3.1 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทที่พักอาศัย..... 34

3.2 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทที่พักอาศัย..... 34

3.3 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทที่พักอาศัย..... 35

3.4 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทอาคารสำนักงาน..... 36

3.5 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทอาคารสำนักงาน..... 36

3.6 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทอาคารสำนักงาน..... 37

3.7 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม..... 38

3.8 อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม..... 38

3.9 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม. 39

3.10 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทสถานพยาบาล..... 40

3.11 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทสถานพยาบาล..... 40

3.12 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทสถานพยาบาล..... 41

3.13 Flowchart การทดสอบที่นำเสนอ..... 42

3.14 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนมกราคม..... 44

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

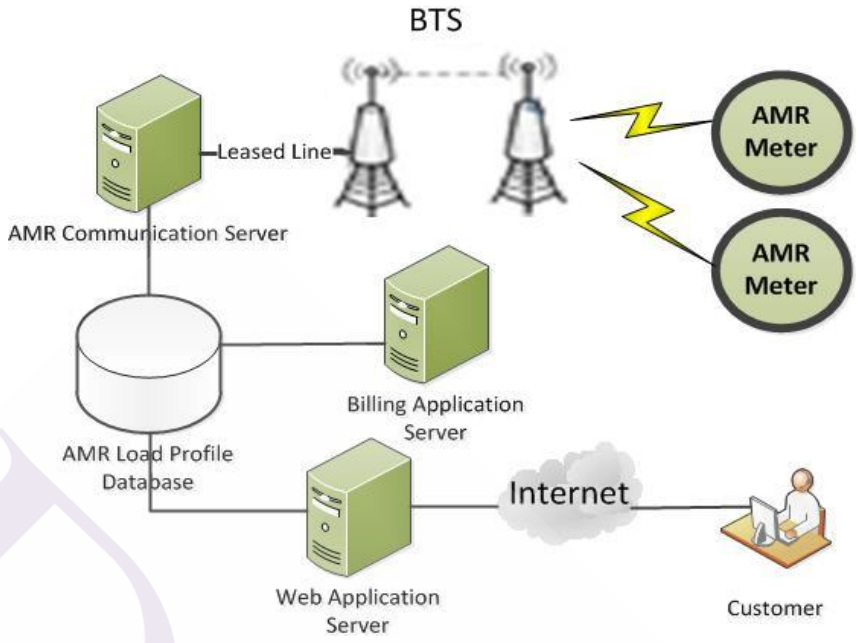
3.15 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนมีนาคม..... 44
3.16 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนกันยายน..... 45
3.17 เปรียบเทียบระหว่าง Working Time Impact กับ Evening Impact ของเดือนมกราคม 47
3.18 เปรียบเทียบระหว่าง Working Time Impact กับ Evening Impact ของเดือนกันยายน 47
3.19 เปรียบเทียบระหว่าง Working Time Impact กับ Night Impact ของเดือนมิถุนายน… 48
3.20 เปรียบเทียบระหว่าง Working Time Impact กับ Night Impact ของเดือนกันยายน… 49
3.21 เปรียบเทียบกำลังไฟฟ้ารีแอคทีฟกับความแตกต่างของกำลังไฟฟ้ารีแอกทีฟ..... 50

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ธุรกิจและอุตสาหกรรมทุกประเภทล้วนมีความต้องการใช้พลังงานไฟฟ้าในการดำเนินกิจการทั้งสิ้น ยิ่งในยุคปัจจุบันที่มีความผันผวนทางเศรษฐกิจอย่างมากนั้น ผู้ให้บริการพลังงานไฟฟ้ายิ่งต้องมีการพัฒนาระบบให้มีเสถียรภาพ มั่นคง และมีความเชื่อถือได้เพื่อรองรับความต้องการพลังงานไฟฟ้าที่เปลี่ยนแปลงไปในทุกเวลา ระบบไฟฟ้าอัจฉริยะ (Smart Grid) จึงเป็นอีกหนึ่งเทคโนโลยีที่สามารถเข้ามาพัฒนาระบบไฟฟ้าให้ดียิ่งขึ้นได้ โดยมีการผนวกรวมระหว่างโครงข่ายไฟฟ้า โครงข่ายการสื่อสารและเทคโนโลยีสารสนเทศเข้าด้วยกัน เพื่อจัดการระบบผลิต ระบบสายส่ง ระบบจำหน่าย รวมถึงการควบคุมการใช้ไฟฟ้าให้เกิดประสิทธิภาพสูงสุด ในส่วนของระบบจำหน่ายและการควบคุมการใช้ไฟฟ้านั้น ประการสำคัญที่ทำให้เกิดการควบคุมได้คือโครงสร้างพื้นฐานสำหรับการวัดค่าขั้นสูง (Advanced Metering Infrastructure: AMI) ซึ่งมีความสามารถในการอ่านหน่วยไฟฟ้าอัตโนมัติ ทำให้ผู้ให้บริการและผู้ใช้ไฟฟ้าสามารถทราบข้อมูลการใช้พลังงานไฟฟ้าได้อย่างเป็นปัจจุบัน นำไปสู่การบริหารจัดการพลังงานอย่างมีประสิทธิภาพ

ในปัจจุบันการไฟฟ้าส่วนภูมิภาค (กฟภ.) ได้มีการใช้งานระบบอ่านหน่วยอัตโนมัติ (Automatic Meter Reading : AMR) โดยติดตั้งมิเตอร์ AMR ให้กับผู้ใช้ไฟฟ้ารายใหญ่เพื่อให้ได้มาซึ่งหน่วยการใช้ไฟฟ้าประจำเดือนโดยไม่ต้องส่งพนักงานไปทำการอ่านหน่วย ณ จุดติดตั้งมิเตอร์ อีกทั้งผู้ใช้ไฟฟ้ายังสามารถตรวจสอบข้อมูลการใช้พลังงานไฟฟ้าของตนเองผ่านทางอินเตอร์เน็ตได้อีกด้วยโดยมิเตอร์จะส่งข้อมูลการใช้พลังงานไฟฟ้า (Electric Load Profile) ผ่านอุปกรณ์สื่อสารให้กับระบบทุกๆ 15 นาที ซึ่งข้อมูลดังกล่าวจะถูกรวบรวมและจัดเก็บในฐานข้อมูล (Database) ซึ่งสามารถนำมาวิเคราะห์และแสดงผลในรูปแบบต่างๆ มากมาย อีกทั้งยังก่อให้เกิดบริหารจัดการข้อมูลอย่างมีประโยชน์ดังภาพที่ 1.1



ภาพที่ 1.1 ระบบการจัดการข้อมูลมิเตอร์ AMR ของการไฟฟ้าส่วนภูมิภาค

การบริหารจัดการพลังงานไฟฟ้าของประเทศที่ดีนั้นจำเป็นต้องมีการจัดเก็บข้อมูลการใช้ไฟฟ้าตามประเภทกิจการที่ถูกต้องแม่นยำ เนื่องจากกิจการแต่ละประเภทจะมีลักษณะการใช้ไฟฟ้าที่แตกต่างกันไป บางประเภทมีความต้องการใช้ไฟฟ้าสูงในช่วงกลางวัน บางประเภทมีความต้องการใช้ไฟฟ้าสูงในช่วงกลางคืน ซึ่งล้วนกระทบต่อกระบวนการผลิตและจัดสรรพลังงานของผู้ให้บริการพลังงานไฟฟ้าทั้งสิ้น นอกจากนี้การศึกษาพฤติกรรมการใช้ไฟฟ้าของกิจการแต่ละประเภทอย่างละเอียดยังเป็นปัจจัยชี้นำการเติบโตทางเศรษฐกิจของประเทศได้อีกด้วย โดยเฉพาะกรณีที่มีการส่งเสริมหรือพัฒนาสินค้าใหม่ให้เกิดขึ้นในตลาด เช่น ส่งเสริมการใช้รถยนต์ไฟฟ้าภายในประเทศ การผลิตพลังงานไฟฟ้าโดยผู้บริโภค (Prosumer) เป็นต้น ดังนั้นการกำหนดหรือจำแนกประเภทกิจการให้ถูกต้องจึงถือเป็นสิ่งสำคัญที่จะทำให้นโยบายหรือมาตรการบริหารจัดการพลังงานไฟฟ้าถูกขับเคลื่อนไปและสัมฤทธิ์ผลตามข้อกำหนดหรือสมมติฐานของภาครัฐ เช่น การหาผู้เข้าร่วมที่เหมาะสมสำหรับมาตรการความร่วมมือลดการใช้ไฟฟ้า (Emergency Demand Response Program : EDRP) รองรับการหยุดจ่ายก๊าซธรรมชาติแหล่งยาดานา[8] การกำหนดอัตราค่าไฟฟ้าที่เหมาะสมสำหรับผู้ใช้ไฟฟ้าที่มีการติดตั้งโซล่าร์เซลล์ในครัวเรือน เป็นต้น นอกจากนี้การทราบประเภทกิจการที่ชัดเจนของผู้ใช้ไฟฟ้ายังเป็นฐานข้อมูลที่สำคัญสำหรับใช้ในการวางแผนก่อสร้างสถานีสำหรับรองรับปริมาณความต้องการพลังงานที่เพิ่มขึ้น รวมทั้งการวางแผนตัดกระแสไฟฟ้าเพื่อแก้ไขหรือบำรุงรักษาระบบไฟฟ้าให้เกิดผลกระทบกับผู้ใช้ไฟฟ้าน้อยที่สุด แต่ด้วย

ข้อมูลประเภทกิจการจะถูกกำหนดไว้ตั้งแต่เริ่มต้นประกอบกิจการและจะมีสถานภาพนั้นตลอดไป โดยไม่มีการปรับเปลี่ยนหรือหมดอายุ ถึงแม้ว่าจะมีการประกอบกิจการอื่นเพิ่มเติมหรือเปลี่ยนแปลงประเภทกิจการไปแล้วก็ตาม ย่อมส่งผลให้การพยากรณ์และการวิเคราะห์ผลสัมฤทธิ์ของมาตรการการบริโภคพลังงานไฟฟ้ามีความคลาดเคลื่อน ไม่เป็นไปตามที่คาดการณ์ไว้ ซึ่งเป็นปัญหาที่ต้องได้รับการแก้ไขเพื่อให้การดำเนินงานของผู้ให้บริการพลังงานไฟฟ้าเป็นไปอย่างมีประสิทธิภาพ

งานวิจัยนี้ได้พัฒนาวิธีการจำแนกประเภทกิจการโดยอาศัยเทคนิคต่างๆ ประกอบด้วย นาอีฟเบย์ และโครงข่ายประสาทเทียม เพื่อใช้จำแนกข้อมูลการใช้ไฟฟ้าที่ได้จากมิเตอร์ระบบ AMR ให้ได้ประเภทกิจการที่ถูกต้อง แม่นยำและเป็นปัจจุบัน

1.2 วัตถุประสงค์

1. เพื่อวิเคราะห์และกำหนดลักษณะประจำ (Attribute) ของแต่ละประเภทกิจการจากข้อมูลการใช้ไฟฟ้า
2. เพื่อพัฒนาวิธีการจำแนกประเภทกิจการโดยอาศัยเทคนิคต่างๆเพื่อใช้จำแนกข้อมูลการใช้ไฟฟ้า

1.3 ขอบเขต

1. วิเคราะห์รูปแบบการใช้ไฟฟ้าจากข้อมูลการใช้ไฟฟ้าที่ได้จากระบบ AMR และกำหนดลักษณะประจำ (Attribute) ของกิจการแต่ละประเภท ซึ่งประกอบไปด้วย ประเภทอุตสาหกรรม ประเภทสถานพยาบาล ประเภทอาคารสำนักงาน และประเภทบ้านที่อยู่อาศัย
2. พัฒนาวิธีการจำแนกประเภทกิจการโดยอาศัยเทคนิคนาอีฟเบย์ และ โครงข่ายประสาทเทียม
3. เปรียบเทียบประสิทธิภาพและผลลัพธ์การทำงานของระหว่างเทคนิคทั้งสองโดยใช้โปรแกรมแรพิดไมเนอร์ สตูดิโอ (RapidMiner Studio)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถจำแนกข้อมูลประเภทกิจการให้เป็นปัจจุบันได้อย่างถูกต้อง และมีความแม่นยำมากกว่าร้อยละ 80
2. ข้อมูลหลังจากผ่านการจำแนกแล้ว เป็นข้อมูลที่มีความน่าเชื่อถือ สามารถนำไปวิเคราะห์หรือขยายผลต่อได้อย่างถูกต้อง สามารถตอบสนองนโยบายภาครัฐได้อย่างมีประสิทธิภาพ
3. ลดค่าใช้จ่ายในการตรวจสอบข้อมูลประเภทกิจการ ณ สถานที่จริง

1.5 การทดสอบระบบ

ในการทดสอบระบบจะทำการทดสอบในแง่ของประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคนาอีฟเบย์และโครงข่ายประสาทเทียม โดยใช้ข้อมูลที่ได้จากการอ่านหน่วยการใช้ไฟฟ้าของผู้ใช้ไฟจากระบบ Automatic Meter Reading (AMR) ในทุกๆ 15 นาที ตั้งแต่เดือนมกราคม ปี พ.ศ. 2559 จนถึง เดือนธันวาคม ปี พ.ศ. 2559 ทั้งนี้ผู้ใช้ไฟฟ้าทุกรายจะมีข้อมูลประเภทกิจการที่ระบุไว้อย่างแน่ชัดกำกับไว้แล้ว โดยจะแบ่งเป็นชุดข้อมูลในการฝึกสอนจำนวน 100 ราย และ ชุดข้อมูลในการทดสอบจำนวน 64 ราย ซึ่งจะแยกออกจากกันอย่างสิ้นเชิง จากนั้นทำการทดสอบโดยเปรียบเทียบผลการจำแนกที่ได้จากเทคนิคนาอีฟเบย์และโครงข่ายประสาทเทียม ซึ่งคิดเป็นร้อยละความถูกต้องของการจำแนก

1.6 วัสดุอุปกรณ์

1. ฮาร์ดแวร์ (Hardware)
   1) คอมพิวเตอร์ส่วนบุคคล (Personal Computer) ใช้สำหรับการทดสอบการทำงาน
2. ซอฟต์แวร์ (Software)
   1) แรพิดไมเนอร์ สตูดิโอ เวอร์ชัน 8.1.000 (RapidMiner Studio 8.1.000) ใช้สำหรับวิเคราะห์และประมวลผลข้อมูลการจำแนกประเภทกิจการ
   2) ไมโครซอฟท์ เอ็กเซล เวอร์ชัน 2010 (Microsoft Excel 2010) ใช้สำหรับคำนวณผลลัพธ์ของค่าการใช้พลังงานให้เป็นไปตามลักษณะประจำ (Attribute) ที่กำหนด

1.7 แผนการดำเนินงาน

ในการดำเนินการวิจัยได้ตั้งแผนดำเนินการดังตารางที่ 1.1

ตารางที่ 1.1 แผนการดำเนินงาน

รายการดำเนินงาน	ระยะเวลา (เดือน)											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. ศึกษารูปแบบข้อมูลการใช้ไฟฟ้าที่ได้จากมิเตอร์ AMR												
2. ศึกษาลักษณะการใช้ไฟฟ้าของกิจการแต่ละประเภท												

รายการดำเนินงาน	ระยะเวลา (เดือน)											
	1	2	3	4	5	6	7	8	9	10	11	12
3. ศึกษาการทำงานการของอัลกอริทึมต่างๆ												
4. จัดทำชุดข้อมูลสำหรับทดสอบ												
5. ทดสอบและปรับปรุงประสิทธิภาพการทำงาน												
6. วิเคราะห์และสรุปผลงาน												
7. รวบรวมข้อมูลที่ได้จัดทำวิทยานิพนธ์												

1.8 ความรู้ใหม่ที่ได้

1. ความรู้ในเรื่องของลักษณะพฤติกรรมการการใช้ไฟฟ้าของกิจการแต่ละประเภท
2. การประยุกต์เทคนิคการจำแนกข้อมูล (Classification) มาใช้กับข้อมูลทางด้านไฟฟ้า

1.9 การตอบรับการตีพิมพ์

งานวิจัยนี้ได้รับการตอบรับให้ตีพิมพ์โดยมีรายละเอียดดังนี้

1. หัวข้อ “การวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์” การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ ครั้งที่ 8 (NCIT 2016) ปี พ.ศ. 2559 หน้า 404-408.

2. หัวข้อ “การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์” งานประชุมวิชาการและนวัตกรรม กฟภ.(PEACON & INNOVATION 2017) ปี พ.ศ. 2560 หน้า 277-282.

1.10 โครงสร้างของรายงานส่วนที่เหลือ

บทที่ 2 กล่าวถึงแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

บทที่ 3 กล่าวถึงการออกแบบ

บทที่ 4 กล่าวถึงผลการทดลองที่จะใช้ในวิทยานิพนธ์

บทที่ 5 กล่าวถึงบทสรุปของงานที่จะใช้ในวิทยานิพนธ์นี้

บทที่ 2
แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่องการจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์และโครงข่ายประสาทเทียมนั้น จำเป็นที่จะต้องศึกษาตั้งแต่ ความรู้ทั่วไปเกี่ยวกับค่าพลังงานไฟฟ้าที่วัดได้จากเครื่องวัด (มิเตอร์ไฟฟ้า) องค์ประกอบทางไฟฟ้า การคิดค่าการใช้พลังงานไฟฟ้า พฤติกรรมการใช้ไฟฟ้าของกิจการแต่ละประเภท วิธีการและอัลกอริทึมที่ใช้ในการจำแนก รวมถึงงานวิจัยที่เกี่ยวข้อง ในบทนี้จะกล่าวถึงประเด็นที่สำคัญโดยมีรายละเอียดดังนี้

2.1 ระบบอ่านหน่วยไฟฟ้าอัตโนมัติ (Automatic Meter Reading : AMR)

โดยทั่วไปการอ่านหน่วยไฟฟ้าจะหมายถึงการส่งเจ้าหน้าที่ไปยังจุดติดตั้งมิเตอร์เพื่อทำการอ่านหน่วยการใช้ไฟฟ้าในเดือนปัจจุบัน แล้วนำข้อมูลดังกล่าวมาประมวลผลบิลค่าไฟฟ้าประจำเดือน ส่วนการอ่านหน่วยไฟฟ้าอัตโนมัตินั้นจะไม่ใช้บุคคลเป็นสื่อกลางในการนำข้อมูลหน่วยการใช้ไฟฟ้ามาใช้ในการประมวลผลบิล แต่จะใช้เทคโนโลยีสารสนเทศมาดำเนินการแทนซึ่งมีรายละเอียดดังนี้

2.1.1 ประวัติความเป็นมาของระบบอ่านหน่วยไฟฟ้าอัตโนมัติ (AMR)

[9] ระบบอ่านหน่วยไฟฟ้าอัตโนมัติ (Automatic Meter Reading : AMR) ได้ถูกคิดค้นและผลิตขึ้นในปี พ.ศ.2510 โดย Mr. Paraskevakos ชาวสหรัฐอเมริกา ซึ่งระบบ AMR สามารถทำงานแบบอัตโนมัติ สามารถสั่งการระยะไกลไปที่ตัวมิเตอร์ และควบคุมการใช้ไฟฟ้าได้อีกด้วย โดย Mr. Paraskevakos ยังเป็นผู้ให้กำเนิดคอมพิวเตอร์ IBM series 1 ในปีเดียวกันนี้ด้วย ซึ่งเป็นยุคก่อนที่จะมีระบบอินเตอร์เน็ตใช้งาน ทำให้ระบบดังกล่าวไม่เป็นที่รู้จักอย่างแพร่หลาย และยังไม่สามารถใช้งานจริงได้ ต่อมาในปี พ.ศ.2548 องค์กร US Energy Policy Act ได้สั่งการให้คณะกรรมการกำกับพลังงานไฟฟ้า หรือ Electric Utility Regulator กำหนดให้มีการคิดค่าไฟฟ้าแบบช่วงเวลา (Time of use rate : TOU) ทำให้ระบบมิเตอร์ AMR เข้ามามีบทบาทในการคิดค่าไฟฟ้าในระบบดังกล่าวและเป็นที่รู้จักมากขึ้น การติดตั้งใช้งานมิเตอร์ AMR ในระยะเริ่มแรกเป็นการติดตั้งให้กับผู้ใช้ไฟฟ้าประเภทครัวเรือนและประเภทอุตสาหกรรมตามลำดับ ซึ่งมักใช้คลื่นความถี่วิทยุหรือโครงข่ายผู้ให้บริการมือถือเป็นตัวกลางในการส่งผ่านข้อมูล ต่อมาองค์กร

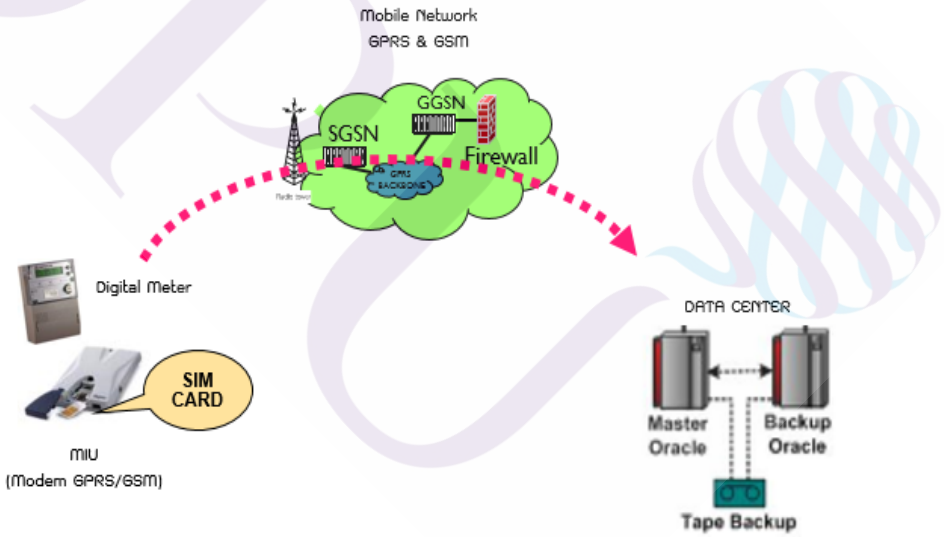
Automatic Meter Reading Association (AMRA) ได้จัดตั้งขึ้นและให้ National Association of Regulatory Utility Commissioners (NARUC) ดำเนินการคิดค้นระบบมิเตอร์ที่มีชื่อว่า Advanced Metering Infrastructure (AMI) ขึ้นซึ่งเป็นระบบที่มีความสามารถเหนือกว่าระบบ AMR เดิมโดยระบบนี้สามารถสื่อสารได้สองทิศทาง นั่นคือสามารถส่งข้อมูลที่วัดค่าและตรวจจับได้เข้าสู่ระบบและยังสามารถโปรแกรมค่าพารามิเตอร์จากระยะไกลได้อีกด้วย มิเตอร์ AMI จึงรองรับการคิดค่าไฟฟ้าในรูปแบบราคาไม่คงที่ (Dynamic pricing) ที่ทำให้ผู้ใช้ไฟฟ้าเข้ามามีบทบาทในการกำหนดราคาค่าไฟฟ้าของตนเองมากยิ่งขึ้น การค้นคว้าและพัฒนาดังกล่าวสำเร็จในปี พ.ศ.2550 ซึ่งเป็นปีที่ประเทศสหรัฐอเมริกาได้ประกาศใช้คำว่า Smart Grid หรือโครงข่ายไฟฟ้าอัจฉริยะนั่นเองก่อให้เกิดการปรับเปลี่ยนการให้บริการพลังงานไฟฟ้ารูปโฉมใหม่ที่ทุกประเทศในโลกกำลังจับตามอง อีกทั้งบริษัทผู้ผลิตระบบมิเตอร์ต่างก็พัฒนาความสามารถของอุปกรณ์ของตนเพื่อให้รองรับระบบโครงข่ายไฟฟ้าอัจฉริยะนั้น ด้วยเหตุนี้จึงถือได้ว่ามิเตอร์ AMI หรือเรียกอีกชื่อหนึ่งว่า Smart Meter เป็นองค์ประกอบสำคัญของโครงข่ายไฟฟ้าอัจฉริยะ

2.1.2 โครงการพัฒนาการอ่านหน่วยด้วยระบบอัตโนมัติของ กฟภ.

การไฟฟ้าส่วนภูมิภาคได้เริ่มศึกษาระบบมิเตอร์ AMR ในช่วงปี พ.ศ.2548 และได้ก่อตั้งโครงการพัฒนาการอ่านหน่วยไฟฟ้าด้วยระบบอัตโนมัติขึ้นในปี พ.ศ.2550 ก่อนที่ประเทศสหรัฐอเมริกาจะประกาศใช้คำว่า Smart Grid หรือโครงข่ายไฟฟ้าอัจฉริยะ ทำให้ดำเนินโครงการโดยใช้ชื่อว่าระบบ AMR โดยมีเป้าหมายหลักของโครงการ คือ ลดค่าใช้จ่ายในการจัดส่งเจ้าหน้าที่ของการไฟฟ้าส่วนภูมิภาคออกไปจดหน่วยและเพิ่มความแม่นยำของหน่วยการใช้ไฟฟ้าทำให้ผู้ใช้ไฟฟ้าเกิดความมั่นใจนั่นเอง ตามแผนที่นำทางการพัฒนาระบบโครงข่ายไฟฟ้าอัจฉริยะของการไฟฟ้าส่วนภูมิภาค (PEA Smart Grid Roadmap) นั้นได้มีการวางแนวทางการเข้าสู่ระบบมิเตอร์อัจฉริยะหรือระบบ AMI ที่แท้จริงโดยเริ่มต้นที่การติดตั้งมิเตอร์ระบบ AMR ให้กับผู้ใช้ไฟฟ้ากลุ่มอุตสาหกรรม หรือเรียกว่าผู้ใช้ไฟฟ้ารายใหญ่ โดยผู้ใช้ไฟฟ้ารายใหญ่นั้นจะต้องมีคุณสมบัติดังนี้ คือ ติดตั้งหม้อแปลงไฟฟ้าที่มีขนาดตั้งแต่ 100 กิโลโวลท์แอมป์ขึ้นไปหรือมีค่ากำลังไฟฟ้าเฉลี่ย 15 นาทีสูงสุดในเดือนไม่น้อยกว่า 30 กิโลวัตต์ เช่น กลุ่มผู้ใช้ไฟฟ้าประเภท 3 กิจการไฟฟ้าขนาดกลาง กลุ่มผู้ใช้ไฟฟ้าประเภท 4 กิจการไฟฟ้าขนาดใหญ่ เป็นต้น เนื่องจากผู้ใช้ไฟฟ้าดังกล่าวมีฐานการใช้พลังงานสูงกว่าผู้ใช้ไฟฟ้าประเภทบ้านที่อยู่อาศัยหรือเรียกว่าผู้ใช้ไฟฟ้ารายย่อย อีกทั้งยังมีรายละเอียดของข้อมูลหน่วยการใช้ไฟฟ้าที่มากกว่าอีกด้วย ดังนั้นเพื่อป้องกันความผิดพลาดในการจดหน่วยไฟฟ้าที่จะเกิดขึ้น การไฟฟ้าส่วนภูมิภาคจึงได้เริ่มติดตั้งมิเตอร์ AMR ให้ผู้ใช้ไฟฟ้าประเภทดังกล่าว

2.1.3 โครงสร้างมิเตอร์ในโครงการพัฒนาระบบการอ่านหน่วยด้วยระบบอัตโนมัติ

ระบบมิเตอร์ AMR หรือมิเตอร์อัจฉริยะประเภท AMR นั้นมีโครงสร้างหลัก 2 ส่วนคือ มิเตอร์อิเล็กทรอนิกส์ที่ใช้สำหรับวัดค่าพารามิเตอร์ทางไฟฟ้าและส่วนของระบบสื่อสารสำหรับส่งข้อมูลที่มิเตอร์วัดค่าได้เข้าสู่ระบบควบคุมมิเตอร์ที่สำนักงานใหญ่ของการไฟฟ้าส่วนภูมิภาคดังภาพที่ 2.1 โดยส่วนประกอบภายในของมิเตอร์อิเล็กทรอนิกส์นั้นประกอบไปด้วย เซนเซอร์วัดค่าพารามิเตอร์ต่างๆ หม้อแปลงแรงดันไฟฟ้า หม้อแปลงกระแสไฟฟ้า หน้าจอแสดงผล แบตเตอร์รี่สำหรับจ่ายไฟเลี้ยงให้กับนาฬิกาและหน่วยความจำในตัวมิเตอร์ หน่วยความจำในมิเตอร์นั้นจะเก็บข้อมูลต่างๆ ซึ่งเป็นหน่วยความจำประเภทแฟลช (Flash) สามารถเก็บบันทึกข้อมูลไว้ในหน่วยความจำเป็นระยะเวลา 45 วัน (ขึ้นอยู่กับการโปรแกรมค่าพารามิเตอร์ที่ต้องการบันทึก หากต้องการบันทึกค่าพารามิเตอร์จำนวนมากก็จะเก็บข้อมูลได้เป็นระยะเวลาสั้นกว่าที่กำหนด แต่หากต้องการบันทึกค่าพารามิเตอร์จำนวนน้อยก็จะสามารถเก็บข้อมูลได้เป็นระยะเวลานานกว่าที่กำหนดนั่นเอง) โดยการเก็บข้อมูลของตัวมิเตอร์ประเภทนี้มีลักษณะเป็นแบบเข้าก่อน - ออกก่อน (First in – First out)



ภาพที่ 2.1 ระบบการสื่อสารของโครงการพัฒนาการอ่านหน่วยด้วยระบบอัตโนมัติ

ที่มา: กองพัฒนาระบบมิเตอร์ การไฟฟ้าส่วนภูมิภาค, พ.ศ. 2557

มิเตอร์ AMR ของการไฟฟ้าส่วนภูมิภาค สามารถแบ่งได้เป็น 3 ประเภท คือ

1. มิเตอร์ที่ใช้วัดไฟฟ้าประเภท 3 เฟส 4 สาย ใช้กับระบบไฟฟ้า 115 กิโลโวลต์
2. มิเตอร์ที่ใช้วัดไฟฟ้าประเภท 3 เฟส 3 สาย ใช้กับระบบไฟฟ้า 22/33 กิโลโวลต์
3. มิเตอร์ที่ใช้วัดไฟฟ้าประเภท 3 เฟส 4 สาย ใช้กับระบบไฟฟ้า 220/380 โวลต์

ส่วนระบบการสื่อสารนั้นจะประกอบไปด้วย พอร์ทสื่อสาร (Modem Port), โมเด็ม (Modem), เสาอากาศ (Antenna) และซิมการ์ด (Sim Card) โดยพอร์ทสื่อสารในระยะแรกจะใช้เป็น RS232 แต่เนื่องจากไม่สะดวกต่อการใช้งานจึงได้ปรับปรุงเป็น RJ45 เพื่อให้ง่ายต่อการใช้งานแต่ยังคงใช้โปรโตคอล RS232 ในการสื่อสารระหว่างมิเตอร์กับโมเด็มเช่นเดิม โมเด็มที่ใช้ในการติดต่อสื่อสารจะติดตั้งภายใต้ฝาครอบตัวมิเตอร์ ต่อเชื่อมกับตัวมิเตอร์ผ่านทางพอร์ทสื่อสาร และส่งข้อมูลผ่านทางเสาอากาศ (ในกรณีทั่วไปจะใช้เสาอากาศขนาด 5 DB แต่ในกรณีที่พื้นที่ติดตั้งมิเตอร์ดังกล่าวมีความแรงของสัญญาณต่ำจะเปลี่ยนเป็นเสาอากาศขนาด 9 DB เพื่อขยายอัตราการรับสัญญาณ (Gain) ให้ดีขึ้น) ผ่านโครงข่ายของผู้ให้บริการเครือข่ายโทรศัพท์เคลื่อนที่ (Mobile Operator) เพื่อส่งไปยังระบบควบคุมมิเตอร์

จากที่กล่าวมานั้นมิเตอร์ AMR จะทำการส่งค่าที่วัดได้ทุกๆ 15 นาที และส่งข้อมูลผ่านเทคโนโลยีแบบไร้สายไปที่ระบบควบคุมข้อมูลมิเตอร์ มิเตอร์ที่อยู่ในพื้นที่ที่มีการให้บริการเทคโนโลยี 3G ครอบคลุมก็จะใช้เทคโนโลยี 3G ในการสื่อสาร ส่วนในพื้นที่ที่ยังไม่ครอบคลุมนั้นจะใช้ GPRS (General Packet Radio Service) เป็นรูปแบบการสื่อสารหลัก และใช้ GSM (Global System for Mobile Communications) เป็นการสื่อสารสำรอง (Backup) ในกรณีที่การสื่อสารหลักมีปัญหา โดยมีเส้นทางการสื่อสารเชื่อมต่อมายังระบบคอมพิวเตอร์แม่ข่าย (Server) ซึ่งตั้งอยู่ที่กองพัฒนาระบบมิเตอร์ การไฟฟ้าส่วนภูมิภาคสำนักงานใหญ่ เนื่องจากข้อมูลมิเตอร์ AMR เป็นข้อมูลที่สำคัญของผู้ใช้ไฟฟ้าทำให้การไฟฟ้าส่วนภูมิภาคจำเป็นต้องมีการดูแลป้องกันข้อมูลเป็นอย่างดี ไม่ให้บุคคลทั่วไปสามารถเข้าถึงข้อมูลได้ จึงกำหนดให้ซิมการ์ดที่ติดตั้งใช้งานอยู่ในโมเด็มจะต้องมีการป้องกันโดยจะระบุ IP address เฉพาะซึ่งเป็นการสื่อสารระหว่างแม่ข่ายกับซิมการ์ดเท่านั้น (Sim-To-Server) เพื่อความปลอดภัยของข้อมูล (Security) ดังนั้นข้อมูลการใช้ไฟฟ้าของมิเตอร์ทุก 15 นาทีจึงสามารถเรียกได้จากระบบคอมพิวเตอร์แม่ข่ายเท่านั้น และเมื่อได้รับข้อมูลเข้าสู่ฐานข้อมูล (Database) แล้วจึงจะนำข้อมูลเหล่านี้ไปใช้ในการสนับสนุนแอพลิเคชันต่างๆ ต่อไป

2.1.4 ข้อมูลที่ได้จากการอ่านหน่วยไฟฟ้าอัตโนมัติ

ข้อมูลที่ได้จากการอ่านหน่วยไฟฟ้าอัตโนมัติประกอบไปด้วยค่าทางไฟฟ้าและค่าการตรวจจับจากเซ็นเซอร์อื่น ประกอบด้วย

1. ค่าทางพลังงานไฟฟ้า

กำลังไฟฟ้า (Electric Power) จะเห็นว่าอุปกรณ์ไฟฟ้าทุกประเภททั้งในอาคารบ้านเรือนหรือโรงงานอุตสาหกรรมล้วนต้องพึ่งพากำลังไฟฟ้าในการทำงาน โดยกำลังไฟฟ้าหมายถึง ปริมาณไฟฟ้าที่มีแรงดันไฟฟ้าและกระแสไฟฟ้าที่ใช้ไปกับอุปกรณ์ ประกอบไปด้วยสองส่วนคือ กำลังไฟฟ้าจริง (Real Power : P) มีหน่วยเป็นวัตต์ (Watt : W) หรือกิโลวัตต์ (kW) เป็นกำลังไฟฟ้าที่สามารถเปลี่ยนแปลงโดยอุปกรณ์ไฟฟ้าไปเป็นพลังงานรูปอื่นได้ เช่น พลังงานความร้อน พลังงานแสงสว่าง หรือพลังงานกล กำลังไฟฟ้าส่วนนี้เกิดจากกระแสไฟฟ้าที่ใช้งาน (Active Current) และอีกส่วนหนึ่งคือ กำลังไฟฟ้ารีแอกทีฟ (Reactive Power : Q) มีหน่วยเป็นวาร์ (Var) หรือกิโลวาร์ (kVar) เป็นกำลังไฟฟ้าที่ไม่สามารถเปลี่ยนแปลงไปเป็นพลังงานรูปอื่นได้แต่มีความจำเป็นต่อการทำงานของอุปกรณ์ไฟฟ้าจำพวกหม้อแปลง มอเตอร์ ฯลฯ เพื่อสร้างสนามแม่เหล็ก หากไม่มีสนามแม่เหล็กอุปกรณ์ดังกล่าวจะไม่สามารถทำงานได้ซึ่งประเภทกิจการจำพวกอุตสาหกรรมที่มีเครื่องจักรกลหรือเครื่องใช้ไฟฟ้าขนาดใหญ่มักมีการใช้กำลังไฟฟ้ารีแอกทีฟสูงกว่าประเภทกิจการอื่นๆ

พลังงานไฟฟ้า (Electric Energy) เป็นปริมาณไฟฟ้าที่อุปกรณ์ไฟฟ้าใช้ไปทั้งหมดหรือแหล่งจ่ายไฟฟ้าจ่ายให้ทั้งหมดซึ่งคิดในช่วงระยะเวลาใดเวลาหนึ่ง โดยแบ่งเป็นสองส่วนเช่นเดียวกับกำลังไฟฟ้า คือ พลังงานไฟฟ้าจริง (Real Energy) มีหน่วยเป็นวัตต์ชั่วโมง (Wh) หรือกิโลวัตต์ชั่วโมง (kWh) และพลังงานไฟฟ้ารีแอกทีฟ (Reactive Energy) มีหน่วยเป็นวาร์ชั่วโมง (Varh) หรือกิโลวาร์ชั่วโมง (kVarh)

พลังงานไฟฟ้าหรือเรียกอีกชื่อว่า หน่วยการใช้ไฟฟ้า (Wh) จากทฤษฎีพบว่า หน่วยการใช้ไฟฟ้าหาได้จากผลคูณของค่ากำลังไฟฟ้าจริงกับเวลา (หน่วยชั่วโมง) โดยสามารถแสดงได้ดังสมการ

$$Wh = w \times time \tag{2.1}$$

แต่ในการใช้งานจริงโหลดนั้นจะมีการเปลี่ยนแปลงตลอดเวลา ทำให้โหลดที่นำมาคิดนั้นเป็นโหลดที่เฉลี่ยในระยะเวลานั้น ดังนั้นค่า W จะสามารถแทนได้ด้วย $W_{average}$ ซึ่งจะแทนด้วย W_{ave} ได้ดังสมการ

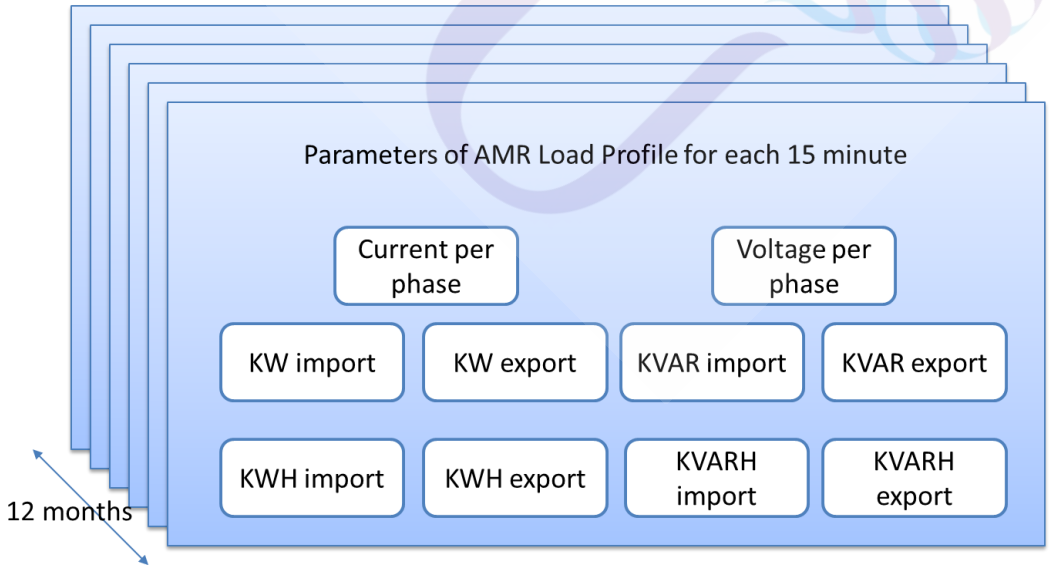
$$W_{average} = \sum_{i=0}^{time} w_i \tag{2.2}$$

การวัดค่าไฟฟ้าของการไฟฟ้าส่วนภูมิภาคจะมีการบันทึกค่ากำลังไฟฟ้าจริงหรือค่าดีมานด์ในทุกๆ 15 นาที ดังนั้นระบบมิเตอร์ AMR จึงทำการวัดหน่วยการใช้ไฟฟ้า 15 นาทีด้วย โดยสามารถเขียนสมการหน่วยการใช้ไฟฟ้า 15 นาที หรือ Wh_{15min} ได้สมการ (3)

$$Wh_{15min} = \sum_{i=0}^{15\ minute} w_i \times 15\ minute \times \frac{1}{60} \quad (2.3)$$

ซึ่งสามารถสรุปรายละเอียดของข้อมูลการใช้ไฟฟ้าที่ กฟภ. รวบรวมในแต่ละ 15 นาที ได้ดังนี้

1. ค่ากระแสไฟฟ้าในแต่ละเฟส (Current per phase หน่วย Ampere : A)
2. ค่าแรงดันไฟฟ้าในแต่ละเฟส (Voltage per phase หน่วย Volt : V)
3. ค่ากำลังไฟฟ้าจริงขารับเข้า (Import Power หน่วย Kilowatt : KW)
4. ค่ากำลังไฟฟ้าจริงขาจ่ายออก (Export Power หน่วย Kilowatt : KW)
5. ค่ากำลังไฟฟ้ารีแอกทีฟขารับเข้า (Import Reactive Power หน่วย Kilovar : KVAR)
6. ค่ากำลังไฟฟ้ารีแอกทีฟขาจ่ายออก (Export Reactive Power หน่วย Kilovar : KVAR)
7. ค่าพลังงานไฟฟ้าจริงขารับเข้า (Import Energy หน่วย Kilowatt-hour : KWH)
8. ค่าพลังงานไฟฟ้าจริงขาจ่ายออก (Export Energy หน่วย Kilowatt-hour : KWH)
9. ค่าพลังงานไฟฟ้ารีแอกทีฟขารับเข้า (Import Reactive Energy หน่วย Kilovar-hour : KVARH)
10. ค่าพลังงานไฟฟ้ารีแอกทีฟขาจ่ายออก (Export Reactive Energy หน่วย Kilovar-hour : KVARH) สามารถพิจารณาได้ดังภาพที่ 2.2



ภาพที่ 2.2 ข้อมูลที่ได้จากระบบ AMR

2. ค่าจากการตรวจจับของเซ็นเซอร์

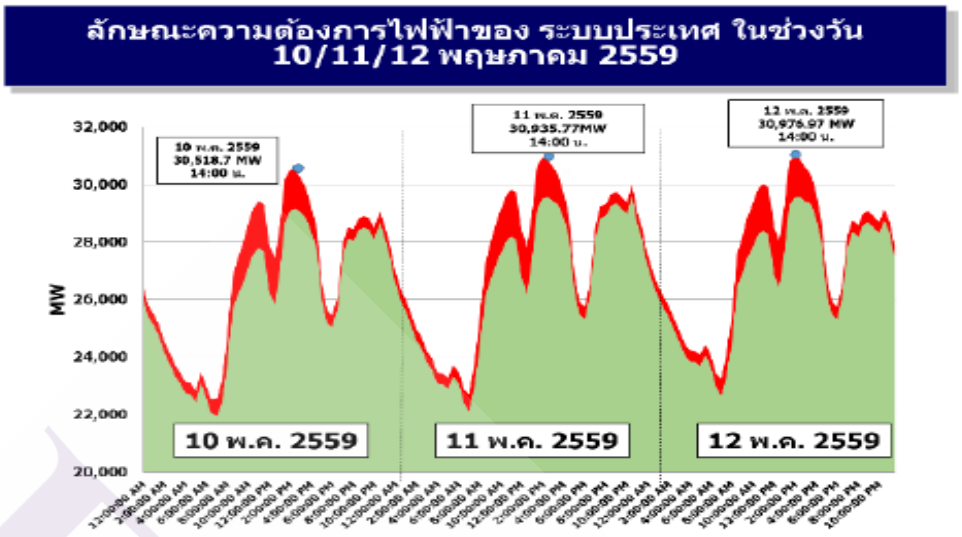
มิเตอร์ AMR จะมีพอร์ต I/O สำหรับรองรับข้อมูลจากเซ็นเซอร์ ซึ่งที่มีการใช้งานในปัจจุบันนั้น คือ รับค่าจากสวิตช์ที่ใช้ตรวจจับการเปิดปิดของตู้มิเตอร์ (Cabinet Limit Switch) เมื่อฝาตู้ปิดอยู่สวิตช์จะมีสถานะเป็น Close และเมื่อเกิดเหตุการณ์ฝาตู้ถูกเปิดสวิตช์ดังกล่าวเปลี่ยนสถานะเป็น Open และมีการบันทึกเหตุการณ์นั้นลงในหน่วยความจำของมิเตอร์ ข้อมูลดังกล่าวจะถูกส่งเข้าสู่ศูนย์ควบคุมมิเตอร์ AMR ตามรอบการเรียกข้อมูลที่ได้กำหนดไว้

2.2 ลักษณะการคิดค่าไฟฟ้าและองค์ประกอบทางไฟฟ้า

ลักษณะการคิดค่าไฟฟ้าและองค์ประกอบทางไฟฟ้าเป็นสิ่งสำคัญที่ทำให้เข้าใจพฤติกรรมการใช้ไฟฟ้าของกิจการแต่ละประเภท เพื่อใช้สำหรับการกำหนดลักษณะประจำ (Attribute) ของวิธีการจำแนกข้อมูล

2.2.1 อัตราค่าไฟฟ้าตามช่วงเวลาของการใช้ (Time of use rate : TOU)

[1] ประเทศไทยกำหนดช่วงที่มีความต้องการไฟฟ้าสูง(On Peak) ตั้งแต่วันจันทร์ – ศุกร์ เวลา 09.00 – 22.00 น. และช่วงที่มีความต้องการไฟฟ้าต่ำ (Off Peak) ตั้งแต่วันจันทร์ – ศุกร์เวลา 22.00 – 09.00 น. รวมถึงวันเสาร์ – อาทิตย์ตลอดทั้งวัน และวันหยุดราชการ (ไม่รวมวันหยุดชดเชย) ตลอดทั้งวันเช่นกัน โดยอัตราค่าไฟฟ้า TOU ที่กำหนดใช้ในปัจจุบันจะสะท้อนถึงต้นทุนไฟฟ้าอย่างแท้จริงดังภาพที่ 2.3 กล่าวคือ ในช่วงที่มีความต้องการไฟฟ้าสูงค่าไฟฟ้าจะสูงตามไปด้วย เนื่องจากผู้ให้บริการต้องลงทุนสร้างโรงไฟฟ้า ระบบสายส่งสายจำหน่ายให้เพียงพอ ต่อความต้องการไฟฟ้าในช่วงนี้ และต้องใช้เชื้อเพลิงทุกชนิดทั้งที่มีราคาถูกและแพงในการผลิตไฟฟ้า แต่ในช่วงที่มีความต้องการไฟฟ้าต่ำค่าไฟฟ้าก็จะต่ำตามไปด้วย เนื่องจากผู้ให้บริการไม่ต้องสร้างโรงไฟฟ้า ระบบสายส่งสายจำหน่ายเพิ่ม เนื่องจากมีการสร้างไว้แล้ว จึงไม่มีต้นทุนค่าไฟฟ้าในส่วนนี้ จะมีเพียงต้นทุนค่าเชื้อเพลิงในการผลิตไฟฟ้า ซึ่งผู้ให้บริการสามารถเลือกใช้เชื้อเพลิงที่มีราคาถูกมาผลิตไฟฟ้าจึงทำให้ต้นทุนพลังงานไฟฟ้าทั้งสองช่วงมีความแตกต่างกันมากกว่าร้อยละ 50



ภาพที่ 2.3 ลักษณะความต้องการไฟฟ้าของประเทศไทย

ที่มา: ฝ่ายวิศวกรรมและวิชาการ คณะกรรมการกำกับกิจการพลังงาน, พ.ศ. 2559

2.2.2 ค่าตัวประกอบการใช้ไฟฟ้า (Load Factor : LF)

ค่าตัวประกอบการใช้ไฟฟ้า หมายถึง อัตราส่วนระหว่างค่าความต้องการกำลังไฟฟ้าเฉลี่ย (Average Demand) กับค่าความต้องการกำลังไฟฟ้าสูงสุด (Maximum Demand) ในช่วงเวลาใดๆ เป็นการแสดงถึงความสม่ำเสมอของการใช้ไฟฟ้า ซึ่งหากมีค่าเข้าใกล้ ร้อยละ 100 แสดงว่ามีการใช้ไฟฟ้าอย่างเต็มประสิทธิภาพเพราะทำให้โรงไฟฟ้าต่างๆ ทำงานได้เต็มที่ คุ้มค่ากับการลงทุน โดยค่าพารามิเตอร์ทั้งสองนั้นสามารถหาได้จากพฤติกรรมการใช้ไฟฟ้า (Electrical Load Profile) โดยค่าความต้องการกำลังไฟฟ้าสูงสุดจะวัดจากค่าดีมานด์โหลดสูงสุดทุกๆ 15 นาที ดังสมการ

$$LF = \frac{P_{AVE}}{P_{MAX}} \tag{2.4}$$

P_{AVE} คือ กำลังไฟฟ้าเฉลี่ยในช่วงเวลานั้น

P_{MAX} คือ กำลังไฟฟ้าสูงสุดในช่วงเวลานั้น

2.2.3 ค่าตัวประกอบความต้องการไฟฟ้า (Demand Factor : DF)

ตัวประกอบความต้องการไฟฟ้า หมายถึง อัตราส่วนระหว่างการใช้กำลังไฟฟ้าสูงสุด (ที่ใช้งานจริง) กับโหลดทั้งหมดที่ต่ออยู่ในระบบ (Connected Load) ในช่วงเวลาหนึ่งๆ ของหน่วยการใช้พลังงานดังสมการ

$$Demand\ Factor\ = \frac{Total\ Demand\ Load}{Total\ Connected\ Load} \tag{2.5}$$

เกิดขึ้นจากมีการใช้กำลังไฟฟ้าจริงเพียงบางส่วนของ Connected Load ในหน่วยใช้ไฟฟ้าหน่วยหนึ่งๆ ในช่วงเวลาหนึ่งๆ ซึ่งตัวประกอบความต้องการไฟฟ้านี้เป็นตัวคูณที่ช่วยให้ทราบถึงการใช้กำลังไฟฟ้าที่จะเกิดขึ้นจริง เหมาะสำหรับการออกแบบระบบไฟฟ้าของบ้านพักที่อยู่อาศัย ทำให้ไม่ต้องมีการเผื่อขนาดอุปกรณ์ที่มากเกินความจำเป็น

2.3 การจำแนกประเภทข้อมูล (Classification) และอัลกอริทึม (Algorithm)

ในยุคปัจจุบันคงเลี่ยงมิได้ที่จะต้องพบกับคำว่า “ปัญญาประดิษฐ์” (Artificial Intelligence) ในแวดวงเทคโนโลยีสารสนเทศ มีการให้นิยามในหลากหลายบริบทไม่ว่าจะเป็นระบบที่คิดเสมือนมนุษย์ ระบบที่กระทำเสมือนมนุษย์ ระบบที่คิดอย่างมีเหตุผล หรือระบบที่กระทำอย่างมีเหตุผล ซึ่งสิ่งที่ทำให้เกิดสิ่งเหล่านี้ล้วนมีพื้นฐานมาจากอัลกอริทึมที่ถูกคิดค้นขึ้นมาจากอดีตนั่นเอง เมื่อกล่าวถึงอัลกอริทึมแล้วสามารถแบ่งออกได้เป็น 4 กลุ่มดังนี้[4]

1. การเรียนรู้แบบมีผู้สอน (Supervised learning) หมายถึง ข้อมูลตัวอย่างและผลลัพธ์ที่ผู้สอนต้องการป้อนเข้าสู่ระบบโดยมีเป้าหมายในการสร้างกฎทั่วไปที่สามารถเชื่อมโยงข้อมูลขาเข้ากับขาออกได้

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) หมายถึงข้อมูลตัวอย่างที่ไม่มีการทำฉลากใดๆและให้ระบบหาโครงสร้างของข้อมูลขาเข้าเอง

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) หมายถึง ระบบมีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนไปตลอดเวลาโดยระบบจะต้องทำงานบางอย่างโดยที่ไม่มีผู้สอนคอยบอกอย่างจริงจังว่าวิธีการที่ทำอยู่นั้นเข้าใกล้เป้าหมายแล้วหรือไม่

4. การเรียนรู้แบบกึ่งมีผู้สอน (Semi supervised learning) หมายถึง การเรียนรู้อีกแบบหนึ่งที่อยู่ระหว่างการเรียนรู้แบบมีผู้สอนกับการเรียนรู้แบบไม่มีผู้สอน โดยที่ผู้สอนจะไม่สอนอย่างสมบูรณ์ นั่นคือ บางข้อมูลในเซ็ตการสอนนั้นขาดข้อมูลขาออก

การจำแนกประเภทข้อมูล (Classification) เป็นกระบวนการทำเหมืองข้อมูลชนิดหนึ่งที่มีการเรียนรู้แบบมีผู้สอน (Supervised learning) โดยมีขั้นตอนหลักๆ อยู่ 2 ขั้นตอนคือ 1.การสร้างแบบจำลอง (Model Construction) โดยเซตของตัวอย่างที่ใช้สร้างแบบจำลองจะเรียกว่าชุดข้อมูลสอน (Training Set) ซึ่งแต่ละตัวอย่างจะมีคุณลักษณะซึ่งบอกค่าประเภทไว้ล่วงหน้า 2. การนำแบบจำลองที่ได้ไปใช้ (Model Usage) สำหรับการจำแนกประเภทตัวอย่างในอนาคต โดยจะต้องมีการประมาณค่าความแม่นยำ (accuracy) ของแบบจำลองที่ได้เสียก่อนการนำไปใช้ ซึ่งอัตราความแม่นยำ (accuracy rate) สามารถคำนวณได้ด้วยร้อยละการจำแนกประเภทได้อย่างถูกต้องของแบบจำลองที่สร้างขึ้น (การเปรียบเทียบค่าประเภทที่ทราบล่วงหน้าของตัวอย่างในชุดทดสอบ (Test Set) กับค่าผลลัพธ์ที่ได้จากการจำแนกประเภทที่ได้จากแบบจำลอง เพื่อจำแนกข้อมูลเป็นกลุ่มตามที่ได้กำหนดไว้) ซึ่งชุดข้อมูลทดสอบนั้นต้องเป็นอิสระไม่ขึ้นต่อชุดข้อมูลสอน มิฉะนั้นจะเกิด Overfitting ได้ วิธีการจำแนกประเภทข้อมูลนั้นมีด้วยกันหลายวิธี ซึ่งงานวิจัยนี้ได้เลือกมา 2 วิธีคือ นาอีฟเบย์ (Naïve Bayes) และ โครงข่ายประสาทเทียม (Neural Networks)

2.3.1 นาอีฟเบย์ (Naïve Bayes)

[4] ในทฤษฎีความน่าจะเป็น สถิติ การอนุมาน และปัญญาประดิษฐ์ บางครั้งมักพบคำว่าแบบเบย์ (Bayesian) มาขยายชื่อทฤษฎีหรือโมเดลต่างๆ โดยทุกครั้งที่พบคำขยายนี้มักหมายความว่าได้มีการนำปรัชญาหรือหลักการของทฤษฎีความน่าจะเป็นแบบเบย์มาใช้กับสาขาความรู้นั้นๆ ชื่อเรียก “แบบเบย์” เริ่มใช้ในช่วงปี ค.ศ. 1950 โดยมีต้นกำเนิดมาจากชื่อของ โทมัส เบย์ ผู้ซึ่งเสนอทฤษฎีบทของเบย์เป็นคนแรก ต่อมาปิแยร์ ซิมง ลาปลาซ ได้เสนอทฤษฎีบทของเบย์เช่นกัน โดยที่ลาปลาซไม่ทราบว่ามีงานของเบย์อยู่ ทฤษฎีบทของเบย์ในแบบลาปลาซถูกนำไปใช้งานอย่างกว้างขวางเนื่องจากการแปลความหมายความน่าจะเป็นของลาปลาซนั้นกว้างมาก โดยลาปลาซได้นำไปประยุกต์ใช้ในการแก้ปัญหาทางกลศาสตร์ ดาราศาสตร์ สถิติการแพทย์ หรือแม้กระทั่งนิติศาสตร์อีกด้วย ลาปลาซได้ใช้ทฤษฎีความน่าจะเป็นแบบเบย์ในการทำนายมวลของดาวเสาร์โดยใช้ข้อมูลวงโคจรของดาวเสาร์ในขณะนั้น โดยในการคำนวณของลาปลาซมีความผิดพลาดไปเพียงร้อยละ 0.63 เท่านั้น ผู้บุกเบิกทฤษฎีความน่าจะเป็นแบบเบย์ที่มีชื่อเสียงคนอื่นๆ เช่น จอห์น เมย์นาร์ด เคนส์, เลโอนาร์ด ซาเวจ, แฟรงค์ แรมซีย์, รูดอร์ฟ คาร์นาพ ซึ่งนักทฤษฎีความน่าจะเป็นแบบเบย์ที่โด่งดังในช่วงปี ค.ศ. 1930-1960 ก็คือ เดนนิส ลินด์ลีย์. เจนส์ ได้ให้ข้อสังเกตไว้ว่าผู้สนับสนุนทฤษฎีแบบเบย์ที่มีชื่อเสียงมักเป็นบุคคลจากสาขาอื่นที่ไม่ใช่คณิตศาสตร์ ไม่ว่าจะเป็นตัวเขา และคอกซ์ที่เป็นนักฟิสิกส์, เซอร์ แฮโรลด์ เจฟฟรีย์ที่เป็นนักธรณีวิทยา, เคนส์ที่เป็นนักเศรษฐศาสตร์ หรือ คาร์นาพที่เป็นนักปรัชญาวิทยาศาสตร์ ทั้งนี้อาจเป็นเพราะบุคคลเหล่านี้

ต้องการนำทฤษฎีความน่าจะเป็นไปใช้งานจริง และต่างก็พบว่าทฤษฎีความน่าจะเป็นเชิงความถี่ไม่กว้างขวางพอที่จะเอาไปใช้จริงได้ อีกทั้งสถิติเชิงความถี่ก็ยังมีความน่าเชื่อถือไม่เพียงพอ บุคคลเหล่านี้จึงต้องพัฒนาทฤษฎีความน่าจะเป็นที่สามารถนำไปใช้ได้จริงขึ้นและต่างก็ค้นพบแนวทางเดียวกันซึ่งก็คือสิ่งที่ลาปลาสได้แสดงไว้แล้วเมื่อราวต้นคริสต์ศตวรรษที่ 19

การเรียนรู้แบบเบย์ (Bayesian Learning) เป็นวิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานจากทฤษฎีของเบย์ (Bayes theorem) เข้ามาช่วยในการเรียนรู้ โดยมีจุดมุ่งหมายเพื่อต้องการสร้างโมเดลที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำโมเดลมาหาว่าสมมติฐานใดถูกต้องที่สุดโดยใช้ความน่าจะเป็นเข้ามาช่วย ความรู้ก่อนหน้า (Prior knowledge) หมายถึงความรู้ที่เรามีเกี่ยวกับสมมติฐานแต่ละตัวก่อนที่เราจะเก็บข้อมูล เมื่อใช้งานเราจะนำความน่าจะเป็นของข้อมูลที่เก็บได้มาปรับสมมติฐานซ้ำอีกครั้ง ข้อดีของวิธีการเรียนรู้แบบนี้คือ สามารถใช้ข้อมูลและความรู้ก่อนหน้าเข้ามาช่วยในการเรียนรู้ได้ ซึ่งพบว่าวิธีนี้ให้ประสิทธิภาพในการเรียนรู้ได้ดีไม่น้อยกว่าวิธีการเรียนรู้ประเภทอื่น เมื่อกล่าวถึงความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้น (A) ถ้ามีเหตุการณ์อีกเหตุการณ์หนึ่งเกิดมาแล้ว (B) สามารถเขียนให้อยู่ในรูปอย่างง่ายดังสมการ

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \qquad (2.6)$$

$P(A \mid B)$ คือ ความน่าจะเป็นที่เหตุการณ์ A จะเกิดขึ้นถ้าเหตุการณ์ B เกิดขึ้นแล้ว

$P(B \mid A)$ คือ ความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้นถ้าเหตุการณ์ A เกิดขึ้นแล้ว

$P(A)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A

$P(B)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ B

วิธีการเรียนรู้เบย์อย่างง่าย (Naïve Bayesian Learning) เป็นวิธีจำแนกข้อมูลที่มีประสิทธิภาพวิธีหนึ่งโดยที่ใช้งานได้ดี เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน มีการนำการจำแนกประเภทของข้อมูลแบบเบย์อย่างง่ายไปประยุกต์ใช้ในการแยกประเภทข้อความ (Text Classification) การวินิจฉัย (Diagnosis) และพบว่าใช้งานได้ดีไม่ต่างจากการจำแนกประเภทด้วยวิธีการอื่นทำให้ผู้วิจัยเลือกวิธีการนี้มาใช้ในงานวิจัย เนื่องจากเป็นวิธีการจำแนกข้อมูลที่มีประสิทธิภาพและมีอัลกอริทึมในการทำงานที่ไม่ซับซ้อน กำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นกลุ่ม Vj สำหรับข้อมูลที่มีคุณสมบัติ n ตัว

$X = \{A_1, A_2, \dots, A_n\}$ หรือใช้สัญลักษณ์ว่า $P(A_1, A_2, \dots, A_n|V_j)$

$$P\left(A_1, A_2, \dots, A_n\middle|V_j\right) = \prod_{i=1}^{n} P(A_i|V_j) \quad (2.7)$$

โดยที่ $\prod$ หมายถึงผลคูณของค่า $P(A_i|V_j)$ ทั้งหมด i = 1,2,3,...,n และ j = 1,2,3,...,n

การนำวิธีการเรียนรู้เบย์อย่างง่ายไปใช้ มีวิธีการดังต่อไปนี้

1. หาความน่าจะเป็นของสิ่งที่พบในในแต่ละกลุ่มโดยนำค่า $P\left(A_1, A_2, \dots, A_n\middle|V_j\right) = \prod_{i=1}^{n} P(A_i|V_j)$ จากสมการ(7) คูณกับค่าความน่าจะเป็นของกลุ่มคือ $P(V_j)$ จะได้เท่ากับ V_{NB}

2. นำค่าที่ได้มาเปรียบเทียบกัน กลุ่มที่มีความน่าจะเป็นสูงสุดคือคำตอบ

ดังนั้นจะได้วิธีการจำแนกประเภทแบบนาอีฟเบย์ดังสมการ

$$V_{NB} = arg\,max P(V_j) \times \prod_{i=1}^{n} P(A_i|V_j) \quad (2.8)$$

นาอีฟเบย์จัดเป็นเทคนิคในการแก้ปัญหาแบบมีผู้สอน (Supervised Learning) ที่สามารถคาดการณ์ผลลัพธ์และสามารถอธิบายได้ด้วยการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ดังสมการ

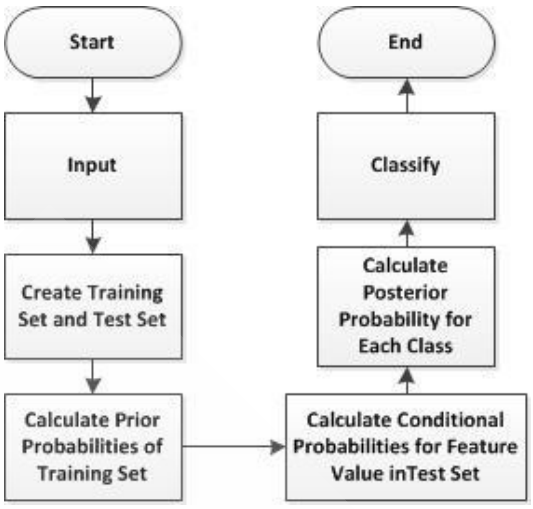
$$P(X_1, X_2, \cdots, X_n | C) = P(X_1 | C) \cdot P(X_2 | C) \cdots P(X_n | C) \quad (2.9)$$

และเมื่อผลคูณของความน่าจะเป็นจากเหตุการณ์ใดมีค่ามากกว่า จะถือว่าเหตุการณ์นั้นคือผลลัพธ์ของนาอีฟเบย์ดังอสมการ

$$[P(x_1 | c^*) \cdots P(x_n | c^*)]P(c^*) > [P(x_1 | c) \cdots P(x_n | c)]P(c) \quad (2.10)$$

เมื่อ $c \neq c^*, c = c_1, \cdots, c_L$

จะสรุปได้ว่าผลลัพธ์ของนาอีฟเบย์นั้นคือ c^* สามารถอธิบายหลักการทำงานอย่างง่ายได้ดังภาพที่ 2.4



ภาพที่ 2.4 Naive Bayes Flow Chart

เพื่อให้เข้าในการสร้างนาอีฟเบย์โมเดลมากยิ่งขึ้นนั้นจึงยกตัวอย่างของ Data set ของสภาพอากาศ ซึ่งประกอบไปด้วย outlook, temperature, humidity และ windy สำหรับการทำนายเงื่อนไขของการเล่นเทนนิสดังดังตารางที่ 2.1 จากนั้นสร้างนาอีฟเบย์โมเดลจากข้อมูลที่ให้มาซึ่งจะได้ผลลัพธ์ดังตารางที่ 2.2

ตารางที่ 2.1 แสดงชุดข้อมูลของสภาพอากาศ

No	outlook	temperature	humidity	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	mild	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

ตารางที่ 2.2 แสดงผลการสร้างนาอีฟเบย์โมเดล

outlook			temperature			humidity			windy			play	
	yes	no		yes	no		yes	no		yes	no	yes	no
Sunny	2	3	Hot	2	2	High	3	4	FALSE	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	TRUE	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/9	Hot	2/9	2/5	High	3/9	4/5	FALSE	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	TRUE	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

จากตารางแสดงให้เห็นความถี่ที่เกิดขึ้นของเหตุการณ์แต่ละส่วนที่ใช้ในการทำนายว่าจะเล่นเทนนิสหรือไม่ ยกตัวอย่างเช่นจาก Data set ทั้งหมดแสดง outlook=sunny เมื่อ play=yes ดังนี้

$$P(outlook = sunny|play = yes) = \frac{2}{9}$$

$$P(play = yes) = \frac{9}{14}$$

เมื่อได้นาอีฟเบย์โมเดลมาแล้ว จะสามารถทำนายเหตุการณ์ของการเล่น “play” ได้จากหลักฐานต่างๆที่ได้เตรียมไว้แล้ว จากตัวอย่างกำหนดให้เหตุการณ์ E มีสภาพอากาศดังนี้ outlook = sunny, temperature = cool, humidity = high และ windy = TRUE แล้วสามารถหาค่าความน่าจะเป็นได้ดังนี้

$$P[yes|E] = P[outlook = sunny|yes] \times P[temperature = cool|yes] \times P[humidity = high|yes] \times P[windy = TRUE|yes] \times \frac{P[yes]}{P[E]}$$

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{P[E]}$$

ในที่นี้สามารถมองข้าม P[E] ได้เนื่องจากเราเพียงต้องการหาค่าความสัมพันธ์ (relatively) เพื่อเปรียบเทียบค่าระหว่าง 2 Class ซึ่งได้ผลดังนี้

Likelihood of two class

For “yes” = 2/9 * 3/9 * 3/9 * 3/9 * 9/14 = 0.0053

For “no” = 3/5 * 1/5 * 4/5 * 3/5 * 5/14 = 0.0206

Conversion into a probability by normalization

P(“yes”) = 0.0053/ (0.0053+0.0206) = 0.205

P(“no”) = 0.0206/ (0.0053+0.0206) = 0.795

สังเกตได้จากนาอีฟเบย์โมเดลข้างต้น จะเห็นได้ว่า P(outlook = overcast|play = no) = 0/5 ซึ่งจะทำให้เกิดปัญหาเมื่อหาค่า P(“no”) เนื่องจากผลลัพธ์ที่ได้จะเท่ากับ 0 สามารถแก้ปัญหาโดยใช้เทคนิคที่เรียกว่า “smoothing technique” ซึ่งหนึ่งในวิธีการ smoothing technique นั้นคือ Laplace estimation

$$P(outlook = sunny|play = no) = \frac{3 + \mu p_1}{5 + \mu}$$

$$P(outlook = overcast|play = no) = \frac{0 + \mu p_2}{5 + \mu}$$

$$P(outlook = rainy|play = no) = \frac{2 + \mu p_3}{5 + \mu}$$

โดยที่ $p_1 + p_2 + p_3 = 1.0$ และ ตามที่ได้มีการกำหนดค่าให้ข้อมูลทั้งหมดมีการกระจายแบบเท่ากัน จะได้ $p_1 = p_2 = p_3 = 1/3$

$$P(outlook = sunny|play = no) = \frac{3 + \mu/3}{5 + \mu} = \frac{3 + 3/3}{5 + 3} = \frac{4}{8}$$

$$P(outlook = overcast|play = no) = \frac{0 + \mu p_2}{5 + \mu} = \frac{0 + 3/3}{5 + 3} = \frac{1}{8}$$

$$P(outlook = rainy|play = no) = \frac{2 + \mu p_3}{5 + \mu} = \frac{2 + 3/3}{5 + 3} = \frac{3}{8}$$

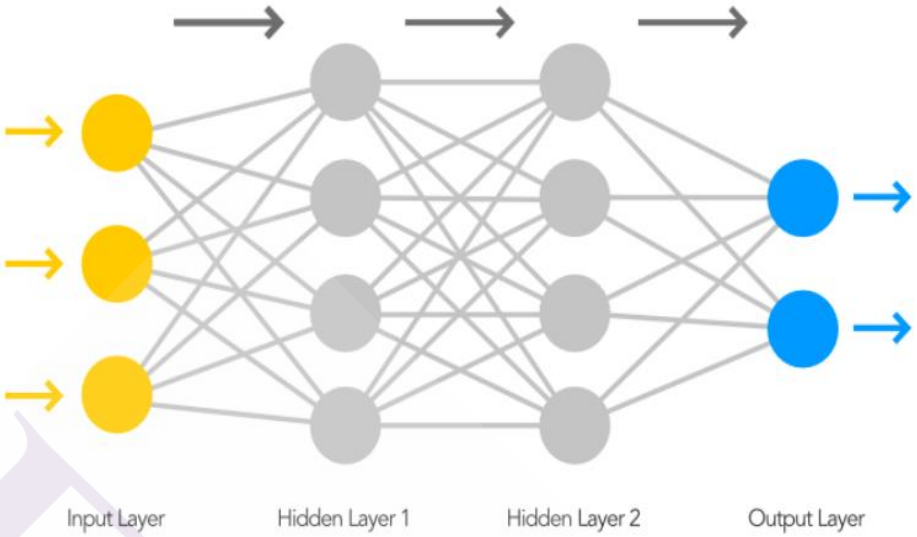
นาอีฟเบย์จึงเป็นการจำแนกข้อมูลวิธีหนึ่งที่มีประสิทธิภาพและมีการทำงานที่ไม่ซับซ้อน เหมาะกับกรณีของเซ็ตตัวอย่างที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างที่ไม่ขึ้นต่อกัน ที่ผ่านมานั้นมีการนำนาอีฟเบย์ไปประยุกต์ใช้งานกันอย่างแพร่หลาย เช่น การกรองข้อความภาษาไทยและ

ภาษาอังกฤษของบริการส่งข้อความสั้นบนเครือข่ายโทรศัพท์เคลื่อนที่ การจำแนกสายพันธุ์ยางพารา เป็นต้น

2.3.2 โครงข่ายประสาทเทียม (Neural Networks)

[10] ข่ายงานประสาทเทียม (Artificial Neural Network) คือ โมเดลทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลข้อมูลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) ซึ่งมีแนวคิดเริ่มต้นมาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ(Bioelectric Network) ในสมองซึ่งประกอบด้วยเซลล์ประสาท (Neurons) และ จุดประสานประสาท (Synapses) ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทที่ทำงานร่วมกัน นักวิจัยส่วนใหญ่ในปัจจุบันเห็นว่าข่ายงานประสาทเทียมมีโครงสร้างแตกต่างจากข่ายงานในสมอง แต่ก็ยังเหมือนกับสมองในแง่ที่ว่าข่ายงานประสาทเทียมคือการรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อยๆ และการเชื่อมต่อนี้เป็นส่วนสำคัญที่ทำให้เกิดสติปัญญาของข่ายงาน เมื่อพิจารณาขนาดแล้ว สมองมีขนาดใหญ่กว่าข่ายงานประสาทเทียมอย่างมาก รวมทั้งเซลล์ประสาทยังมีความซับซ้อนกว่าหน่วยย่อยของข่ายงาน อย่างไรก็ดีหน้าที่สำคัญของสมอง เช่นการเรียนรู้ ยังคงสามารถถูกจำลองขึ้นอย่างง่ายด้วยข่ายงานประสาท

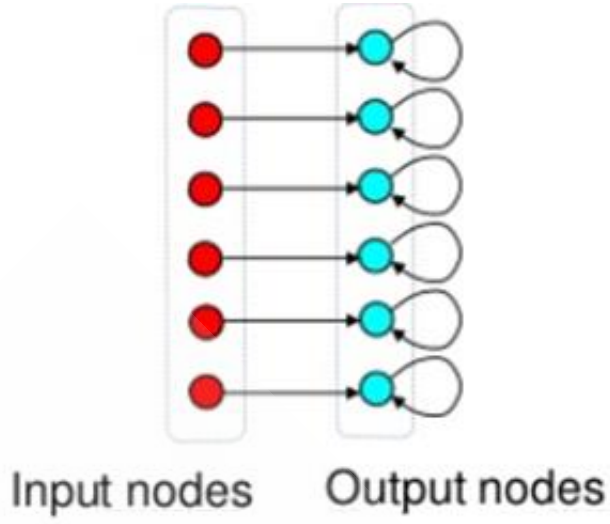
ข่ายงานประสาทสามารถแบ่งตามสถาปัตยกรรมมได้เป็น 2 ประเภท คือ โครงข่ายการส่งข้อมูลแบบไม่ย้อนกลับ (Feed forward Network) ประกอบด้วยเซตของบัพ (node) ซึ่งอาจจะถูกกำหนดให้เป็นโหนดขาเข้า (input nodes) โหนดขาออก (output nodes) หรือ โหนดระหว่างกลาง (hidden nodes) มีการเชื่อมต่อระหว่างโหนด (หรือนิวรอน) โดยกำหนดค่าน้ำหนัก (weight) กำกับอยู่ที่เส้นเชื่อมทุกเส้น เมื่อข่ายงานเริ่มทำงาน จะมีการกำหนดค่าให้แก่โหนดขาเข้าซึ่งค่าเหล่านี้อาจจะได้มาจากการกำหนดโดยมนุษย์ เซนเซอร์ที่วัดค่าต่างๆ หรือผลจากโปรแกรมอื่นๆ ก็ได้ จากนั้นโหนดขาเข้าจะส่งค่าที่ได้รับไปตามเส้นเชื่อมขาออกโดยที่ค่าที่ส่งออกไปจะถูกคูณกับค่าน้ำหนักของเส้นเชื่อม ต่อมาโหนดในชั้นถัดไปจะรับค่าซึ่งเป็นผลรวมจากโหนดต่างๆ แล้วจึงคำนวณผลอย่างง่าย โดยทั่วไปจะใช้ฟังก์ชันซิกมอยด์ (sigmoid function) แล้วส่งค่าไปยังชั้นถัดไป โดยไม่มีการย้อนกลับหรือแม้แต่โหนดในชั้นเดียวกันก็ยังไม่มีการเชื่อมต่อถึงกัน การคำนวณเช่นนี้จะเกิดขึ้นไปเรื่อยๆ ทีละชั้นจนถึงโหนดขาออก ดังภาพที่ 2.5



ภาพที่ 2.5 โครงข่ายการส่งข้อมูลแบบไม่ย้อนกลับ (Feedforward Network)

ที่มา: https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic

โครงข่ายการส่งข้อมูลแบบย้อนกลับ (Feedback Network) เป็นสถาปัตยกรรมที่กำหนดให้มีการส่งข้อมูลจากข้อมูลในชั้นข้อมูลขาเข้า (Input Layer) เข้ามาภายในชั้นซ่อน (Hidden Layer) และส่งไปยังข้อมูลขาออก (Output Layer) จะมีทิศทางในการไหลของข้อมูลไปในทิศทางเดียวกัน ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจากชั้นขาเข้าส่งต่อมาเรื่อยๆ จนถึงชั้นขาออกโดยมีการย้อนกลับของข้อมูลดังภาพที่ 2.6 การส่งผ่านย้อนกลับนั้นค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด (Error-Correction) คือผลต่างของผลตอบที่แท้จริง (Actual Response) กับผลตอบเป้าหมาย (Target Response) เกิดเป็นสัญญาณผิดพลาด (Error Signal) ซึ่งสัญญาณผิดพลาดนี้จะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ และค่าน้ำหนักของการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย เพื่อให้การจำแนกนั้นมีประสิทธิภาพมากขึ้นนั้นเอง



ภาพที่ 2.6 โครงข่ายการส่งข้อมูลแบบย้อนกลับ (Feedback Network)

ที่มา: https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic

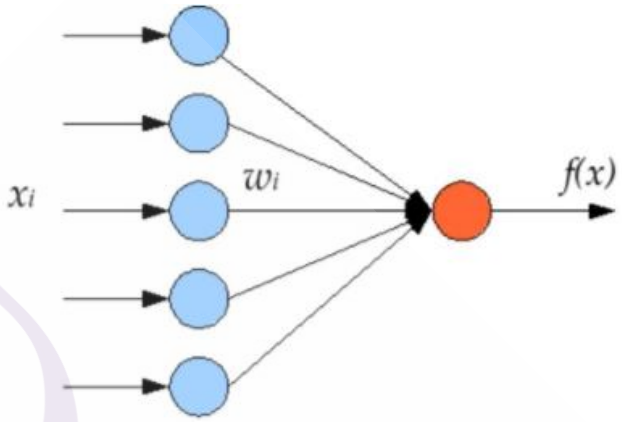
นอกจากนี้ยังสามารถแบ่งประเภทข่ายงานประสาทตามระดับชั้นของโครงข่ายได้เป็นสองประเภท คือ เพอร์เซปตรอนแบบชั้นเดียว (Single-layer Perceptron : SLP) และเพอร์เซปตรอนแบบหลายชั้น (Multi-layer Perceptron : MLP)

เพอร์เซปตรอนแบบชั้นเดียว (SLP) เป็นโครงข่ายประสาทเทียมแบบง่ายมีหน่วยเดียวที่จำลองลักษณะของเซลล์ประสาท (neuron) ของมนุษย์ ซึ่งถือเป็นหน่วยย่อยที่สุดของข่ายงานประสาทเทียมดังภาพที่ 2.7 ทำหน้าที่รับข้อมูล Input เป็นเวกเตอร์จำนวนจริงเข้ามาโดยจำนวนข้อมูล Input จะขึ้นอยู่กับจำนวนแอทริบิวต์ของข้อมูลขาเข้า (Xi) มีค่าน้ำหนักของข้อมูล Input แต่ละตัว (Wi) แล้วคำนวณหาผลรวมเชิงเส้น (linear combination) และจะต้องมีฟังก์ชันกระตุ้น (Activation function) ขึ้นอยู่กับลักษณะข้อมูล Output เช่น ถ้าข้อมูล Output ที่ต้องการมีลักษณะเป็น “ใช่” หรือ “ไม่ใช่” จะต้องใช้ Threshold function แต่ถ้าหากเป็นค่าตัวเลขที่ต่อเนื่องต้องใช้ Continuous function จำพวก Sigmoid function ดังสมการ

$$f(x) = \begin{cases} 1 \; if \; x \geq T \\ 0 \; if \; x < T \end{cases} \tag{2.11}$$

$$f(x) = \frac{1}{1+e^{-\alpha x u}} \tag{2.12}$$

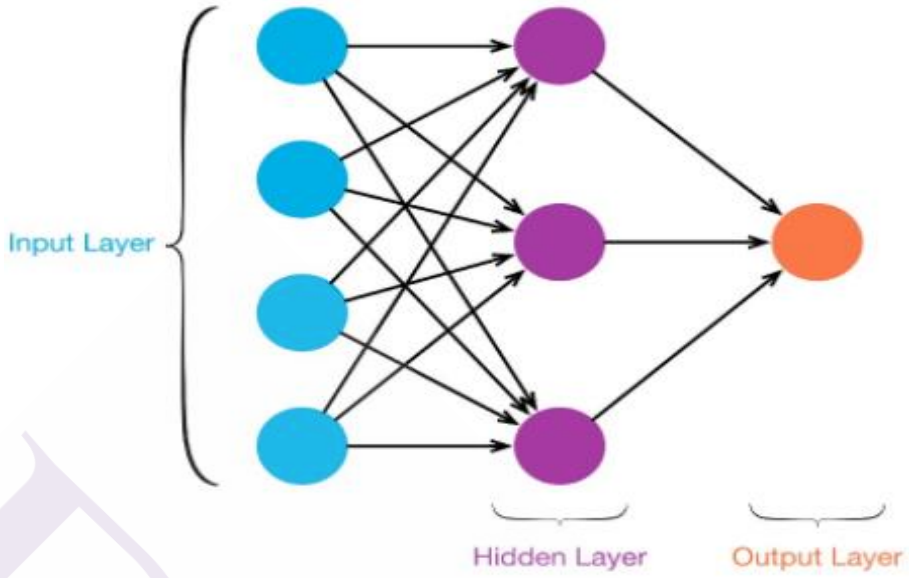
ซึ่งมีข้อจำกัดคือ Concept หรือ Function ที่สามารถเรียนได้โดยอาศัยเพอร์เซปตรอนแบบชั้นเดียว จะต้องเป็นฟังก์ชันที่สามารถแยกเชิงเส้นได้ (linearly separate function) เท่านั้น



ภาพที่ 2.7 เพอร์เซปตรอนแบบชั้นเดียว (Single-layer Perceptron : SLP)

ที่มา: https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic

เพอร์เซปตรอนแบบหลายชั้น (multi-layer perceptron : MPL) เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบหลายๆชั้น ใช้สำหรับงานที่มีความซับซ้อนได้ผลเป็นอย่างดี โดยมีกระบวนการฝึกฝนเป็นแบบมีผู้สอน (Supervise) และใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) ดังภาพที่ 2.7



ภาพที่ 2.8 เพอร์เซปตรอนแบบหลายชั้น (Multi-layer Perceptron : MPL)

ที่มา: https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic

สัญญาณที่มีโครงข่ายประสาทเทียมแบบ MLP มี 2 ประเภทคือ Function Signal และ Error Signal

1. Function Signal เป็นสัญญาณขาเข้าที่มาจากโหนดในชั้นก่อนหน้า และจะส่งผ่านไปข้างหน้าจากโหนดหนึ่งไปสู่อีกโหนดหนึ่ง

2. Error Signal เป็นสัญญาณย้อนกลับที่เกิดขึ้นที่โหนดในชั้นข้อมูลขาออกของโครงข่ายประสาทเทียม และถูกส่งผ่านย้อนกลับจากชั้นหนึ่งไปสู่อีกชั้นหนึ่ง

หลักการทำงานของ MLP คือในแต่ละชั้นของชั้นซ่อนตัว (Hidden Layer) จะมีฟังก์ชันสำหรับคำนวณเมื่อได้รับสัญญาณ (Output) จากโหนดในชั้นก่อนหน้านี้ที่เรียกว่า Activation Function โดยในแต่ละชั้นไม่จำเป็นต้องเป็นฟังก์ชันเดียวกันก็ได้ ชั้นซ่อนตัวนั้นจะมีหน้าที่สำคัญคือพยายามแปลงข้อมูลที่เข้ามาในชั้น (Layer) นั้นๆให้สามารถแยกแยะความแตกต่างได้โดยใช้ฟังก์ชันที่สามารถแยกเชิงเส้นได้ (linearly separate function) และก่อนที่ข้อมูลจะถูกส่งไปถึงชั้นข้อมูลออก (Output Layer) ในบางครั้งอาจจำเป็นต้องใช้ชั้นซ่อนตัวมากกว่า 1 ชั้นในการแปลงข้อมูลให้อยู่ในรูป Linearly Separable ในการคำนวณหา Output เพื่อแก้ปัญหาการจำแนกทำได้โดยการใส่ข้อมูล Input เข้าไปในโครงข่ายประสาทเทียมที่เราได้ทำการหาไว้แล้ว จากนั้นให้ทำการเปรียบเทียบค่าของ Output ใน Output Layer และให้ทำการเลือกค่าของ Output ที่มีค่าสูงกว่า

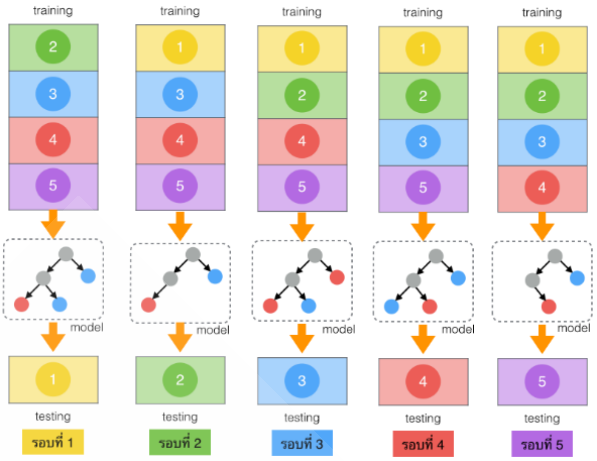
(Neuron ที่มีค่าสูงกว่า) และทำการรับค่าของพยากรณ์ที่ตรงกับ Neuron ที่เลือก จากนั้นให้นำค่าของ Output มาเปรียบเทียบกับค่าที่ยอมรับได้ หากค่าของ อยู่ในช่วงที่รับได้ (Error น้อยกว่า Error ที่เรากำหนด) ก็ให้ทำการรับข้อมูลชุดถัดไป แต่หากค่าของ มากกว่าค่าที่ยอมรับได้ ให้ทำการปรับค่าน้ำหนักและ Biased ตามขั้นตอนที่ได้กล่าวไว้ข้างต้น เมื่อทำการปรับน้ำหนักเรียบร้อยแล้ว ให้ทำการรับข้อมูลชุดถัดไปและทำตามขั้นตอนซ้ำอีกรอบจนกระทั่งถึงข้อมูลชุดสุดท้าย และเมื่อทำข้อมูลชุดสุดท้ายเสร็จจะนับเป็น 1 รอบของการคำนวณ (1 Epoch) จากนั้นจะทำการหาค่าผิดพลาดรวมเฉลี่ย จากค่าเฉลี่ยที่ได้เก็บค่าเอาไว้เพื่อใช้ในการตรวจสอบว่าค่าโดยเฉลี่ยในการจำแนกนั้น มีค่าน้อยกว่าค่าผิดพลาดที่ยอมรับได้หรือไม่ ถ้าใช่แสดงว่าโครงข่ายประสาทเทียมที่สร้างขึ้นนั้นสามารถให้ผลลัพธ์ที่ถูกต้องของทุกๆ ข้อมูลแล้ว จึงทำการจบการเรียนรู้ได้ แต่ถ้าไม่ใช่ให้กลับไปทำตามขั้นตอนแรกโดยเริ่มรับข้อมูลชุดที่ 1 ใหม่

2.3.2 การวัดประสิทธิภาพของโมเดลการจำแนกข้อมูล

การวัดประสิทธิภาพของโมเดลที่นิยมใช้ประกอบไปด้วย

1. วิธีแยกทดสอบ (Split Test) เพื่อใช้วัดประสิทธิภาพการจำแนกข้อมูลสำหรับชุดข้อมูลทดสอบ โดยจะแบ่งชุดข้อมูลโดยวิธีการสุ่มเป็น 2 ส่วน เช่น 70% ต่อ 30% โดย 70% เป็นชุดข้อมูลสำหรับทำการสอนและ 30% เป็นชุดข้อมูลสำหรับทำการทดสอบ

2. วิธีการตรวจสอบไขว้ (Cross-validate Test) ใช้สร้างแบบจำลองเพื่อวัดประสิทธิภาพของชุดข้อมูลฝึกสอนโดยจะแบ่งข้อมูลออกเป็นหลายส่วน k-fold cross-validation โดยจะแบ่งข้อมูลออกเป็น k ชุด ชุดละเท่าๆกัน และในการวนรอบทดสอบจะมีการใช้ชุดข้อมูล 1 ชุดสำหรับทดสอบและอีก k-1 ชุดใช้สำหรับสอน เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ โดยจะมีการนำผลลัพธ์ของแต่ละรอบมาคำนวณหาค่าเฉลี่ยร่วมกันเพื่อใช้เป็นค่าประสิทธิภาพของแบบจำลอง Cross-validation Test เป็นวิธีการทดสอบประสิทธิภาพของโมเดลที่นิยมในการทำงานวิจัยเนื่องจากผลที่ได้มีความน่าเชื่อถือ



ภาพที่ 2.9 ลักษณะการทำ Cross-validation

ที่มา: http://dataminingtrend.com

3. การวัดประสิทธิภาพของโมเดลจากสัดส่วนการจำแนกข้อมูลที่ถูกต้องจากการจำแนกทั้งหมด

$$\text{ร้อยละความถูกต้อง} = \frac{\text{TA}}{\text{TA+FA}} \times 100$$

โดย TA คือ จำนวนตัวอย่างที่อยู่ใน Class A และทำนายได้ว่าอยู่ใน Class A

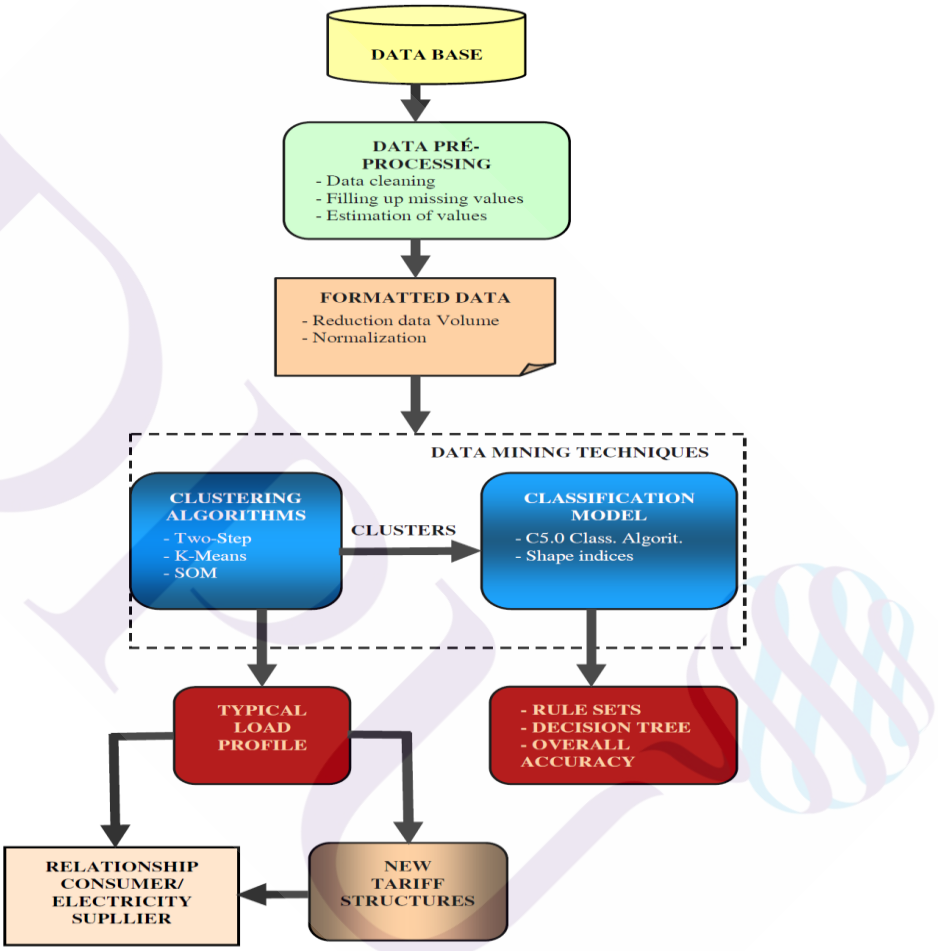
FA คือ จำนวนตัวอย่างที่อยู่ใน Class A แต่ทำนายได้ว่าอยู่ใน Class อื่นๆ ที่ไม่ใช่ Class A

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 การประยุกต์เหมืองข้อมูลเพื่อจำแนกกลุ่มผู้ใช้ไฟฟ้า

[13] Sergio Ramos, Zita Vale, “Data Mining Techniques to Support the Classification of MV Electricity Customer”, IEEE2008:1-7, เป็นงานวิจัยที่ได้นำเสนอวิธีการจำแนกกลุ่มผู้ใช้ไฟฟ้าโดยอาศัยข้อมูลการใช้ไฟฟ้าทุก 15 นาที เป็นเวลา 3 เดือนของผู้ใช้ไฟฟ้าจำนวน 229 ราย โดยมีแผนผังการทำ Classification ดังภาพที่ 2.9 แผนผังการทำ Classification ของงานวิจัยที่นำเสนอ โดยใช้อัลกอริทึ่ม 3 ประเภท ได้แก่ Two-Step Cluster Analysis, K-means, และ Kohonen Net – Self Organizing Features Map เพื่อให้ได้โหลดไดอะแกรมสำหรับผู้ใช้ไฟฟ้าแต่ละรายซึ่งเมื่อเปรียบเทียบประสิทธิภาพจาก Mean Index Adequacy (MIA) และ Clustering Dispersion Indicator (CDI) แล้วงานวิจัยนี้ได้เลือกใช้ Two-Step Cluster Analysis เพื่อให้ได้โหลดไดอะแกรม

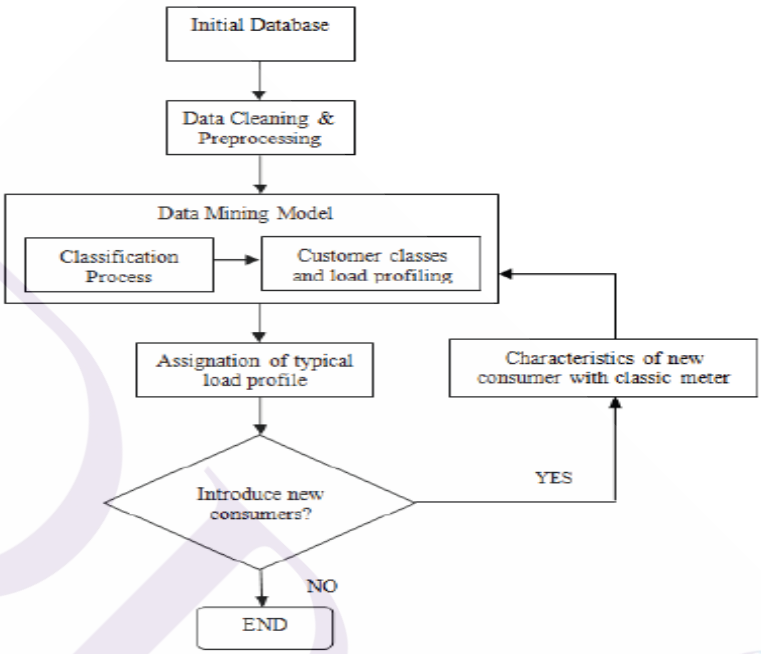
สำหรับกรณีศึกษา จากนั้นทำการสกัดข้อมูลให้ได้ Load Shape Index สำหรับ Classification Model ซึ่งใช้อัลกอริทึ่ม C5.0 classification ซึ่งได้ออกมาเป็น 9 Cluster มีความเที่ยงตรงของการจำแนกเป็นที่น่าพอใจ คือ 94% สำหรับวันทำการ และ 95% สำหรับวันหยุด



ภาพที่ 2.10 แผนผังการทำ Classification ของงานวิจัยที่นำเสนอ

[7] Gheorghe Grigoras, Ovidiu Ivanov, and Mihai Gavrilas, “Customer Classification and Load Profiling using Data from Smart Meters”,NEUREL2014 งานวิจัยนี้นำเสนอโมเดลสำหรับจำแนกประเภทของผู้ใช้ไฟฟ้าในระบบจำหน่ายที่อยู่ห่างไกล จำนวน 180 ราย ด้วยวิธีการ Classification และ Load Profiling โดยใช้ข้อมูลจาก Smart Meters สำหรับหาลักษณะเฉพาะของ

ผู้ใช้ไฟฟ้าในแต่ละประเภท ซึ่งกำหนด Class ตามข้อมูลหลักคือ ปริมาณการใช้ไฟฟ้ารายวัน รายเดือน (รวม, สูงสุด และต่ำสุด) และมีการใช้นิวรัลเน็ทเวิร์คจำแนกผู้ใช้ไฟฟ้าดังกล่าวออกจากกันเป็น 5 กลุ่มที่มีลักษณะการใช้พลังงานแตกต่างกัน



ภาพที่ 2.11 แผนผังอัลกอริทึมของงานวิจัยที่นำเสนอ

2.4.2 การใช้นาอีฟเบย์ในการแก้ปัญหาต่างๆ

[3] นนท์ บุญนิธิประเสริฐ และ ดร. ชัยพร เขมะภาตะพันธ์, “Short Message Service Filtering for Thai & English Language on Mobile Phone Network”, NCCIT2009:34-39 งานวิจัยนี้ได้นำเสนอวิธีการกรองสแปมในระบบส่งข้อความในระบบโทรศัพท์เคลื่อนที่ ซึ่งในระบบการส่งข้อความสั้น Short Message Service (SMS) ของประเทศไทยนั้นยังไม่มีการศึกษาและพัฒนาอย่างจริงจัง ซึ่งปัญหา Spam ในระบบส่งข้อความของผู้ให้บริการโทรศัพท์เคลื่อนที่มีความรุนแรงเพิ่มมากขึ้น จึงได้ทำการศึกษาการกรองข้อความ Spam ด้วยวิธีการกรองแบบ Support Vector Machine (SVM) และ Naïve Bayesian (NB) โดยผลทดสอบด้วยวิธีการกรองแบบ SVM มีความถูกต้องในการกรองข้อมูลสูงกว่า NB แต่วิธีการกรองแบบ NB ใช้เวลาในการประมวลผลน้อยกว่า

[5] ภูรดา นนทวาสี, “การแบ่งกลุ่มข้อความ SMS ตามลักษณะการให้บริการ”, วิทยานิพนธ์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม มหาวิทยาลัยธุรกิจบัณฑิตย์ ได้นำเสนอการแบ่งประเภทข้อความ SMS โดยใช้วิธีการ Naïve Bayesian โดยพิจารณาจากเนื้อหาของข้อความ เพื่อใช้สำหรับแก้ปัญหาความคับคั่งของ SMSC เมื่อมีปริมาณผู้ส่งจำนวนมาก โดยมีประโยชน์ต่อระบบในการสร้างลำดับการส่งใหม่ตามระดับความสำคัญของข้อความ ซึ่งช่วยลดความเสี่ยงที่ระบบจะเกิด Overload จนไม่สามารถให้บริการได้ อีกทั้งยังสามารถแยกระดับความสำคัญของข้อความ SMS เพื่อไปใช้งานเชิงพาณิชย์สำหรับผู้ใบริการโทรศัพท์เคลื่อนที่ต่อไปในอนาคต ซึ่งวิธีที่นำเสนอนี้สามารถคัดแบ่งประเภทข้อความโดยใช้เวลาในการทำงานน้อยกว่าแบบเดิมถึง 6% และยังมีความถูกต้องสูงกว่าถึง 13.59%

สรุป

จากงานวิจัยที่ได้ศึกษามายังมีส่วนที่สามารถพัฒนาต่อได้ดังนี้คือ การจำแนกประเภทของผู้ใช้ไฟฟ้าโดยอาศัยข้อมูลการใช้ไฟฟ้านั้นสามารถกระทำได้หลายวิธีทั้งการทำเหมืองข้อมูลและประยุกต์ใช้เทคนิคต่างๆ เพื่อเสริมให้กระบวนการมีประสิทธิภาพที่ดียิ่งขึ้น แต่ยังมีปัจจัยหลากหลายที่จำเป็นต้องพิจารณา เช่น ลักษณะเชิงพื้นที่ ข้อกำหนดหรือกฎหมายของแต่ละประเทศ แม้กระทั่งสภาพอากาศก็ล้วนแต่เป็นตัวแปรที่ทำให้ข้อมูลตั้งต้นแตกต่างกันออกไป ด้วยเหตุนี้ผู้วิจัยจึงเลือกที่จะดำเนินการกับข้อมูลการใช้ไฟฟ้าในขอบเขตของประเทศไทย นอกจากนี้แล้วยังมีส่วนของการนำตัวแบบนาอีฟเบย์และโครงข่ายประสาทเทียมมาประยุกต์ในงานวิจัยนี้ด้วย

บทที่ 3
การดำเนินงาน

เนื้อหาในส่วนนี้จะประกอบไปด้วยแนวทางการวิจัยและพัฒนา แผนการดำเนินงาน เครื่องมือที่ใช้ในการวิจัย การศึกษารูปแบบข้อมูลแต่ละประเภท รวมถึงออกแบบและเตรียมข้อมูล

3.1 แนวทางการวิจัยและพัฒนา

ประเภทกิจการของผู้ใช้ไฟฟ้าแต่ละรายของ กฟภ. นั้นเป็นการระบุข้อมูลตั้งต้นตั้งแต่เริ่มมีการขอใช้ไฟฟ้า โดยที่ไม่ว่าเวลาจะผ่านไปนานสักเพียงใด หากไม่มีการร้องขอเพิ่ม-ลดขนาดอุปกรณ์ก็จะไม่มีการตรวจสอบข้อมูลประเภทกิจการใหม่ ด้วยเหตุนี้ กฟภ. จึงไม่สามารถรู้ได้ว่าผู้ใช้ไฟฟ้ารายใดมีรูปแบบการใช้ไฟฟ้าที่แตกต่างไปจากเดิม ทำให้ยากต่อบริหารจัดการการจ่ายพลังงานไฟฟ้าอย่างเหมาะสม ผู้วิจัยจึงได้นำข้อมูลโหลดโปรไฟล์ที่ได้จากระบบ AMR ที่มีการอัพเดตทุก 15 นาทีมาทำการวิเคราะห์เพื่อให้เกิดการจำแนกประเภทกิจการให้เป็นปัจจุบัน ซึ่งการวิเคราะห์นี้ประกอบด้วย 2 ส่วนคือ

1. การวิเคราะห์ข้อมูลการใช้ไฟฟ้า (Load Profile) สำหรับหาลักษณะประจำ (Attribute) ของกิจการแต่ละประเภท เนื่องจาก Attribute เป็นเสมือนอัตลักษณ์บ่งชี้ลักษณะเฉพาะ และยังเป็นปัจจัยสำคัญที่จะทำให้ผลลัพธ์การจำแนกมีความถูกต้องจำเป็นต้องใช้วิธีการที่ก่อให้เกิดผลลัพธ์ที่เชื่อถือได้

2. การประยุกต์ใช้เทคนิคการจำแนกข้อมูล (Classification) โดยงานวิจัยนี้ได้เลือกเทคนิคที่ได้รับความนิยมในใช้งาน ได้แก่ นาอีฟเบย์ (Naïve Bayes) และโครงข่ายประสาทเทียม (Neural Network) สำหรับพัฒนาวิธีการจำแนกประเภทกิจการ

3.2 แผนการดำเนินงาน

1. ศึกษารูปแบบข้อมูลการใช้ไฟฟ้าที่ได้จากมิเตอร์ระบบ AMR
2. ศึกษาลักษณะการใช้ไฟฟ้าของกิจการแต่ละประเภท
3. ศึกษาการทำงานของอัลกอริทึมต่างๆ

4. กำหนดลักษณะประจำ (Attribute) ของข้อมูลเพื่อสร้างชุดข้อมูลสำหรับทำการทดสอบ

5. กำหนดลักษณะของการทดลอง

6. กำหนดวิธีการวัดประสิทธิภาพที่ได้จากการทดลองวิธีการต่างๆ

7. ดำเนินการทดลองตามที่ได้กำหนดไว้ เพื่อทดสอบประสิทธิภาพ

8. เก็บข้อมูล และวิเคราะห์ผลการทดลอง

3.3 เครื่องมือที่ใช้ในการวิจัย

Rapid Miner Studio คือ ซอฟท์แวร์ที่ใช้ในการวิเคราะห์ข้อมูล (Data Analysis) ที่ได้รับความนิยมมากโดย Gartner Inc. ซึ่งเป็นบริษัทเพื่อการวิจัยและให้คำปรึกษาของประเทศสหรัฐอเมริกา ได้จัดให้ Rapid Miner Studio อยู่ในกลุ่มผู้นำซอฟท์แวร์สำหรับการวิเคราะห์ข้อมูลประจำปี 2016 โดยซอฟท์แวร์นี้สามารถรองรับการใช้งานไฟล์ได้หลายประเภท สามารถวิเคราะห์และแสดงผลได้หลายรูปแบบ อีกทั้งยังสามารถแสดงผลโมเดลได้สวยงาม[11]

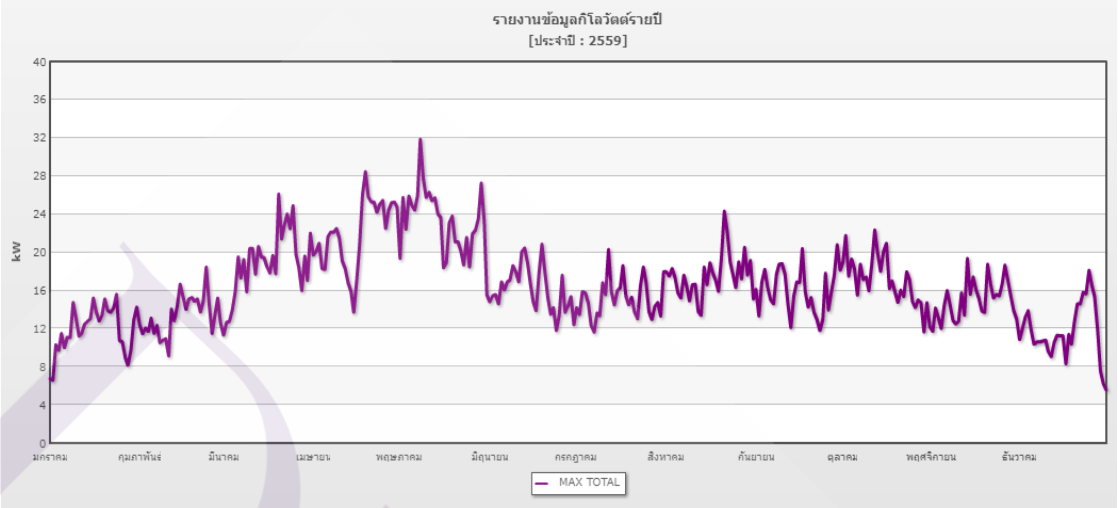
3.4 การศึกษาข้อมูลการใช้ไฟฟ้าของผู้ใช้ไฟแต่ละประเภท

งานวิจัยนี้ใช้ข้อมูลที่ได้จากระบบ AMR ของการไฟฟ้าส่วนภูมิภาคที่มีความน่าเชื่อถือและเป็นข้อมูลสถิติการใช้พลังงานไฟฟ้าจริงตลอดระยะเวลา 12 เดือน ของผู้ใช้ไฟฟ้าจำนวน 100 ราย สำหรับสร้างโมเดลในการจำแนกประเภทกิจการ แต่ข้อมูลที่ได้เป็นเพียงค่าการใช้พลังงานไฟฟ้าเท่านั้น การจะนำข้อมูลดังกล่าวมาใช้ประโยชน์ในการจำแนกข้อมูลจำเป็นต้องทราบถึงเอกลักษณ์ (Identity) ของกิจการแต่ละประเภท สำหรับงานวิจัยนี้ได้นำข้อมูลการใช้ไฟฟ้า (Load Profile) ของผู้ใช้ไฟฟ้ามาเป็นข้อมูลตั้งต้น โดยข้อมูลของผู้ใช้ไฟฟ้าแต่ละรายจะมีการระบุประเภทกิจการไว้ก่อนแล้วซึ่งแบ่งออกได้เป็น 4 ประเภท คือ ประเภทที่พักอาศัย (Residential customer) ประเภทอาคารสำนักงาน (Official customer) ประเภทโรงงานอุตสาหกรรม (Industrial customer) และประเภทสถานพยาบาล (Hospital) เมื่อพิจารณาจากข้อมูลของผู้ใช้ไฟฟ้าแต่ละประเภทพบว่ามีค่าหน่วยการใช้ไฟฟ้าที่แตกต่างกันอย่างมีนัยสำคัญ ซึ่งอธิบายได้หน่วยการใช้ไฟฟ้าดังนี้

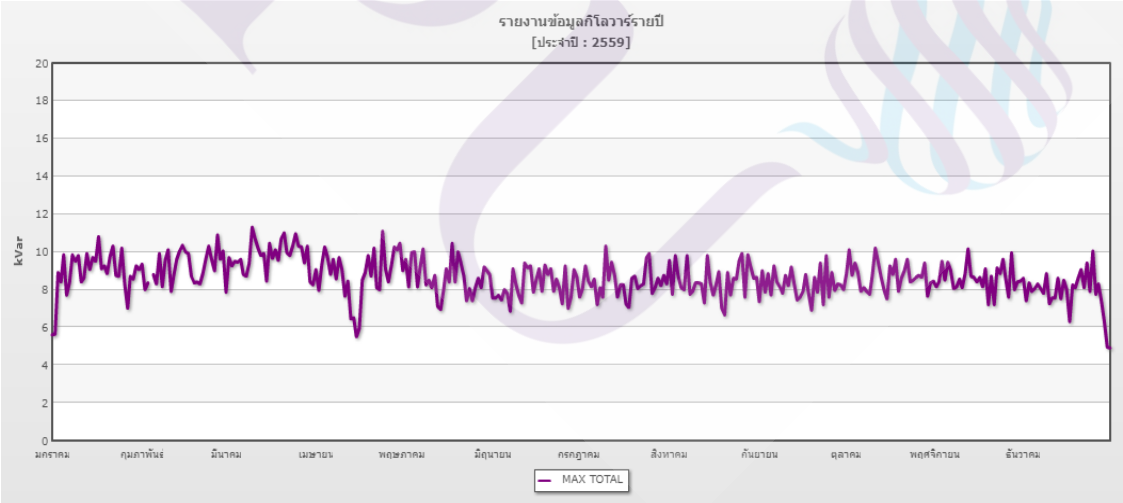
3.4.1 ประเภทที่พักอาศัย (Residential customer)

เป็นผู้ใช้ไฟฟ้าที่มีการใช้พลังงานเป็นไปในแนวทางเดียวกับสภาพอากาศหรืออุณหภูมิที่เพิ่มขึ้น เนื่องจากโหลดส่วนมากเกิดจากเครื่องใช้ไฟฟ้าที่มีหน้าที่ทำความเย็น เช่น ตู้เย็น เครื่องปรับอากาศ เป็นต้น นอกจากนี้ยังมีแนวโน้มของการใช้ไฟฟ้าเพิ่มขึ้นตามวันหยุดประจำปี เนื่องจากอาจมีการรวมตัวของสมาชิกในครอบครัวทำให้มีจำนวนผู้ใช้ไฟในบ้านเพิ่มขึ้นทำให้

มีการใช้ไฟฟ้าเพิ่มขึ้นตามไปด้วย สามารถพิจารณาได้จากข้อมูลการใช้กำลังไฟฟ้าจริงและกำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทที่พักอาศัยดังภาพที่ 3.1 และภาพที่ 3.2 ตามลำดับ

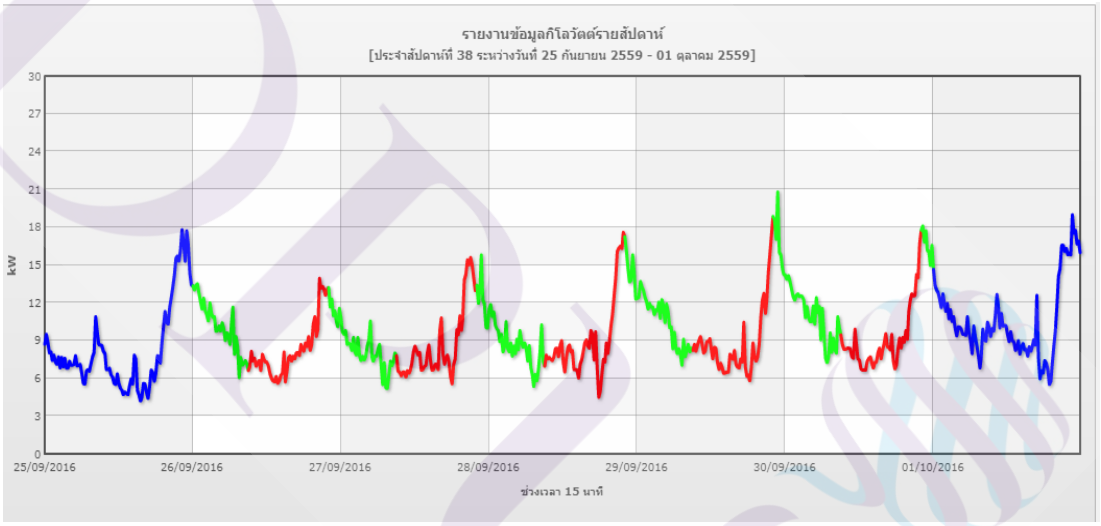


ภาพที่ 3.1 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทที่พักอาศัย



ภาพที่ 3.2 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทที่พักอาศัย

และเมื่อสังเกตตามช่วงเวลาแล้ว จะพบว่ามีการใช้ไฟฟ้าในช่วงเย็นของทุกวันสูงกว่าการใช้ไฟฟ้าในช่วงเวลาอื่นของวันเนื่องจากเป็นช่วงเวลาที่มีผู้คนอยู่บ้านและเปิดเครื่องใช้ไฟฟ้าพร้อมกันมากที่สุดทำให้มีการใช้ไฟฟ้าสูงตามไปด้วย ดังภาพที่ 3.3 จะเห็นได้ว่ากราฟนี้ประกอบไปด้วย 3 สี คือ สีแดง หมายถึง การใช้ไฟฟ้าที่มีอัตราการคิดค่าไฟในอัตรา Peak rate ตั้งแต่ 09.00 – 22.00 น. ของวันจันทร์-วันศุกร์ ยกเว้นวันหยุดนักขัตฤกษ์, สีเขียว หมายถึง การใช้ไฟฟ้าที่มีอัตราการคิดค่าไฟในอัตรา Off-Peak rate ตั้งแต่ 22.00 – 09.00 น. ของวันจันทร์-วันศุกร์ ยกเว้นวันหยุดนักขัตฤกษ์ และสีน้ำเงิน หมายถึง การใช้ไฟฟ้าที่มีอัตราการคิดค่าไฟในอัตรา Holiday rate ตลอดทั้งวันของวันเสาร์-วันอาทิตย์ รวมทั้งวันหยุดนักขัตฤกษ์ โดยอัตราการคิดค่าไฟฟ้าในปัจจุบันอัตรา Off-Peak และอัตรา Holiday rate นั้นถูกกำหนดให้มีอัตราเท่ากัน

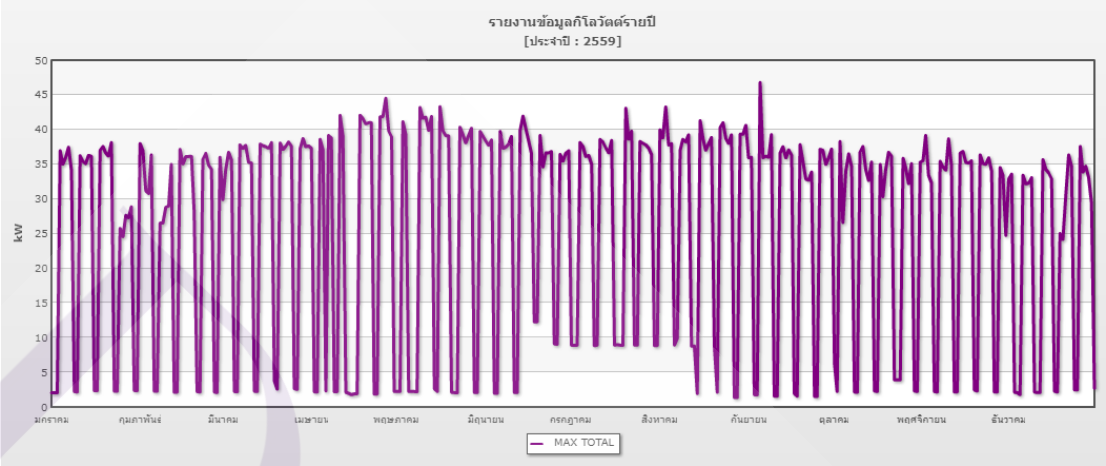


ภาพที่ 3.3 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทที่พักอาศัย

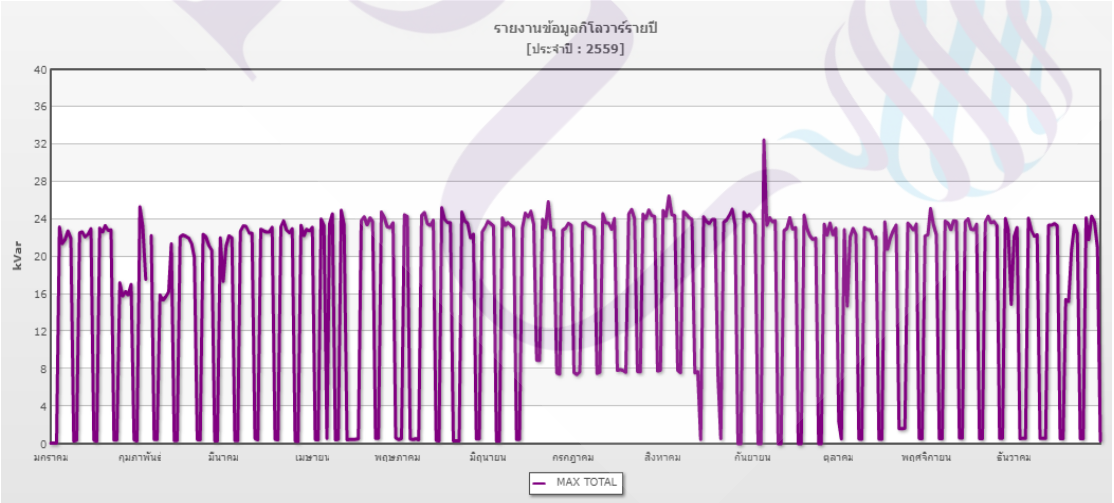
3.4.2 ประเภทอาคารสำนักงาน (Official customer)

อาคารสำนักงานโดยทั่วไปมักจะมีการใช้ไฟฟ้าที่มีปริมาณเท่ากันในทุกๆวัน จะเห็นได้จากรูปที่ ในแต่ละเดือนจะมีการใช้ไฟฟ้าที่ค่อนข้างสม่ำเสมอ ไม่แปรปรวนมากนัก สามารถอธิบายได้จากการใช้ไฟฟ้าโดยทั่วไปของอาคารสำนักงานเช่น ธนาคาร ที่ทำการไปรษณีย์ สำนักงานต่างๆของรัฐ มักมีการเปิดใช้เครื่องปรับอากาศเป็นประจำทุกวันในเวลาทำการ ซึ่งเครื่องปรับอากาศที่ใช้ในอาคารถือเป็นโหลดขนาดใหญ่ ทำให้แนวโน้มการใช้ไฟฟ้ามีลักษณะใกล้เคียงกันในทุกๆวัน

สามารถพิจารณาได้จากข้อมูลการใช้กำลังไฟฟ้าจริงและกำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทอาคารสำนักงานดังภาพที่ 3.3 และภาพที่ 3.4 ตามลำดับ

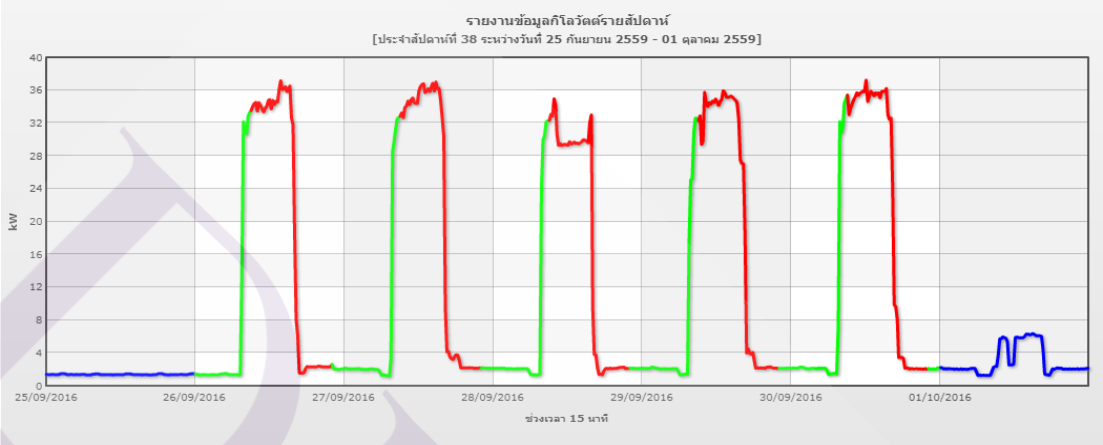


ภาพที่ 3.4 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทอาคารสำนักงาน



ภาพที่ 3.5 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทอาคารสำนักงาน

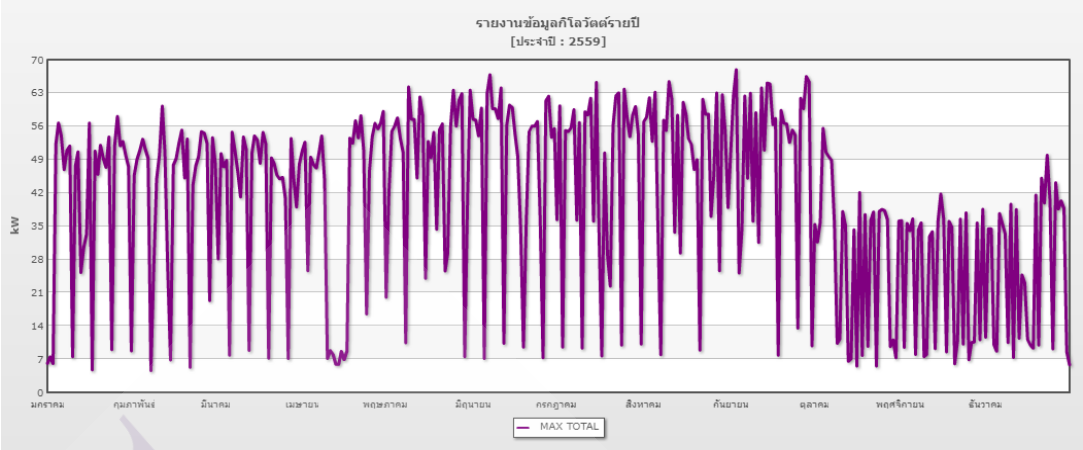
จะเห็นได้ว่ากราฟมีลักษณะตกลงเป็นประจำในทุกๆเดือน เดือนละสี่ครั้งโดยประมาณนั้น เนื่องมาจากวันหยุดทำการและวันหยุดนักขัตรฤกษ์ ที่ไม่มีการทำงานทำให้ไม่มีการใช้ไฟฟ้าหรือหากมีใช้ก็เป็นปริมาณที่ไม่สูงมากนัก สามารถอธิบายได้จากข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทอาคารสำนักงานดังภาพที่ 3.6



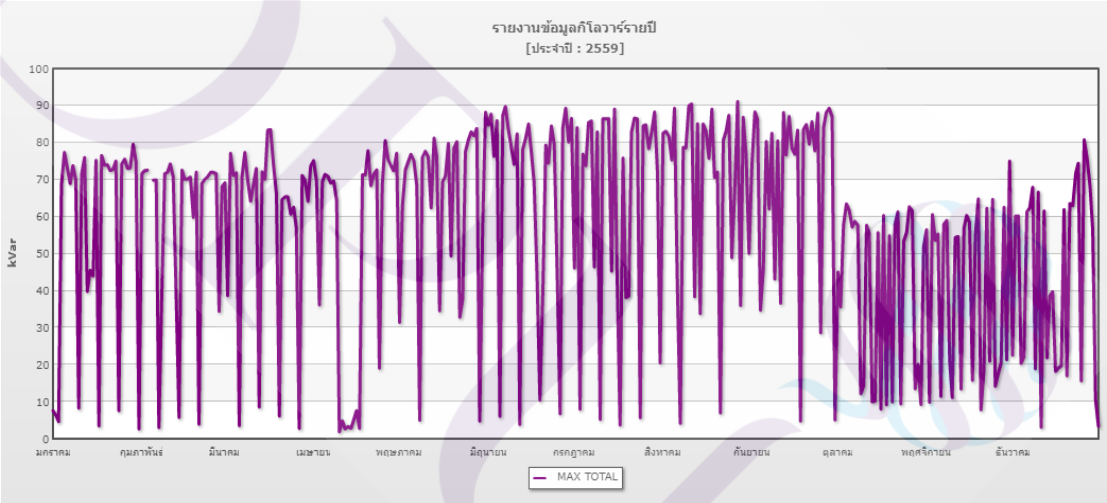
ภาพที่ 3.6 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทอาคารสำนักงาน

3.4.3 ประเภทโรงงานอุตสาหกรรม (Industrial customer)

โรงงานอุตสาหกรรมนั้นมีหลากหลายรูปแบบ ทั้งแบบที่มีการใช้ไฟฟ้าขึ้นกับกำลังการผลิตที่มีลักษณะประจำ เช่น โรงงานน้ำตาล โรงงานแป้งมัน โรงงานสับปะรดกระป๋อง เป็นต้น เนื่องจากมีวัตถุดิบในการดำเนินงานในช่วงเดิมเป็นประจำทุกปี และแบบมีการใช้ไฟฟ้าไม่เป็นรูปแบบประจำขึ้นกับยอดการสั่งซื้อหรือยอดการผลิต จากรูปที่ เป็นลักษณะการใช้ไฟฟ้าของโรงงานแปรรูปไม้ซึ่งมีการใช้ไฟฟ้าไม่แน่นอนเป็นไปตามยอดการสั่งผลิต แต่ที่เห็นได้ชัดเจนคือโรงงานนี้มีการหยุดผลิตในช่วงเทศกาลสงกรานต์ และเทศกาลปีใหม่ซึ่งเป็นวันหยุดประจำปี สามารถพิจารณาได้จากข้อมูลการใช้กำลังไฟฟ้าจริงและกำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทอาคารสำนักงานดังภาพที่ 3.7 และภาพที่ 3.8 ตามลำดับ

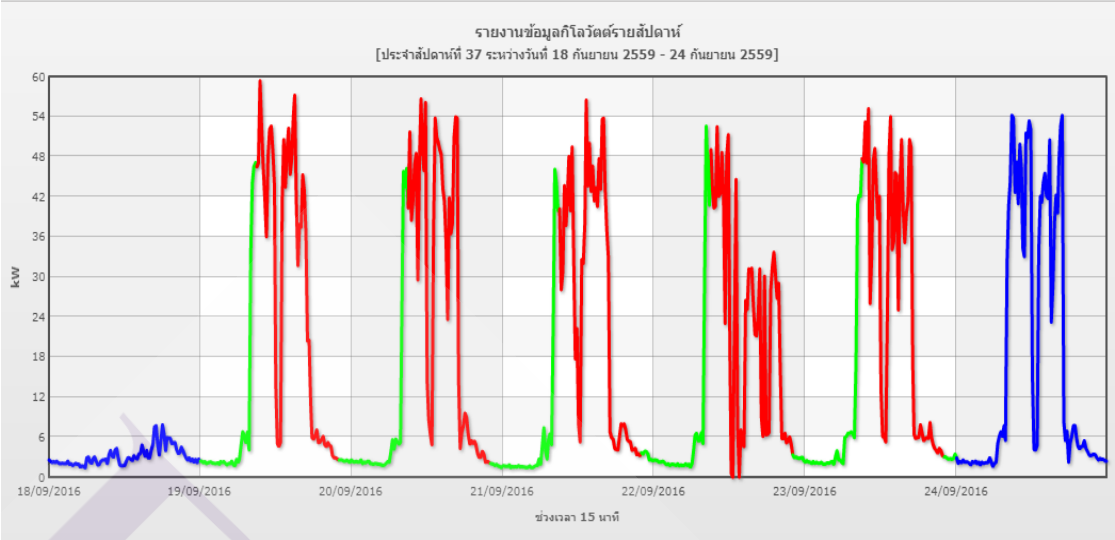


ภาพที่ 3.7 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม



ภาพที่ 3.8 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม

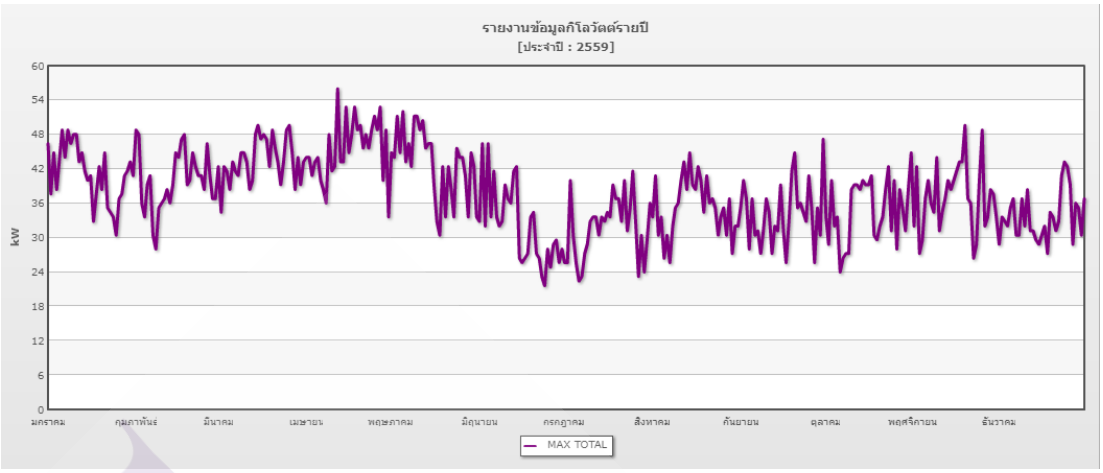
เมื่อสังเกตการณ์ใช้ไฟฟ้าในแต่ละวัน พบว่ามีการทำงานในช่วงเช้า และมีการพักกลางวันในช่วงเที่ยงแล้วกลับมาทำงานใหม่ในช่วงบ่ายจนถึงช่วงเย็น นอกจากนี้ยังมีการทำงานในวันเสาร์ตามเวลารูปแบบเดียวกันด้วยดังภาพที่ 3.9



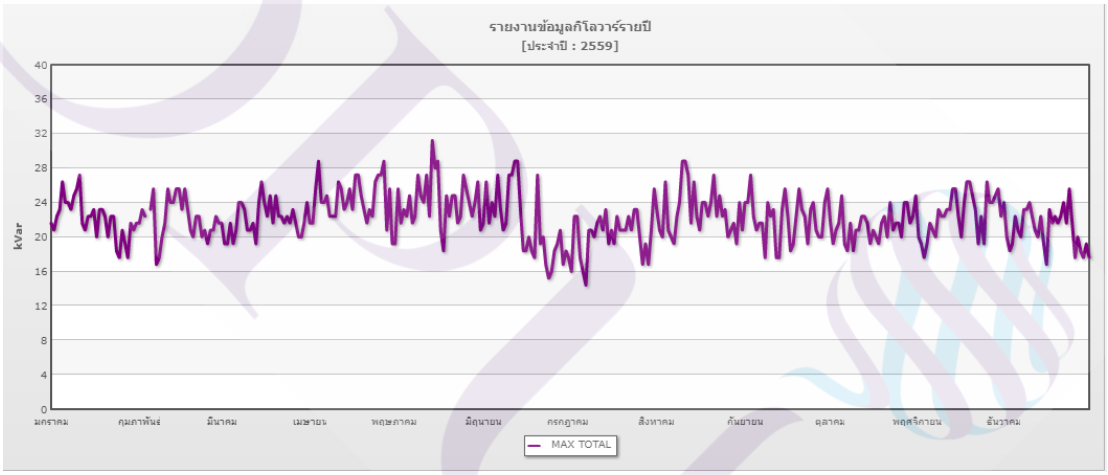
ภาพที่ 3.9 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทโรงงานอุตสาหกรรม

3.4.3 ประเภทสถานพยาบาล (Hospital)

สถานพยาบาล เป็นอีกหนึ่งประเภทของผู้ใช้ไฟฟ้าที่หารูปแบบประจำได้ยาก เนื่องจากในแต่ละวันหรือแต่ละเดือนจะมีจำนวนผู้เข้ารับบริการที่ไม่แน่นอน อย่างไรก็ตามสถานพยาบาลนั้นก็มีโหลดอยู่หลายประเภทที่ต้องเดินเครื่องอยู่ตลอดเวลา ทำให้สถานพยาบาลที่เปิดทำการตามเวลาคล้ายอาคารสำนักงานมีความต่างออกไป เช่น เครื่องทำออกซิเจน เครื่องปรับอากาศของห้องพัก เป็นต้น สามารถพิจารณาได้จากข้อมูลการใช้กำลังไฟฟ้าจริงและกำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทอาคารสำนักงานดังภาพที่ 3.10 และภาพที่ 3.11 ตามลำดับ

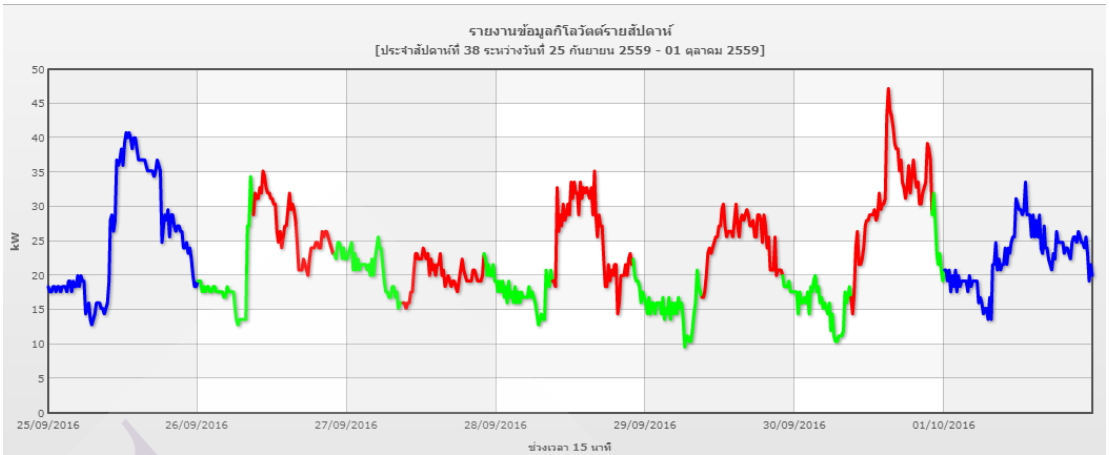


ภาพที่ 3.10 ข้อมูลการใช้กำลังไฟฟ้าจริงรายปีของผู้ใช้ไฟประเภทสถานพยาบาล



ภาพที่ 3.11 ข้อมูลการใช้กำลังไฟฟ้ารีแอกทีฟรายปีของผู้ใช้ไฟประเภทสถานพยาบาล

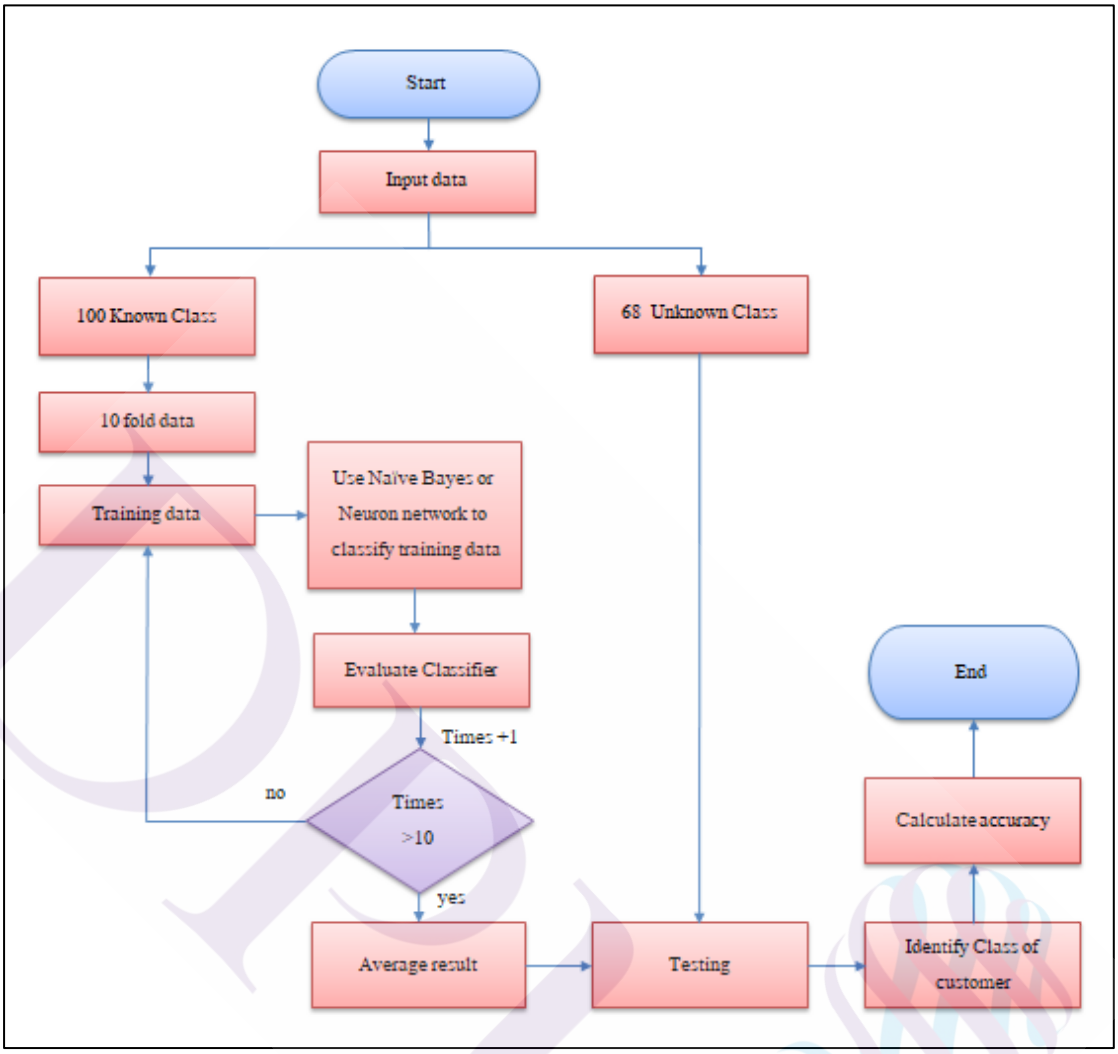
นอกจากนั้นยังเห็นได้ว่า ทุกวันทุกช่วงเวลาในแต่ละสัปดาห์ล้วนมีการใช้ไฟฟ้าดังภาพที่ 3.12



ภาพที่ 3.12 ข้อมูลการใช้กำลังไฟฟ้าจริงรายสัปดาห์ของผู้ใช้ไฟประเภทสถานพยาบาล

3.5 การออกแบบและกำหนดลักษณะประจำ (Attribute)

เมื่อทราบถึงลักษณะการใช้ไฟฟ้าของกิจการแต่ละประเภทแล้ว ต่อมาคือการออกแบบข้อมูลที่มีอยู่ และกำหนดลักษณะประจำที่เหมาะสมเพื่อใช้ในการจำแนกต่อไป โดยการออกแบบการจำแนกประเภทของผู้ใช้ไฟฟ้าสามารถแสดงเป็น Flowchart ได้ดังภาพที่ 3.13



ภาพที่ 3.13 Flowchart การทดสอบที่นำเสนอ

การทดสอบการจำแนกประเภทของผู้ใช้ไฟฟ้าในงานวิจัยนี้ ได้นำข้อมูลการใช้ไฟฟ้าของผู้ใช้ไฟฟ้าจำนวน 168 ราย โดยแบ่งออกเป็นสองส่วน คือ ผู้ใช้ไฟฟ้าจำนวน 100 รายสำหรับทำการฝึกสอนด้วยอัลกอริทึมที่กำหนด ได้แก่ นาอีฟเบย์และโครงข่ายประสาทเทียม ซึ่งจะต้องอาศัยลักษณะประจำที่ผู้วิจัยได้ออกแบบประกอบไปด้วย โหลดแฟกเตอร์ของวันทำการ, โหลดแฟกเตอร์ของวันหยุด, สัดส่วนของการใช้ไฟฟ้าตามช่วงเวลาในวันทำการ, ปริมาณการใช้กำลังไฟฟ้ารีแอคทีฟ และการเปลี่ยนแปลงการใช้กำลังไฟฟ้ารีแอคทีฟโดยเฉลี่ยในหนึ่งปี ซึ่งจะกล่าวในลำดับต่อไป จากนั้นทำการฝึกฝนทั้งหมด 10 รอบ เนื่องจากผู้วิจัยได้เลือกการคำนวณในลักษณะ 10 fold – validation เพื่อใช้ในการทดสอบนี้ แล้วทำการหาค่าเฉลี่ยของผลลัพธ์ดังกล่าวว่าเป็นค่าที่รับได้ตาม

วัตถุประสงค์หรือไม่ จากนั้นนำข้อมูลผู้ใช้ไฟฟ้าจำนวน 68 รายที่เหลือเพื่อใช้เป็นข้อมูลสำหรับทดสอบการจำแนกประเภทผู้ใช้ไฟฟ้าว่าสามารถจำแนกได้อย่างถูกต้องหรือไม่

3.5.1 ค่าตัวประกอบการใช้ไฟฟ้า (Load Factor : LF_{WORK} , LF_{HOL})

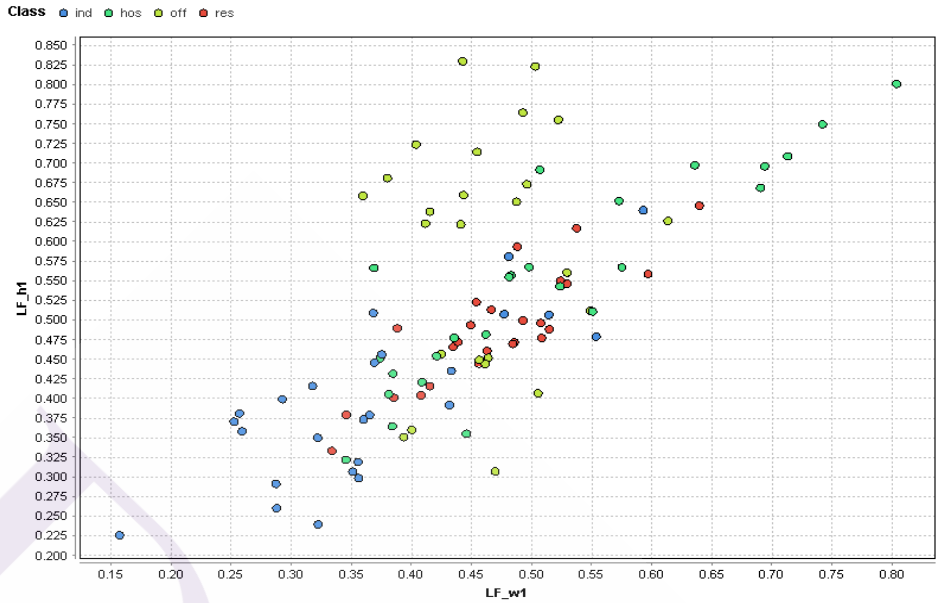
ค่าตัวประกอบการใช้ไฟฟ้า (Load Factor) เป็น Attribute ที่เหมาะสมสำหรับใช้ในการทดลอง โดยจะแบ่ง Load Factor นี้ออกเป็นสองค่าคือ Load Factor ในช่วงวันทำการ (Working day Load Factor : LF_{WORK}) และ Load Factor ในช่วงวันหยุด (Holiday Load Factor : LF_{HOL}) ของแต่ละวันดังสมการ

$$LF = \frac{P_{AVE}}{P_{MAX}} \tag{3.1}$$

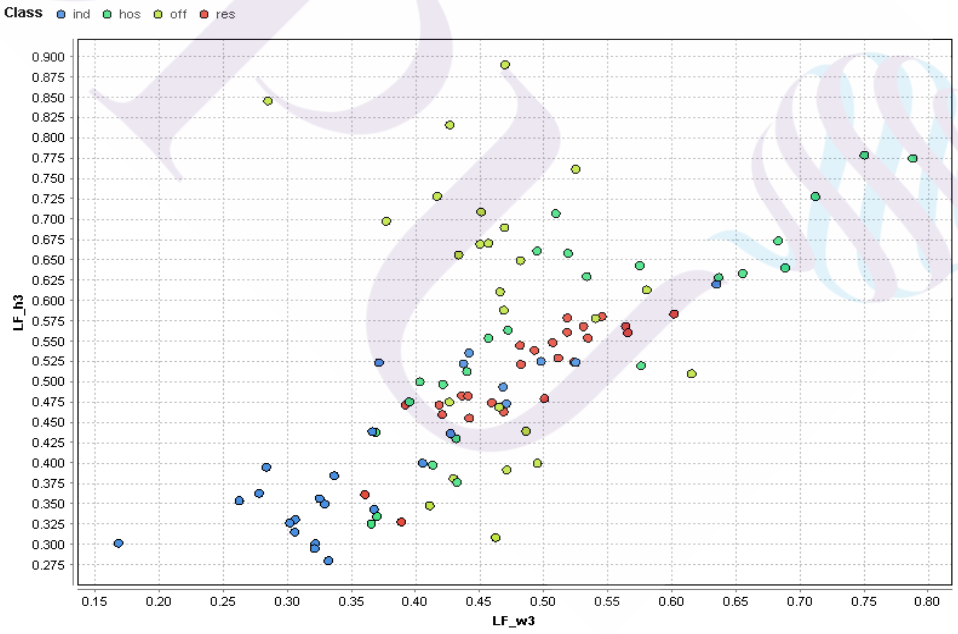
P_{AVE} คือ กำลังไฟฟ้าเฉลี่ยในช่วงเวลานั้น

P_{MAX} คือ กำลังไฟฟ้าสูงสุดในช่วงเวลานั้น

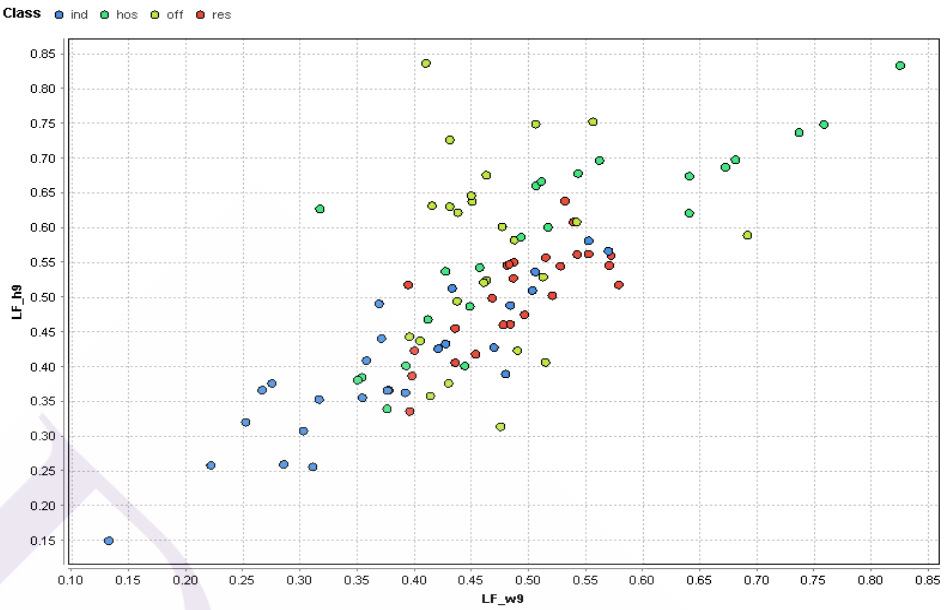
จากนั้นหาค่าเฉลี่ยของทั้งสองค่าในแต่ละเดือนของผู้ใช้ไฟฟ้าแต่ละราย กำหนดให้ Load Factor ของวันทำการอยู่ในแกนแนวนอน และ Load Factor ของวันหยุดอยู่ในแกนแนวตั้ง จุดสีแดงหมายถึงประเภทที่พักอาศัย จุดสีแหลืองหมายถึงประเภทอาคารสำนักงาน จุดสีเขียวหมายถึงประเภทสถานพยาบาล และจุดสีฟ้าหมายถึงประเภทโรงงานอุตสาหกรรม เมื่อเปรียบเทียบ Load Factor ของวันทำการและ Load Factor ของวันหยุดในแต่ละเดือนดังภาพที่ 3.13 ภาพที่ 3.14 และภาพที่ 3.15 จะสังเกตได้ว่า จุดสีเหลืองคือประเภทอาคารสำนักงานนั้นจะไม่สอดคล้องกันไม่ว่าจะเป็นฤดูกาลใดก็ตาม เนื่องจากในวันหยุดนั้นมีปริมาณการใช้ไฟฟ้าต่ำมากทำให้ Load Factor ของวันหยุดที่เกิดขึ้นนั้นมีค่าสูง



ภาพที่ 3.14 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนมกราคม



ภาพที่ 3.15 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนมีนาคม



ภาพที่ 3.16 เปรียบเทียบ Load Factor ระหว่างวันทำการและวันหยุดของเดือนกันยายน

3.5.2 การใช้กำลังไฟฟ้าจริงตามช่วงเวลาในวันทำการ (Load Shape Impact)

โดยแบ่งออก เป็น 3 ส่วนตามช่วงเวลา ดังนี้

ตารางที่ 3.1 กำหนดสัดส่วนของเวลาในวันทำการ

ลำดับที่	วัน	เวลา	จำนวน (ชั่วโมง)
1	วันจันทร์-วันศุกร์ (ไม่รวมวันหยุดนักขัตฤกษ์)	09.00 – 17.00 น.	8
2	วันจันทร์-วันศุกร์ (ไม่รวมวันหยุดนักขัตฤกษ์)	17.00 – 22.00 น.	5
3	วันจันทร์-วันศุกร์ (ไม่รวมวันหยุดนักขัตฤกษ์)	22.00 – 09.00 น.	11

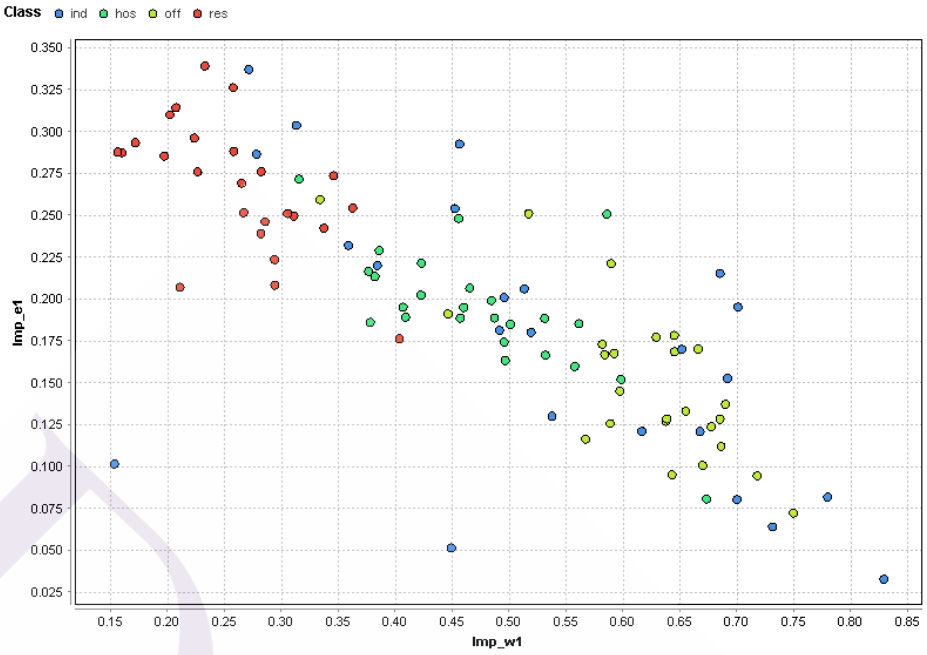
จะเห็นได้ว่าสัดส่วนการพิจารณาที่กำหนดจะมีความคล้ายคลึงกับการแบ่งอัตราค่าไฟฟ้าแบบ TOU แต่จะถูกแบ่งให้ละเอียดเพิ่มขึ้น เนื่องจากสมมติฐานที่ว่ากิจการประเภทอาคารสำนักงานจะเลิกงานหลัง 17.00 น. ซึ่งจะทำให้มีปริมาณการใช้ไฟฟ้าลดลง สามารถสรุปเป็นตาราง Load

Shape Impact ได้ดังตารางที่ 3.2 จากนั้นหาค่าเฉลี่ยของค่า Impact ทั้งสามค่าในแต่ละเดือนของผู้ใช้ไฟฟ้าแต่ละราย

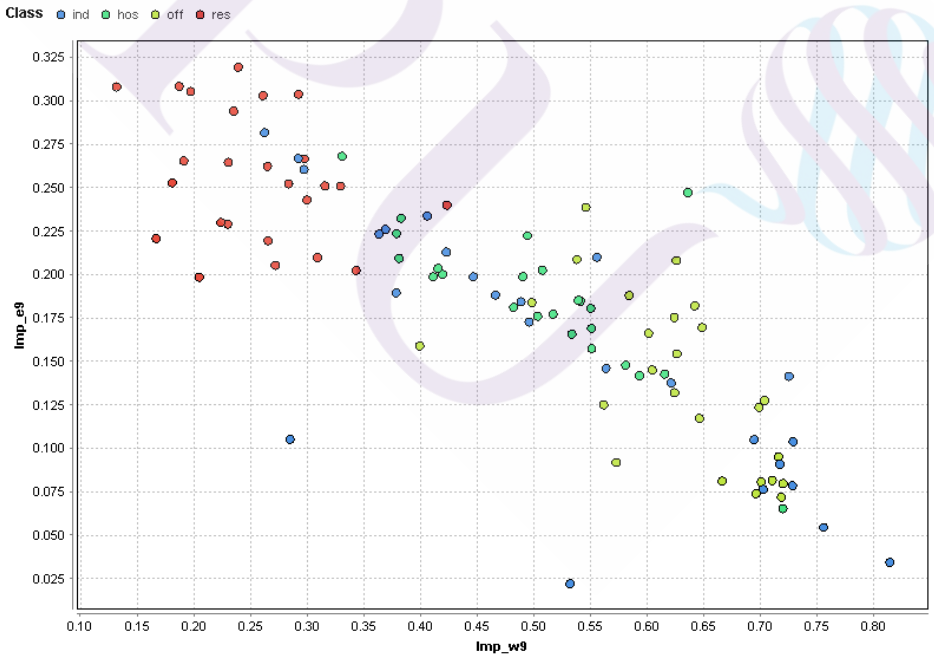
ตารางที่ 3.2 Load Shape Impact

ที่	พารามิเตอร์	นิยาม	หมายเหตุ
1	Working Time Impact (W)	$W = \frac{P_{AVE,work}}{P_{AVE,day}}$	09.00 – 17.00 น.
2	Evening Impact (E)	$E = \frac{P_{AVE,eve}}{P_{AVE,day}}$	17.00 – 22.00 น.
3	Night Impact (N)	$N = \frac{P_{AVE,night}}{P_{AVE,day}}$	22.00 – 09.00 น.

เมื่อเปรียบเทียบระหว่าง Working Time Impact และ Evening Impact ในทุกๆเดือน โดยกำหนดให้ Working Time Impact อยู่ในแกนนอนและ Evening Impact อยู่ในแกนแนวตั้ง จะสังเกตเห็นจุดสีแดงกระจุกตัวซึ่งหมายถึงกลุ่มที่พักอาศัย (Residential customer) สามารถแบ่งออกด้วยสายตาอย่างชัดเจน เนื่องจากสัดส่วนการใช้ไฟฟ้าของผู้ใช้ไฟประเภทที่อยู่อาศัยจะใช้ปริมาณมากในช่วงเย็นนั่นเอง ดังภาพที่ 3.16 และ ภาพที่ 3.17

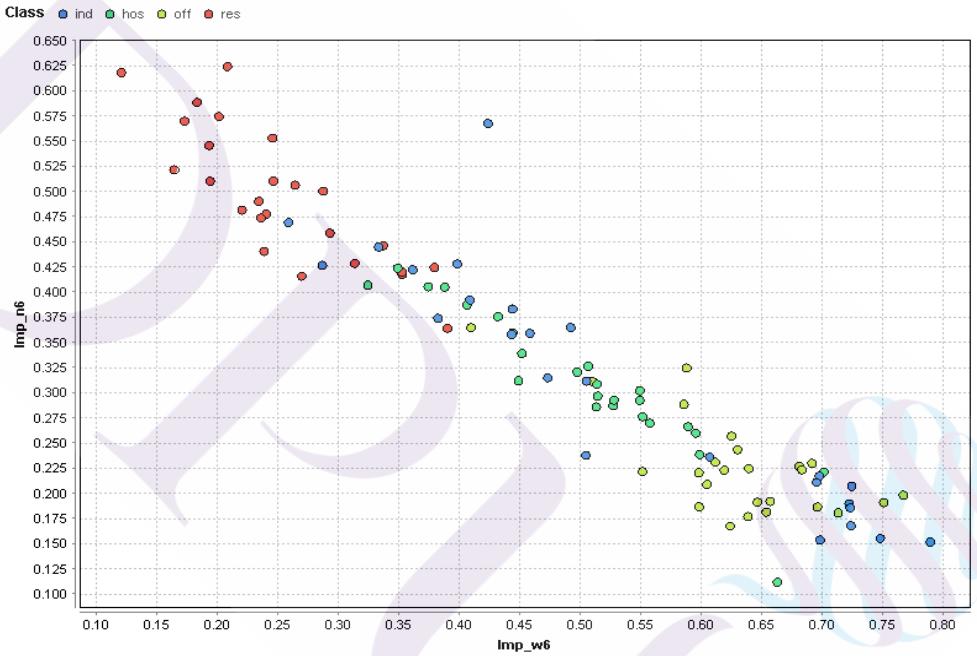


ภาพที่ 3.17 เปรียบเทียบระหว่าง Working Time Impact กับ Evening Impact ของเดือนมกราคม

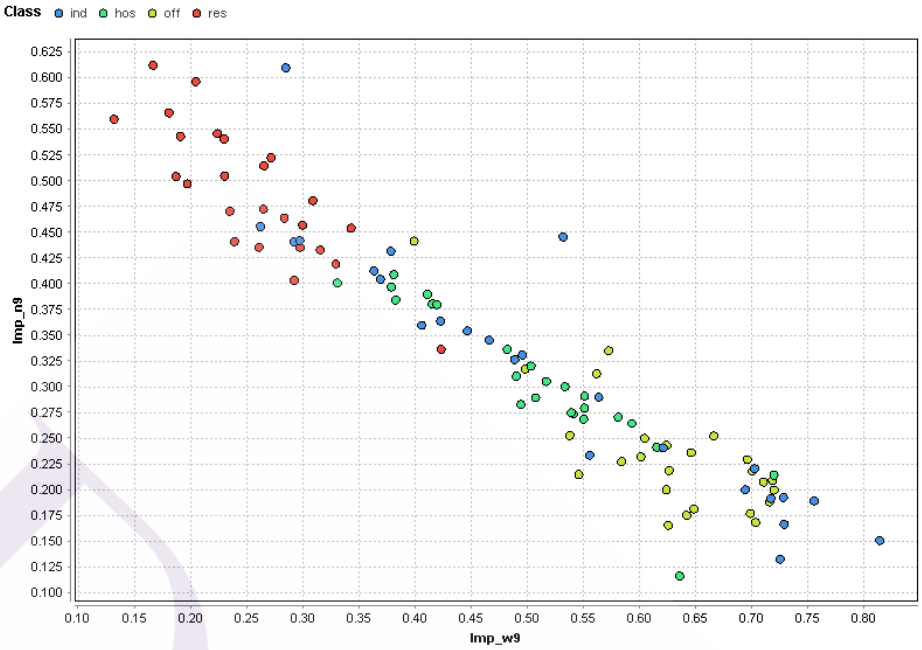


ภาพที่ 3.18 เปรียบเทียบระหว่าง Working Time Impact กับ Evening Impact ของเดือนกันยายน

และเมื่อเปรียบเทียบระหว่าง Working Time Impact และ Night Impact ในทุกๆเดือน กำหนดให้ Working Time Impact อยู่ในแกนนอนและ Night Impact อยู่ในแกนตั้ง จะสังเกตเห็นจุดสีแดงกระจุกตัวซึ่งหมายถึงกลุ่มที่พักอาศัย (Residential customer) เนื่องจากการใช้ไฟฟ้าของผู้ใช้ไฟประเภทที่อยู่อาศัยจะใช้ปริมาณมากในช่วงกลางคืนเมื่อเทียบกับช่วงกลางวัน และจุดสีเหลืองที่กระจุกตัวหมายถึงกลุ่มอาคารสำนักงาน (Official customer) ที่สามารถแบ่งออกได้ด้วยสายตาเช่นกัน เนื่องจากการใช้ไฟฟ้าของผู้ใช้ไฟประเภทอาคารสำนักงานจะใช้ไฟฟ้าปริมาณน้อยในช่วงกลางคืนเมื่อเทียบกับช่วงกลางวันนั่นเอง ดังภาพที่ 3.18 และ ภาพที่ 3.19



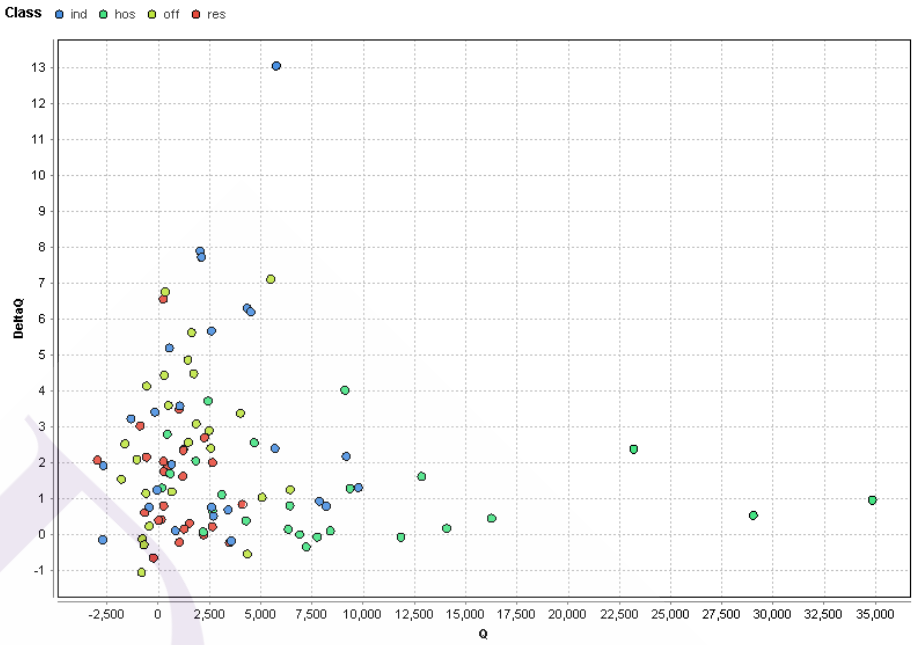
ภาพที่ 3.19 เปรียบเทียบระหว่าง Working Time Impact กับ Night Impact ของเดือนมิถุนายน



ภาพที่ 3.20 เปรียบเทียบระหว่าง Working Time Impact กับ Night Impact ของเดือนกันยายน

3.5.3 การใช้กำลังไฟฟ้ารีแอกทีฟ (Reactive Power : Q)

เนื่องจากเป็นลักษณะเด่นที่สังเกตได้จากการใช้ไฟฟ้าของกิจการประเภทอุตสาหกรรมและสถานพยาบาล สาเหตุมาจากเครื่องจักรกลหรือมอเตอร์ขนาดใหญ่มักมีการบริโภคกำลังไฟฟ้ารีแอกทีฟปริมาณมากเพื่อใช้สร้างสนามไฟฟ้าของต้นกำลัง โดยหาค่าเฉลี่ยในแต่ละเดือนของผู้ใช้ไฟฟ้าแต่ละราย ส่วนการแยกระหว่างประเภทสถานพยาบาลและโรงงานอุตสาหกรรมจำเป็นต้องใช้อีกหนึ่งตัวแปรคือ ความแตกต่างระหว่างเดือนของกำลังไฟฟ้ารีแอกทีฟ (ΔQ) เนื่องจากสถานพยาบาลไม่ว่าจะเป็นเดือนใดของปีย่อมต้องมีการใช้โหลดที่มีองค์ประกอบของมอเตอร์เป็นประจำ ซึ่งต่างกับโรงงานอุตสาหกรรมที่จะใช้กำลังไฟฟ้ารีแอกทีฟมากน้อยขึ้นอยู่กับกำลังการผลิต เมื่อเปรียบเทียบการใช้กำลังไฟฟ้ารีแอคทีฟกับความแตกต่างของการใช้กำลังไฟฟ้ารีแอกทีฟในแต่ละเดือน โดยกำหนดให้การใช้กำลังไฟฟ้ารีแอคทีฟอยู่ในแกนนอน และความแตกต่างของการใช้กำลังไฟฟ้ารีแอกทีฟอยู่ในแกนตั้ง จะเห็นได้ว่าจุดสีฟ้าหมายถึงผู้ใช้ไฟฟ้าประเภทโรงงานอุตสาหกรรม (Industrial customer) นั้นจะมีความเปลี่ยนแปลงของการใช้กำลังไฟฟ้ารีแอกทีฟสูงกว่าจุดสีเขียวหรือผู้ใช้ไฟประเภทสถานพยาบาล (Hospital) แม้ว่าจะมีการใช้กำลังไฟฟ้ารีแอกทีฟเนื่องจากมีโหลดขนาดใหญ่เช่นเดียวกันก็ตามดังภาพที่ 3.20



ภาพที่ 3.21 เปรียบเทียบกำลังไฟฟ้ารีแอคทีฟกับความแตกต่างของกำลังไฟฟ้ารีแอกทีฟ

สามารถสรุป Attribute ของการทดสอบที่ได้กำหนดขึ้นดังตารางที่ 3.3

ตารางที่ 3.3 Attributes ที่ใช้ในการทดสอบ

No	**Class**	LFwork_1	LFhol_1	W_1	E_1	N_1	...	LFwork_12	LFhol_12	W_12	E_12	N_12	**Q**	**ΔQ**
1	Residence						...							
2	Hospital						...							
3	Industry						...							
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	Office						...							

โดยที่ หมายเลข 1 หมายถึงเดือน ม.ค. ,..., หมายเลข 12 หมายถึงเดือน ธ.ค.

ในส่วนของการทดสอบจะทำการแปลงข้อมูลการใช้ไฟฟ้าที่ได้จากระบบ AMR ให้อยู่ในรูปแบบของ Attribute ที่กำหนดไว้ รวมถึงจะทำการเปรียบเทียบค่าความถูกต้องของการจำแนกประเภทกิจการของผู้ใช้ไฟฟ้า ดังจะกล่าวในบทต่อไป

DPU

บทที่ 4
ผลการทดสอบ

สำหรับเนื้อหาในบทนี้ จะเป็นการอธิบายและแสดงผลการทดลองของโมเดลตามแนวทางที่เสนอมาซึ่งได้จากการวิเคราะห์ข้อมูลการใช้ไฟฟ้า (Load Profile) สำหรับหาลักษณะประจำ (Attribute) ของกิจการแต่ละประเภท เนื่องจาก Attribute เป็นเสมือนอัตลักษณ์บ่งชี้ลักษณะเฉพาะ และยังเป็นปัจจัยสำคัญที่จะทำให้ผลลัพธ์การจำแนกมีความถูกต้องจำเป็นต้องใช้วิธีการที่ก่อให้เกิดผลลัพธ์ที่เชื่อถือได้ โดยผู้วิจัยได้กำหนดให้มี Attribute สำหรับการใช้ในการทดสอบดังนี้

1. Load Factor สำหรับวันทำการ จำนวน 12 เดือน
2. Load Factor สำหรับวันหยุด จำนวน 12 เดือน
3. Working Time Impact (W) จำนวน 12 เดือน
4. Evening Impact (E) จำนวน 12 เดือน
5. Night Impact (N) จำนวน 12 เดือน
6. Reactive Power โดยเฉลี่ยของทั้งปี
7. Reactive Power ที่เปลี่ยนแปลงไปในแต่ละเดือนโดยเฉลี่ยของใน 1 ปี

จากนั้นนำ Attribute ทั้งหมดมาทำการจำแนกประเภทกิจการของผู้ใช้ไฟฟ้า โดยใช้เทคนิคการจำแนกข้อมูล (Classification) ซึ่งงานวิจัยนี้ได้เลือกเทคนิคที่ได้รับความนิยมในใช้งาน ได้แก่ นาอีฟเบย์ (Naïve Bay) และโครงข่ายประสาทเทียม (Neural Network) สำหรับพัฒนาวิธีการจำแนกประเภทกิจการตามกรณีศึกษา

4.1 การทดสอบและวัดประสิทธิภาพ

วิธีการทดลองและวัดประสิทธิภาพได้อธิบายวิธีการโดยละเอียดในภาคผนวก

4.2 ผลการทดสอบโมเดลที่สร้างด้วยนาอีฟเบย์

จากการทดสอบโมเดลที่สร้างนาอีฟเบย์สามารถแสดงผลการทดสอบด้วยค่าความเที่ยงตรงและค่าความคลาดเคลื่อนได้ดังนี้

ตารางที่ 4.1 ความเที่ยงตรงของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยนาอีฟเบย์

Accuracy: 78.00% +/- 17.78% (mikro: 78.00%)					
	True industry	True hospital	True office	True residence	Class precision
Pred. industry	16	2	0	1	84.21%
Pred. hospital	4	17	3	1	68.00%
Pred. office	2	5	22	0	75.86%
Pred. residence	3	1	0	23	85.19%
Class recall	64.00%	68.00%	88.00%	92.00%	

ตารางที่ 4.2 ความคลาดเคลื่อนของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน **(**Training set) ด้วยนาอีฟเบย์

Classification error: 22.00% +/- 17.78% (mikro: 22.00%)					
	True industry	True hospital	True office	True residence	Class precision
Pred. industry	16	2	0	1	84.21%
Pred. hospital	4	17	3	1	68.00%
Pred. office	2	5	22	0	75.86%
Pred. residence	3	1	0	23	85.19%
Class recall	64.00%	68.00%	88.00%	92.00%	

จะเห็นได้ว่าหลังจากการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยนาอีฟเบย์แล้ว ให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 78 และเมื่อนำโมเดลดังกล่าวไปใช้ใน

การจำแนกข้อมูลสำหรับทดสอบ (Testing set) จำนวน 64 ราย พบว่าให้ผลความเที่ยงตรงที่ร้อยละ 81.25

ตารางที่ 4.3 ความเที่ยงตรงของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนาอีฟเบย์

Accuracy: 81.25%					
	True hospital	True industry	True office	True residence	Class precision
Pred. hospital	13	1	1	1	81.25%
Pred. industry	2	11	2	1	68.75%
Pred. office	2	1	19	0	86.36%
Pred. residence	0	1	0	9	90.00%
Class recall	76.47%	78.57%	86.36%	81.82%	

ตารางที่ 4.4 ความคลาดเคลื่อนของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนาอีฟเบย์

Classification error: 18.75%					
	True hospital	True industry	True office	True residence	Class precision
Pred. hospital	13	1	1	1	81.25%
Pred. industry	2	11	2	1	68.75%
Pred. office	2	1	19	0	86.36%
Pred. residence	0	1	0	9	90.00%
Class recall	76.47%	78.57%	86.36%	81.82%	

4.3 ผลการทดสอบโมเดลที่สร้างด้วยนิวรัลเน็ตเวิร์ค

จากการทดสอบโมเดลที่สร้างด้วยนิวรัลเน็ตเวิร์คสามารถแสดงผลการทดสอบด้วยค่าความเที่ยงตรงและค่าความคลาดเคลื่อนได้ดังนี้

ตารางที่ 4.5 ความเที่ยงตรงของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยนิวรัลเน็ตเวิร์ค

Accuracy: 75.00% +/- 13.60% (mikro: 75.00%)					
	True industry	True hospital	True office	True residence	Class precision
Pred. industry	14	3	0	3	70.00%
Pred. hospital	3	20	5	1	68.97%
Pred. office	1	2	20	0	86.96%
Pred. residence	7	0	0	21	75.00%
Class recall	56.00%	80.00%	80.00%	84.00%	

ตารางที่ 4.6 ความคลาดเคลื่อนของผลการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วยด้วยนิวรัลเน็ตเวิร์ค

Classification error: 25.00% +/- 13.61% (mikro: 25.00%)					
	True industry	True hospital	True office	True residence	Class precision
Pred. industry	14	3	0	3	70.00%
Pred. hospital	3	20	5	1	68.97%
Pred. office	1	2	20	0	86.96%
Pred. residence	7	0	0	21	75.00%
Class recall	56.00%	80.00%	80.00%	84.00%	

จะเห็นได้ว่าหลังจากการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับสอน (Training set) ด้วยนิวรัลเน็ตเวิร์คแล้ว ให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 75 และเมื่อนำโมเดลดังกล่าวไปใช้ในการจำแนกข้อมูลสำหรับทดสอบ (Testing set) จำนวน 64 ราย พบว่าให้ผลความเที่ยงตรงที่ร้อยละ 87.50

ตารางที่ 4.7 ความเที่ยงตรงของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนิวรัลเน็ตเวิร์ค

Accuracy: 87.50%					
	True hospital	True industry	True office	True residence	Class precision
Pred. hospital	14	0	1	0	93.33%
Pred. industry	2	11	0	0	84.62%
Pred. office	1	1	21	1	87.50%
Pred. residence	0	2	0	10	83.33%
Class recall	82.35%	78.57%	95.45%	90.91%	

ตารางที่ 4.8 ความคลาดเคลื่อนของการจำแนกชุดข้อมูลทดสอบ (Testing set) ด้วยนิวรัลเน็ตเวิร์ค

Classification error: 12.50%					
	True hospital	True industry	True office	True residence	Class precision
Pred. hospital	14	0	1	0	93.33%
Pred. industry	2	11	0	0	84.62%
Pred. office	1	1	21	1	87.50%
Pred. residence	0	2	0	10	83.33%
Class recall	82.35%	78.57%	95.45%	90.91%	

สรุป

จากการทดลองแสดงให้เห็นว่าลักษณะประจำ (Attribute) ที่ได้ทำการสังเคราะห์จากข้อมูลการใช้ไฟฟ้าที่ได้จากระบบ AMR นั้น สามารถนำมาใช้ในการสร้างโมเดลสำหรับการจำแนกผู้ใช้ไฟฟ้าตามประเภทกิจการต่างๆ ได้ โดยอาศัยเทคนิคการ Classification ทั้งนาอีฟเบย์และนิวรัลเน็ทเวิร์ค ซึ่งสามารถทำการจำแนกได้จริง และให้ผลลัพธ์อย่างถูกต้องตามวัตถุประสงค์การวิจัยนี้ ในบทต่อไปจะกล่าวถึงบทสรุปและการวิเคราะห์ผลการทดลอง รวมถึงข้อเสนอแนะจากการทดลอง

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

เนื้อหาในบทนี้จะเป็นการอภิปรายเพื่อสรุปผลที่ได้จากการทดลองงานวิจัย รวมทั้งข้อจำกัดของโมเดลที่เสนอ และข้อเสนอแนะสำหรับแนวทางในการพัฒนางานวิจัยนี้ต่อไปเพื่อแก้ไขข้อบกพร่องของระบบให้มีประสิทธิภาพมากยิ่งขึ้น

5.1 สรุปผลการทดลอง

จากการวิจัยและดำเนินการทดลองจำแนกประเภทกิจการโดยอาศัยลักษณะเฉพาะที่ได้สังเคราะห์ขึ้นนั้นสามารถจำแนกประเภทได้จริงและได้ผลลัพธ์ความถูกต้องของการจำแนกแบ่งตามอัลกอริทึมที่ใช้งานได้ดังตารางที่ 5.1

ตารางที่ 5.1 สรุปผลการทดลอง

Algorithms	Accuracy
Naïve Bayes	81.25%
Neural Network	87.5%

ทั้งสองโมเดลให้ผลลัพธ์ที่อยู่ในเกณฑ์ดี ซึ่งการจำแนกด้วยนิวรัลเน็ทเวิร์คนั้นให้ค่าเที่ยงตรงที่สูงกว่า แต่ด้วยอัลกอริทึมที่มีความซับซ้อนกว่าจึงใช้เวลาในการประมวลผลนานกว่าเช่นกัน หากนำไปประยุกต์ใช้กับการจำแนกฐานข้อมูลขนาดใหญ่ อาจเป็นวิธีการที่ไม่เหมาะสมเพราะต้องใช้ทรัพยากรในการประมวลมากและใช้เวลานาน เมื่อเปรียบเทียบกับนาอีฟเบย์ซึ่งเป็นอัลกอริทึมที่ไม่ซับซ้อน สามารถประมวลผลได้รวดเร็วและให้ผลลัพธ์ที่อยู่ในเกณฑ์ดีเช่นกัน จึงเห็นควรนำนาอีฟเบย์ไปประยุกต์ใช้งานจริงจะมีความเหมาะสมกับสภาวการณ์มากกว่า

ดังนั้นการจำแนกประเภทกิจการด้วยนาอีฟเบย์โดยอาศัยลักษณะเฉพาะ (Attribute) ที่ได้จากการสังเคราะห์ข้อมูลการใช้ไฟฟ้าของมิเตอร์ระบบ AMR สามารถช่วยให้การไฟฟ้าส่วนภูมิภาคลด

ภาระงานและค่าใช้จ่ายในการสำรวจข้อมูลประเภทกิจการแต่ละรายเพื่อปรับปรุงฐานข้อมูลให้เป็นปัจจุบัน ซึ่งข้อมูลที่ได้จากวิธีการที่เสนอนี้ยังให้ผลลัพธ์ความถูกต้องสูงถึงร้อยละ 81.25

5.2 ปัญหาและข้อเสนอแนะ

5.2.1 ปัญหาที่พบในงานวิจัย

จะเห็นได้จากผลการทดลองนั้นยังพบข้อผิดพลาดจากการจำแนกประเภทกิจการอยู่ เนื่องจากผู้ใช้ไฟฟ้าในแต่ละประเภทยังสามารถแตกรายละเอียดลงได้อีก เช่น ประเภทโรงงานอุตสาหกรรมมีหลายหลายประเภท ทำให้มีลักษณะการใช้ไฟฟ้าที่แตกต่างกัน ส่งผลให้โมเดลที่สร้างขึ้นมานั้นยังไม่สามารถครอบคลุมรูปแบบทั้งหมดได้

5.2.2 ข้อเสนอแนะ

สามารถปรับปรุงเพิ่มเติมในงานวิจัยต่อไป ในการจำแนกแต่ละประเภทย่อยในกลุ่มใหญ่ เช่น ประเภทอุตสาหกรรม ประกอบไปด้วย อุตสาหกรรมเหล็ก อุตสาหกรรมแปรรูปไม้ อุตสาหกรรมสิ่งทอ เป็นต้น นอกจากนี้ในการประยุกต์ใช้งานกับระบบจริงอาจพบปัญหาข้อมูลที่ได้จากระบบ AMR มีความคลาดเคลื่อนเนื่องจากอุปกรณ์ชำรุด ซึ่งอาจต้องใช้วิธีการในการปรับปรุงข้อมูลหน่วยการใช้ไฟฟ้าเสียก่อนทำการจำแนกประเภทกิจการต่อไป

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

กองการไฟฟ้า คณะกรรมการนโยบายพลังงานแห่งชาติ สำนักนายกรัฐมนตรี. (2545). ทำอย่างไรให้ได้ประโยชน์จากอัตราค่าไฟฟ้าแบบ TOU. สืบค้นจาก www.eppo.go.th/images/Power/pdf/tou-intro.pdf

โครงการพัฒนาการอ่านหน่วยด้วยระบบอัตโนมัติ การไฟฟ้าส่วนภูมิภาค. (2557), AMR system [Online]. สืบค้นจาก http://service.amr.pea. co.th

ชัยพร เขมะภาตะพันธ์ และ นนท์ บุญนิธิประเสริฐ. (2552). การกรองข้อความภาษาไทยและภาษาอังกฤษของบริการส่งข้อความสั้นบนเครือข่ายโทรศัพท์เคลื่อนที่. National Conference on Computing and Information Technology (NCCIT) 5th

บุญเสริม กิจศิริกุล. (2546). อัลกอริทึมการทำเหมืองข้อมูล. ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ภูรดา นนทวาสี. (2554). การแบ่งกลุ่มข้อความ SMS ตามลักษณะการให้บริการ. สาขาวิศวกรรมคอมพิวเตอร์และโทรคมนาคม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์.

เอกสิทธิ์ พัชรวงศ์ศักดา. (2559). Practical Data Mining with RapidMiner Studio 7. กรุงเทพฯ: บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด.

ภาษาต่างประเทศ

Gheorghe Grigoras, Ovidiu Ivanov, and Mihai Gavrilas. (2014). Customer Classification and Load Profiling using Data from Smart Meters. NEUREL2014.

Published online in www.eppo.go.th/index.php/th/planpolicy/climatechange/thailand/eqmplan/item/11983-announce-edrp

Published online in http://mcodesmart.wordpress.com/2011/04/14/history-of-automated-meter-reading-amr

Published online in https://coladev.com/machine-learning/neural-network/2017/02/22/neural-network-basic

Published online in http://dataminingtrend.com

Published online in www.slideshare.net/NontawatB/07-classification-3-neural-network

Sergio Ramos, Zita Vale. (2008). Data Mining Techniques to Support the Classification of MV Electricity Customer. IEEE2008:1-7.

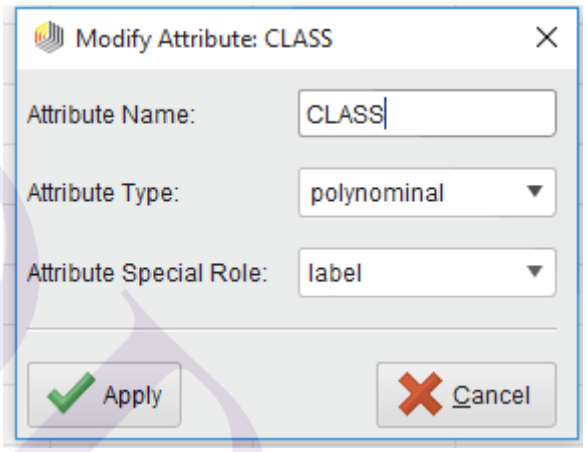
Worawat Ladarat and Sumate Naetiladdanon. (2015). PEA Automatic Meter Reading System : Progress and Lessons Learned. IEEE2015.

ภาคผนวก

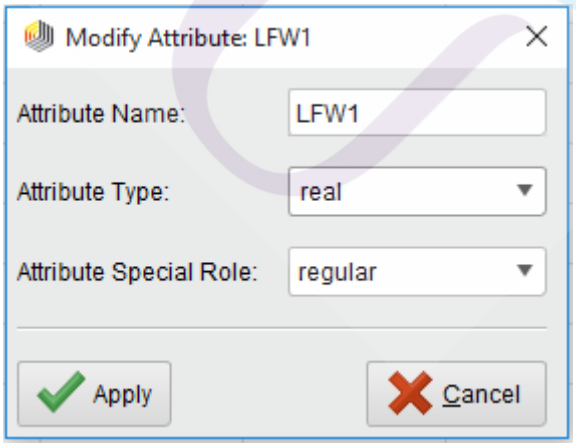
การดำเนินการในซอฟท์แวร์ RapidMiner Studio

1. การนำเข้าข้อมูลที่ได้เตรียมไว้ตามรูปแบบที่กำหนด ซึ่งจะต้องกำหนดชนิดของข้อมูลให้ถูกต้อง ดังนี้

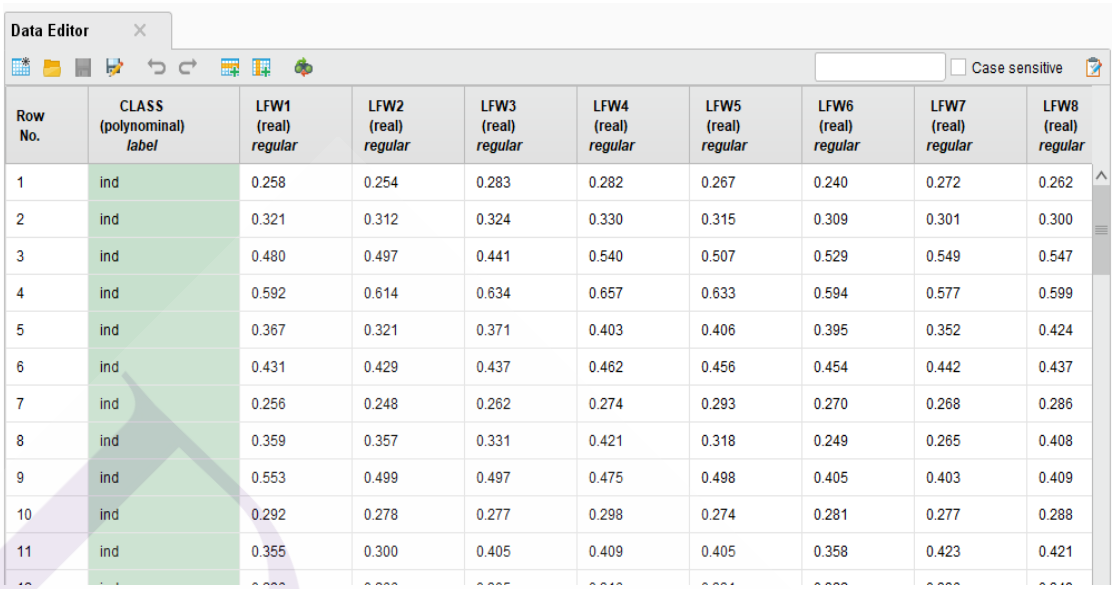
1.1 ประเภท หรือ Class ของผู้ใช้ไฟฟ้าแต่ละราย ให้กำหนด Attribute Name เป็น Class ส่วน Attribute Type และ Attribute Special Role ให้กำหนดเป็น Polynominal และ Label ตามลำดับ ดังนี้



1.2 ลักษณะประจำ (Attribute) ให้กำหนด Attribute Name ดังที่กำหนด ส่วน Attribute Type และ Attribute Special Role ให้กำหนดเป็น Real และ Regular ตามลำดับ ดังตัวอย่าง



เมื่อกำหนดชนิดของข้อมูลทั้งหมดเรียบร้อยแล้วจะได้ผลลัพธ์ดังภาพ



Data Editor

Case sensitive

Row No.	CLASS (polynominal) label	LFW1 (real) regular	LFW2 (real) regular	LFW3 (real) regular	LFW4 (real) regular	LFW5 (real) regular	LFW6 (real) regular	LFW7 (real) regular	LFW8 (real) regular
1	ind	0.258	0.254	0.283	0.282	0.267	0.240	0.272	0.262
2	ind	0.321	0.312	0.324	0.330	0.315	0.309	0.301	0.300
3	ind	0.480	0.497	0.441	0.540	0.507	0.529	0.549	0.547
4	ind	0.592	0.614	0.634	0.657	0.633	0.594	0.577	0.599
5	ind	0.367	0.321	0.371	0.403	0.406	0.395	0.352	0.424
6	ind	0.431	0.429	0.437	0.462	0.456	0.454	0.442	0.437
7	ind	0.256	0.248	0.262	0.274	0.293	0.270	0.268	0.286
8	ind	0.359	0.357	0.331	0.421	0.318	0.249	0.265	0.408
9	ind	0.553	0.499	0.497	0.475	0.498	0.405	0.403	0.409
10	ind	0.292	0.278	0.277	0.298	0.274	0.281	0.277	0.288
11	ind	0.355	0.300	0.405	0.409	0.405	0.358	0.423	0.421

2. การตั้งค่าพารามิเตอร์ในซอฟท์แวร์ RapidMiner Studio สำหรับทำการทดสอบและวัดค่าความเที่ยงตรงของชุดข้อมูลสำหรับสอน (Training set) ด้วยนาอีฟเบย์

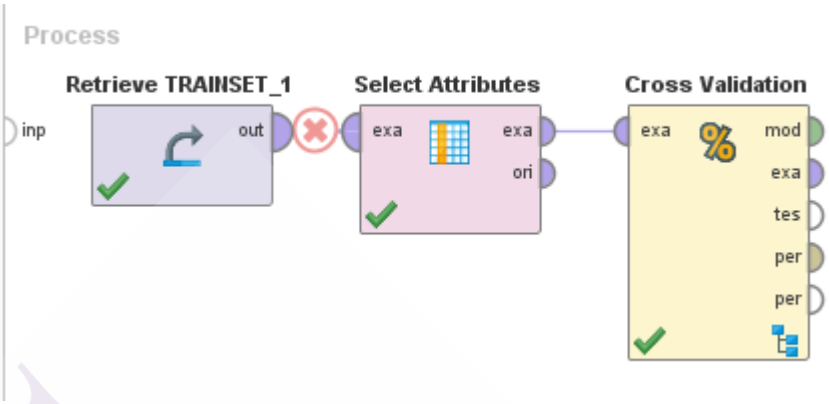
2.1 เลือก Operator ซึ่งประกอบด้วย

(1) Data Retrieve เป็นการนำเข้าชุดข้อมูลที่สร้างขึ้นสำหรับสร้างโมเดล

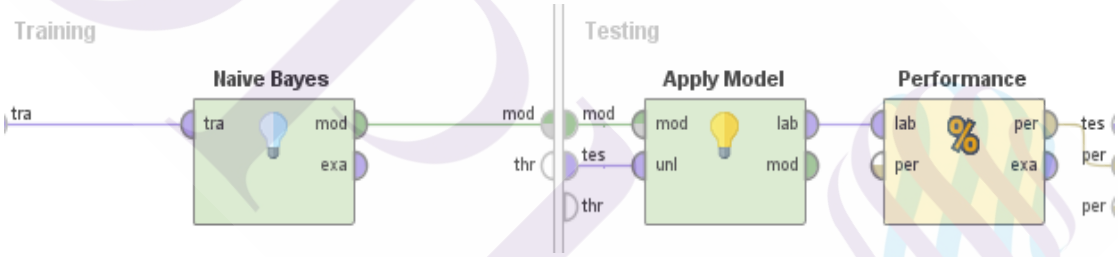
(2) Select Attribute เป็นการเลือก subset ของ Attribute เข้าหรือย้าย subset ของ Attribute ออกจากการคำนวณ

(3) Cross Validation เป็น operator สำหรับหาประสิทธิภาพการจำแนกของชุดข้อมูลที่ใช้สร้างเป็นโมเดล โดย Cross-validation Test เป็นวิธีการทดสอบประสิทธิภาพของโมเดลที่นิยมในการทำงานวิจัยเนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี Cross-validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายๆส่วน เช่น 5-fold-cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นหนึ่งในห้าส่วนของข้อมูลจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ ดังภาพที่ 4.3 ซึ่งในงานวิจัยนี้ได้เลือกใช้ 10-fold cross-validation

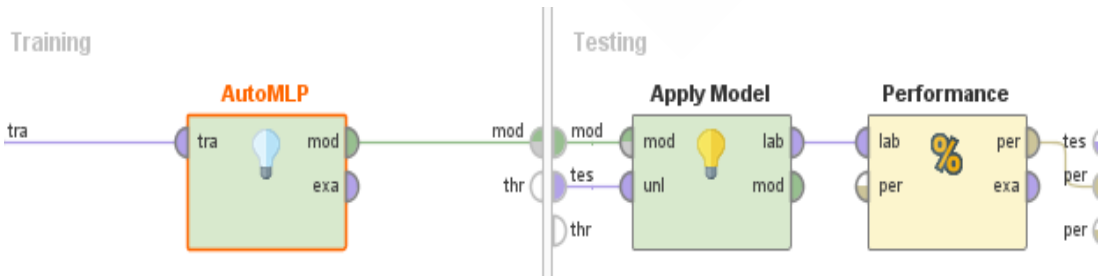
2.2 ลากเส้นเชื่อมโยง operator ต่างๆ เข้าด้วยกันเพื่อทำให้เกิด process



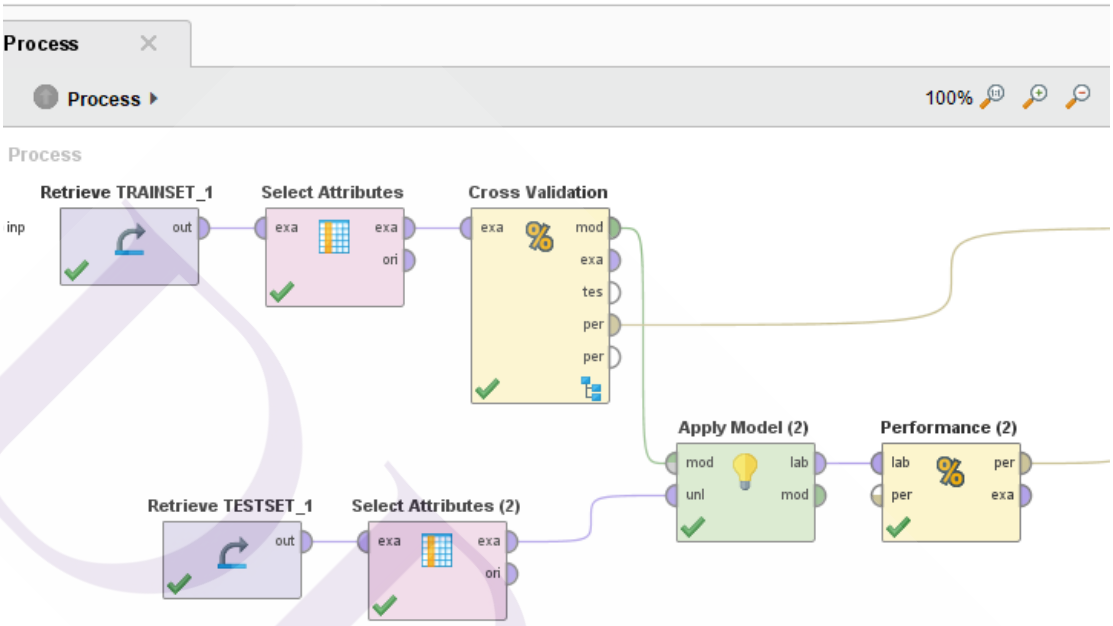
2.3 ทำการกำหนดค่าของ operator Cross Validation ซึ่งภายในจะแบ่งออกเป็น Training และ Testing ทำการเลือก operator Naïve Bayes ลงในฝั่ง Training และเลือก operator Apply Model และ Performance ลงในฝั่ง Testing จากนั้นลากเส้นเชื่อมโยง operator ดังภาพ



2.4 การตั้งค่าพารามิเตอร์ในซอฟท์แวร์ RapidMiner Studio สำหรับทำการทดสอบและวัดค่าความเที่ยงตรงของชุดข้อมูลสำหรับสอน (Training set) ด้วยนิวรัลเน็ทเวิร์ค ให้กำหนดทุกประการเช่นเดียวกับขั้นตอนที่ 2.3 ยกเว้น การกำหนด operator ภายใน Cross Validation ให้เปลี่ยนจาก operator Naïve Bayes เป็น Auto MLP ดังภาพ



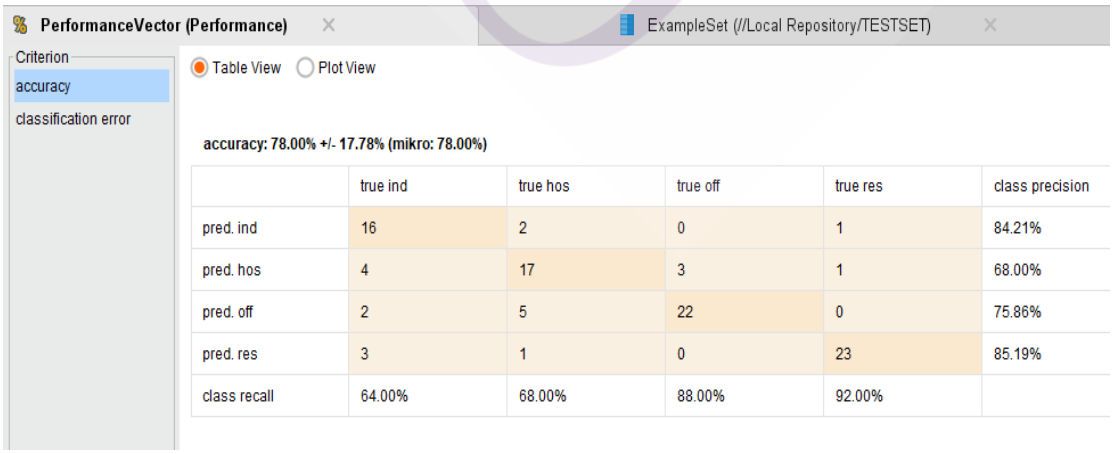
2.5 นำโมเดลที่สร้างไปทดสอบกับข้อมูลทดสอบ (Testing set) เพื่อวัดประสิทธิภาพการทำงานของโมเดลกับชุดข้อมูลทดสอบจำนวน 64 รายซึ่งไม่ซ้ำกับชุดข้อมูลฝึกสอน โดยเลือก operator และเชื่อมโยงเพื่อสร้าง process ดังภาพ



3. ผลการทดสอบ

3.1 ผลการทดสอบโมเดลที่สร้างด้วยนาอีฟเบย์

จากการทดสอบโมเดลที่สร้างนาอีฟเบย์สามารถแสดงผลการทดสอบด้วยค่าความเที่ยงตรงและค่าความคลาดเคลื่อนได้ดังนี้



	true ind	true hos	true off	true res	class precision
pred. ind	16	2	0	1	84.21%
pred. hos	4	17	3	1	68.00%
pred. off	2	5	22	0	75.86%
pred. res	3	1	0	23	85.19%
class recall	64.00%	68.00%	88.00%	92.00%	

PerformanceVector (Performance) | ExampleSet (//Local Repository/TESTSET)

Criterion
accuracy
classification error

Table View | Plot View

classification_error: 22.00% +/- 17.78% (mikro: 22.00%)

	true ind	true hos	true off	true res	class precision
pred. ind	16	2	0	1	84.21%
pred. hos	4	17	3	1	68.00%
pred. off	2	5	22	0	75.86%
pred. res	3	1	0	23	85.19%
class recall	64.00%	68.00%	88.00%	92.00%	

จะเห็นได้ว่าหลังจากการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับฝึกสอน (Training set) ด้วย นาอีฟเบย์แล้ว ให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 78 เกิดค่าความคลาดเคลื่อน (error) ที่ร้อยละ 22 และเมื่อนำโมเดลดังกล่าวไปใช้ในการจำแนกข้อมูลสำหรับทดสอบ (Testing set) จำนวน 64 ราย พบว่าให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 81.25 เกิดค่าความคลาดเคลื่อน (error) ที่ร้อยละ 18.75

PerformanceVector (Performance (2))

Table View | Plot View

accuracy: 81.25%

	true hos	true ind	true off	true res	class precision
pred. hos	13	1	1	1	81.25%
pred. ind	2	11	2	1	68.75%
pred. off	2	1	19	0	86.36%
pred. res	0	1	0	9	90.00%
class recall	76.47%	78.57%	86.36%	81.82%	

PerformanceVector (Performance (2))

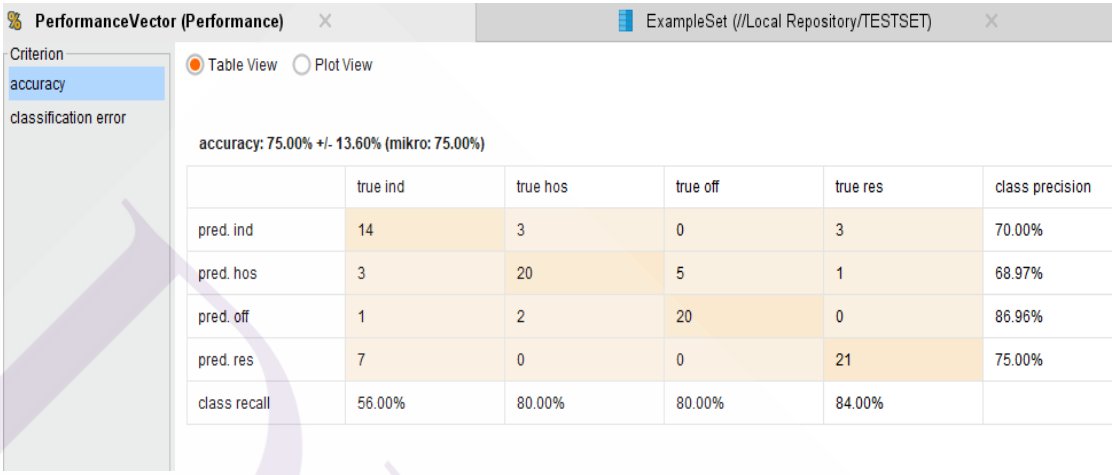
Table View | Plot View

classification_error: 18.75%

	true hos	true ind	true off	true res	class precision
pred. hos	13	1	1	1	81.25%
pred. ind	2	11	2	1	68.75%
pred. off	2	1	19	0	86.36%
pred. res	0	1	0	9	90.00%
class recall	76.47%	78.57%	86.36%	81.82%	

3.2 ผลการทดสอบโมเดลที่สร้างด้วยนิวรัลเน็ทเวิร์ค

จากการทดสอบโมเดลที่สร้างด้วยนิวรัลเน็ทเวิร์คสามารถแสดงผลการทดสอบด้วยค่าความเที่ยงตรงและค่าความคลาดเคลื่อนได้ดังนี้



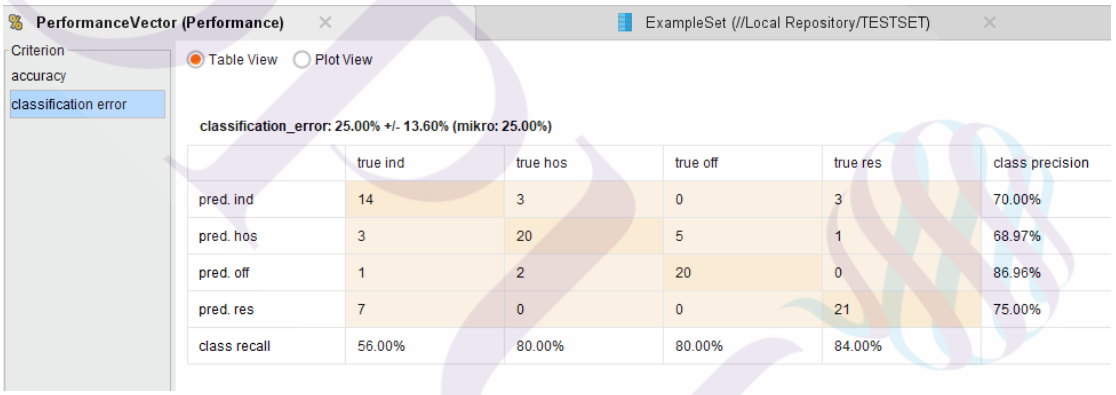
PerformanceVector (Performance) | ExampleSet (//Local Repository/TESTSET)

Criterion: accuracy | classification error

Table View | Plot View

accuracy: 75.00% +/- 13.60% (mikro: 75.00%)

	true ind	true hos	true off	true res	class precision
pred. ind	14	3	0	3	70.00%
pred. hos	3	20	5	1	68.97%
pred. off	1	2	20	0	86.96%
pred. res	7	0	0	21	75.00%
class recall	56.00%	80.00%	80.00%	84.00%	



PerformanceVector (Performance) | ExampleSet (//Local Repository/TESTSET)

Criterion: accuracy | classification error

Table View | Plot View

classification_error: 25.00% +/- 13.60% (mikro: 25.00%)

	true ind	true hos	true off	true res	class precision
pred. ind	14	3	0	3	70.00%
pred. hos	3	20	5	1	68.97%
pred. off	1	2	20	0	86.96%
pred. res	7	0	0	21	75.00%
class recall	56.00%	80.00%	80.00%	84.00%	

จะเห็นได้ว่าหลังจากการทดสอบโมเดลที่สร้างด้วยชุดข้อมูลสำหรับสอน (Training set) ด้วยนิวรัลเน็ทเวิร์คแล้ว ให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 75 เกิดค่าความคลาดเคลื่อน (error) ที่ร้อยละ 25 และเมื่อนำโมเดลดังกล่าวไปใช้ในการจำแนกข้อมูลสำหรับทดสอบ (Testing set) จำนวน 64 ราย พบว่าให้ผลความเที่ยงตรง (accuracy) ที่ร้อยละ 87.50 เกิดค่าความคลาดเคลื่อน (error) ที่ร้อยละ 12.50

PerformanceVector (Performance (2))

Table View Plot View

accuracy: 87.50%

	true hos	true ind	true off	true res	class precision
pred. hos	14	0	1	0	93.33%
pred. ind	2	11	0	0	84.62%
pred. off	1	1	21	1	87.50%
pred. res	0	2	0	10	83.33%
class recall	82.35%	78.57%	95.45%	90.91%	

PerformanceVector (Performance (2))

Table View Plot View

classification_error: 12.50%

	true hos	true ind	true off	true res	class precision
pred. hos	14	0	1	0	93.33%
pred. ind	2	11	0	0	84.62%
pred. off	1	1	21	1	87.50%
pred. res	0	2	0	10	83.33%
class recall	82.35%	78.57%	95.45%	90.91%	

ประวัติผู้เขียน

ชื่อ-นามสกุล	วุฒิชัย กำจรกิตติคุณ
ประวัติการศึกษา	พ.ศ. 2555 วิศวกรรมศาสตรบัณฑิต (เกียรตินิยมอันดับ 2) สาขาวิศวกรรมไฟฟ้า มหาวิทยาลัยธรรมศาสตร์
ตำแหน่งและสถานที่ทำงานปัจจุบัน	วิศวกร ระดับ 6 แผนกพัฒนาระบบมิเตอร์ กองพัฒนาระบบมิเตอร์ ฝ่ายระบบมิเตอร์และหม้อแปลง สายงานปฏิบัติการและบำรุงรักษา การไฟฟ้าส่วนภูมิภาค
ผลงานทางวิชาการ	พ.ศ. 2559 การวิเคราะห์ความผิดปกติข้อมูลการใช้ไฟฟ้าด้วยนาอีฟเบย์ การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ ครั้งที่ 8 (NCIT 2016) พ.ศ. 2560 การจำแนกประเภทผู้ใช้ไฟฟ้าด้วยนาอีฟเบย์ งานประชุมวิชาการและนวัตกรรม กฟภ. ปี 2560 (PEACON 2017)