



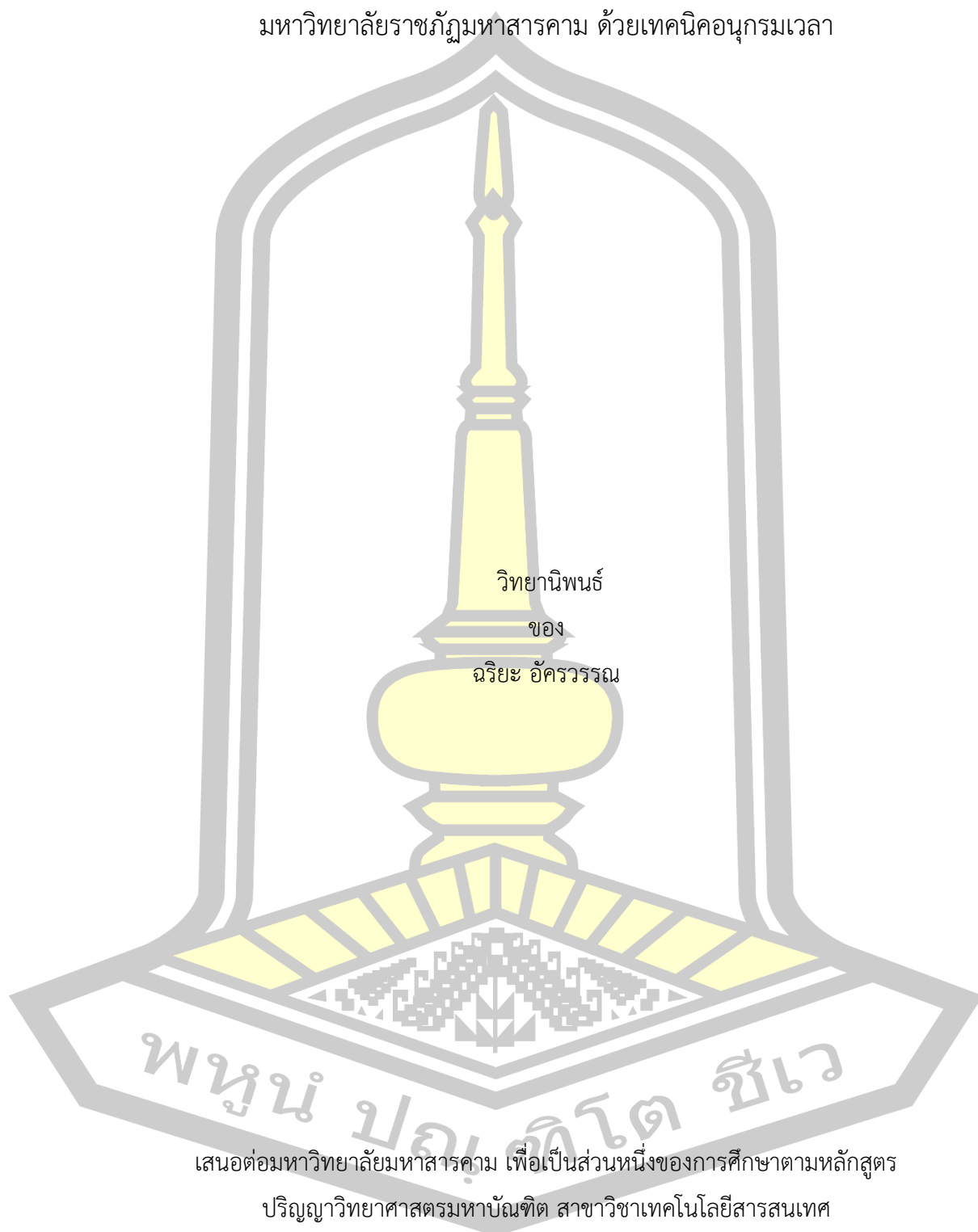
การพยากรณ์ปริมาณการใช้หนังสือในสำนักวิทยบริการและเทคโนโลยีสารสนเทศ  
มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคอนุกรมเวลา

วิทยานิพนธ์  
ของ  
ฉริยะ อัครวรรณ

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
ปีการศึกษา 2560

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การพยากรณ์ปริมาณการใช้หนังสือในสำนักวิทยบริการและเทคโนโลยีสารสนเทศ  
มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคอนุกรมเวลา



พูน ปอญติโต ชีเว

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

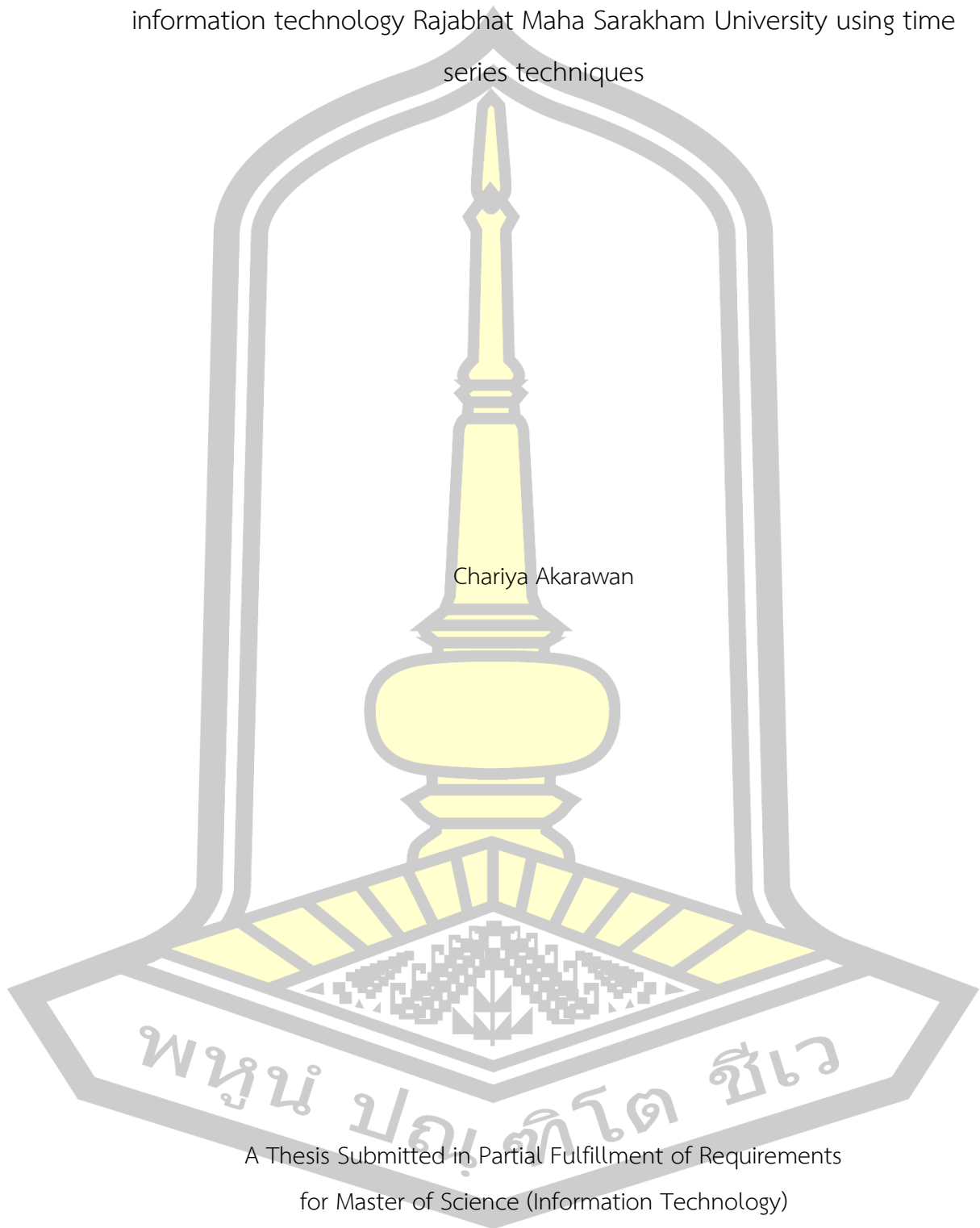
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ปีการศึกษา 2560

สงวนลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Forecasting the amount of book usage in academic resources and  
information technology Rajabhat Maha Sarakham University using time  
series techniques

Chariya Akarawan



A Thesis Submitted in Partial Fulfillment of Requirements  
for Master of Science (Information Technology)

Academic Year 2017

Copyright of Mahasarakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายฉริยะ อัครวรรณ แล้ว  
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตร์มหาบัณฑิต สาขาวิชา  
เทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการ

(ผศ. ดร. แกมกาญจน์ สมประเสริฐศรี )

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. จารีย์ ทองคำ )

กรรมการ

(อ. ดร. สาธิต แสงประดิษฐ์ )

กรรมการผู้ทรงคุณวุฒิภายนอก

(รศ. ดร. สิทธิชัย บุขหมั่น )

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญา วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

(ผศ. ดร. สุจิน บุตรดีสุวรรณ)

คณบดีคณะวิทยาการสารสนเทศ

(ผศ. ดร. กริสน์ ชัยมูล )

คณบดีบัณฑิตวิทยาลัย

วัน.....เดือน.....ปี.....

|                         |   |                   |                   |
|-------------------------|---|-------------------|-------------------|
| <b>ชื่อเรื่อง</b>       | การพยากรณ์ปริมาณการใช้หนังสือในสำนักวิทยบริการและเทคโนโลยี<br>สารสนเทศ<br>มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคอนุกรมเวลา |                   |                   |
| <b>ผู้วิจัย</b>         | ฉริยะ อัครวรรณ  |                   |                   |
| <b>อาจารย์ที่ปรึกษา</b> | ผู้ช่วยศาสตราจารย์ ดร. จาริ ทองคำ   |                   |                   |
| <b>ปริญญา</b>           | วิทยาศาสตรมหาบัณฑิต   | <b>สาขาวิชา</b>   | เทคโนโลยีสารสนเทศ |
| <b>มหาวิทยาลัย</b>      | มหาวิทยาลัยมหาสารคาม  | <b>ปีการศึกษา</b> | 2560              |

### บทคัดย่อ

การสร้างแบบจำลองเพื่อการพยากรณ์ปริมาณการใช้หนังสือในสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคอนุกรมเวลา เป็นข้อมูลที่สำคัญในการกำหนดนโยบายการจัดหาหนังสือใหม่ของผู้บริหารห้องสมุด ภายใต้งบประมาณที่มีอย่างจำกัด เพื่อทราบถึงปริมาณของหนังสือที่ความเหมาะสมของแต่ละหมวดหนังสือ ซึ่งส่งผลต่อความพึงพอใจของผู้ใช้บริการต่อคุณภาพการให้บริการของสำนักวิทยบริการ เปรียบเสมือนเป็นตัวชี้วัดว่าการจัดหาหนังสือประสบความสำเร็จหรือล้มเหลว ด้วยการนำข้อมูลการยืมหนังสือแบบรายวัน ระหว่างปี พ.ศ. 2557 – 2560 โดยได้รับการจัดหมู่หนังสือตามระบบทศนิยมของดิวอี้ มาสร้างแบบจำลองที่มีประสิทธิภาพ ประกอบด้วย 1) เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นแบบถดถอย 2) เทคนิคโครงข่ายประสาทเทียมแบบถดถอย 3) เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย 4) การวิเคราะห์การถดถอยโลจิสติก และ 5) การเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ แล้วเปรียบเทียบประสิทธิภาพแบบจำลอง ด้วยหลักการหน้าต่างเคลื่อนที่ แล้ววัดค่าด้วยการหาค่าเฉลี่ยคลาดเคลื่อนสมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง จากการศึกษาพบว่า เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย มีค่าต่ำสุด เท่ากับ  $9.42 \pm 2.57$  และ  $11.46 \pm 2.96$  อยู่ในระดับดีมากและดีตามลำดับ ซึ่งมีค่าใกล้เคียงค่าจริงมากที่สุด

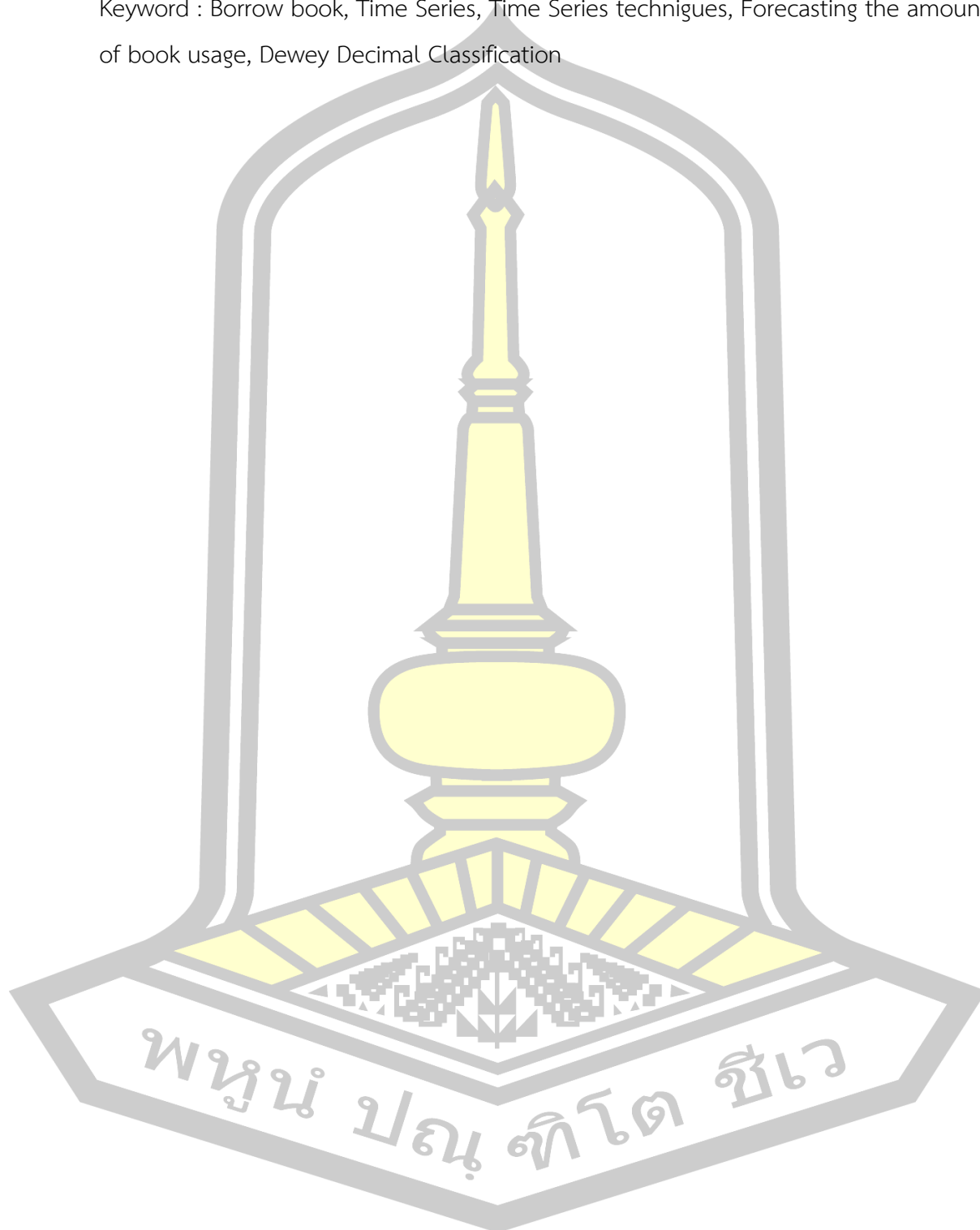
คำสำคัญ : การยืมหนังสือ, ข้อมูลอนุกรมเวลา, เทคนิคอนุกรมเวลา, การพยากรณ์การใช้หนังสือ, ระบบทศนิยมของดิวอี้

|                   |  |              |                        |
|-------------------|--|--------------|------------------------|
| <b>TITLE</b>      | Forecasting the amount of book usage in academic resources and information technology Rajabhat Maha Sarakham University using time series techniques |              |                        |
| <b>AUTHOR</b>     | Chariya Akarawan   |              |                        |
| <b>ADVISORS</b>   | Assistant Professor Jaree Thongkam , Ph.D.   |              |                        |
| <b>DEGREE</b>     | Master of Science  | <b>MAJOR</b> | Information Technology |
| <b>UNIVERSITY</b> | Maharakham University  | <b>YEAR</b>  | 2017                   |

### ABSTRACT

Modeling for Forecasting the amount of book usage in academic resources and information technology Rajabhat Maha Sarakham University using time series techniques. This technique is important to predict the new book supply of library managers. Therefore, it is virtuous to be established in the policy to manage academic resources, especially under a limited budget. Moreover, this technique is utilized to know the volume of books which can be suitable for each category. As a result, the users' satisfaction towards the service quality of the academic resources and information technology can indicate whether the book supply is successful or failed. The data of borrowed books were collected from 2014 to 2017. The Dewey Decimal of Classification was used to construct the effective models that consist of 1) Multi-Layer Perceptron Regression (MLPR) 2) Artificial Neural Network Regression (ANNR) 3) Support Vector Machine for Regression (SVMR) 4) Logistic Regression Analysis (LR) and 5) Reduced Error Pruning Tree (REPT). The models were compared with Sliding Windows method measured by values of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results showed that the borrowed books prospective of library members with the Support Vector Machine for Regression was in the highest efficiency rank. They were as low as 9.42 and 11.46, respectively

Keyword : Borrow book, Time Series, Time Series techniques, Forecasting the amount of book usage, Dewey Decimal Classification



## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ ได้ด้วยกรุณาอย่างสูงจาก ผู้ช่วยศาสตราจารย์ ดร.จารี ทองคำ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำปรึกษาตลอดจนปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่อย่างยิ่ง ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม และคณาจารย์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ที่ให้ความอนุเคราะห์ข้อมูลและคำปรึกษาในการดำเนินงานวิจัยครั้งนี้ให้สำเร็จลุล่วงด้วยดี

ผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยฉบับนี้จะมีประโยชน์ประโยชน์ต่อผู้ที่เกี่ยวข้อง จึงขอมอบส่วนดีทั้งหมดนี้ ให้แก่ คณาจารย์ที่ได้ประสิทธิ์ประสาทวิชาจนทำให้ผลงานวิจัยสำเร็จและขอมอบความกตัญญูทเวทิตาคุณ แต่บิดา มารดา และผู้มีพระคุณทุกท่านที่ห่วงใย ส่งเสริม เกื้อหนุน และเป็นกำลังใจให้เสมอมา

ฉวีระ อัครวรรณ

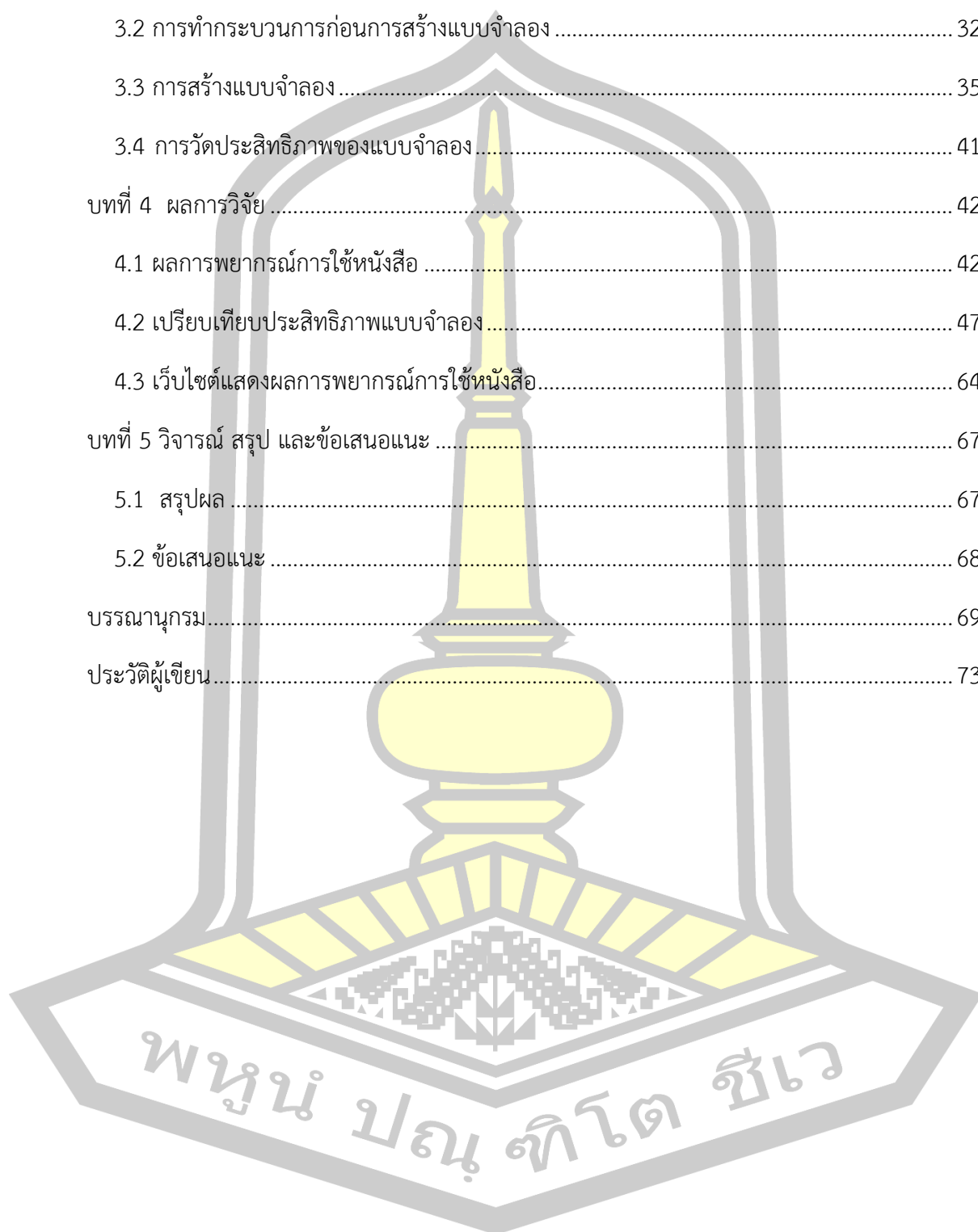




## สารบัญ

|  | หน้า |
|--|------|
| บทคัดย่อภาษาไทย.....   | ง    |
| บทคัดย่อภาษาอังกฤษ.....  | จ    |
| กิตติกรรมประกาศ.....   | ช    |
| สารบัญ.....  | ช    |
| สารบัญตาราง.....   | ญ    |
| สารบัญภาพประกอบ.....   | ฎ    |
| บทที่ 1 บทนำ.....  | 1    |
| 1.1 หลักการและเหตุผล .....   | 1    |
| 1.2 วัตถุประสงค์ของงานวิจัย .....  | 3    |
| 1.3 ความสำคัญของงานวิจัย.....  | 3    |
| 1.4 ขอบเขตของงานวิจัย.....   | 3    |
| 1.5 นิยามศัพท์เฉพาะ .....  | 4    |
| บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....                               | 5    |
| 2.1 สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ..... | 5    |
| 2.2 ความหมายของการจัดหมู่หนังสือ .....                                   | 6    |
| 2.3 ข้อมูลอนุกรมเวลา.....  | 9    |
| 2.4 การทำเหมืองข้อมูล.....   | 12   |
| 2.5 เทคนิคในการสร้างแบบจำลองข้อมูลเหมืองข้อมูล.....                      | 13   |
| 2.8 การวัดประสิทธิภาพแบบจำลอง.....                                       | 22   |
| 2.9 งานวิจัยที่เกี่ยวข้อง.....   | 25   |
| บทที่ 3 วิธีการดำเนินการวิจัย .....                                      | 29   |

|   |    |
|---|----|
| 3.1 การเตรียมข้อมูล .....                       | 29 |
| 3.2 การทำกระบวนการก่อนการสร้างแบบจำลอง .....    | 32 |
| 3.3 การสร้างแบบจำลอง .....                      | 35 |
| 3.4 การวัดประสิทธิภาพของแบบจำลอง .....          | 41 |
| บทที่ 4 ผลการวิจัย .....                        | 42 |
| 4.1 ผลการพยากรณ์การใช้หนังสือ .....             | 42 |
| 4.2 เปรียบเทียบประสิทธิภาพแบบจำลอง .....        | 47 |
| 4.3 เว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ ..... | 64 |
| บทที่ 5 วิจารณ์ สรุป และข้อเสนอแนะ .....        | 67 |
| 5.1 สรุปผล .....                                | 67 |
| 5.2 ข้อเสนอแนะ .....                            | 68 |
| บรรณานุกรม .....                                | 69 |
| ประวัติผู้เขียน .....                           | 73 |



สารบัญตาราง

|   | หน้า |
|---|------|
| ตารางที่ 3.1 ตัวอย่างข้อมูลการใช้หนังสือเป็นรายวัน .....  | 30   |
| ตารางที่ 3.2 ค่าทางสถิติตามเลขหมวดหมู่ในระบบทศนิยมของดิวี่.....   | 32   |
| ตารางที่ 3.3 Class เลขหมวดหมู่ตัวเลขแทนชนิดของหนังสือ .....   | 33   |
| ตารางที่ 3.4 คำอธิบายตัวแปรและชนิดของตัวแปร.....  | 33   |
| ตารางที่ 3.5 การแปลงข้อมูลการใช้หนังสือแบบรายวัน.....   | 34   |
| ตารางที่ 4.1 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ ของแบบจำลองในการพยากรณ์<br>การใช้หนังสือ..... | 54   |
| ตารางที่ 4.2 การเปรียบเทียบค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สองของแบบจำลอง<br>ในการพยากรณ์.....     | 62   |



## สารบัญภาพประกอบ

หน้า

|                   |  |          |
|-------------------|--|----------|
| ภาพประกอบที่ 2.1  | สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม                                       | .....6   |
| ภาพประกอบที่ 2.2  | นายเมลวิล ดิวอี้ ผู้คิดค้นระบบทศนิยมของดิวอี้  | ..... 8  |
| ภาพประกอบที่ 2.3  | กราฟอนุกรมเวลาที่มีค่าแนวโน้มเป็นส่วนประกอบ  | ..... 10 |
| ภาพประกอบที่ 2.4  | กราฟอนุกรมเวลาที่มีความผันแปรตามฤดูกาล   | ..... 10 |
| ภาพประกอบที่ 2.5  | กราฟอนุกรมเวลาที่มีความผันแปรของวัฏจักรเป็นส่วนประกอบ  | ..... 11 |
| ภาพประกอบที่ 2.6  | กราฟอนุกรมเวลาที่มีความผันแปรที่ผิดปกติเป็นส่วนประกอบ  | ..... 12 |
| ภาพประกอบที่ 2.7  | แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย              | ..... 14 |
| ภาพประกอบที่ 2.8  | แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยโครงข่ายประสาทเทียมแบบถดถอย                            | ..... 15 |
| ภาพประกอบที่ 2.9  | แสดง Feature Space และการแบ่งข้อมูล  | ..... 17 |
| ภาพประกอบที่ 2.10 | แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์การถดถอยโลจิสติก   | ..... 19 |
| ภาพประกอบที่ 2.11 | แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยการตัดกิ่งต้นไม้ตัดสินใจด้วยวิธี Reduced-error pruning | ..... 21 |
| ภาพประกอบที่ 2.12 | แสดงการเคลื่อนที่ที่คล้ายหน้าต่างไปตามเวลาของ Sliding Window   | ..... 23 |
| ภาพประกอบที่ 3.1  | ขั้นตอนการวิเคราะห์อนุกรมเวลา  | .....29  |
| ภาพประกอบที่ 3.2  | กราฟแสดงปริมาณการใช้หนังสือแบบรายเดือน พ.ศ. 2557 - 2560  | ..... 31 |
| ภาพประกอบที่ 3.3  | ค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย   | ..... 36 |
| ภาพประกอบที่ 3.4  | ค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมแบบถดถอย   | ..... 37 |
| ภาพประกอบที่ 3.5  | ค่าพารามิเตอร์ของเทคนิคสนับสนุนเส้นสมมติแบบถดถอย   | ..... 38 |
| ภาพประกอบที่ 3.6  | ค่าพารามิเตอร์ของเทคนิคการวิเคราะห์การถดถอยโลจิสติก  | ..... 39 |
| ภาพประกอบที่ 3.7  | ค่าพารามิเตอร์ของเทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้                  | ..... 40 |
| ภาพประกอบที่ 3.8  | การพยากรณ์ข้อมูลการใช้ด้วยหลักการหน้าต่างเคลื่อนที่  | ..... 41 |





|  |    |
|--|----|
| ภาพประกอบที่ 4.27 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 500.....  | 59 |
| ภาพประกอบที่ 4.28 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 600.....  | 59 |
| ภาพประกอบที่ 4.29 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 700.....  | 60 |
| ภาพประกอบที่ 4.30 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 800.....  | 60 |
| ภาพประกอบที่ 4.31 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 900.....  | 61 |
| ภาพประกอบที่ 4.32 แสดงค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ของแบบจำลองในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด.....         | 63 |
| ภาพประกอบที่ 4.33 เว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม..... | 64 |
| ภาพประกอบที่ 4.34 การแสดงผลการพยากรณ์การใช้หนังสือเป็นกราฟ แบบรายเดือนแยกตามหมวด.....  | 65 |
| ภาพประกอบที่ 4.35 การแสดงผลแบบรายปี.....   | 65 |
| ภาพประกอบที่ 4.36 การเพิ่มข้อมูลด้วยการเพิ่มไฟล์นามสกุล CSV.....   | 66 |
| ภาพประกอบที่ 4.37 การเพิ่มข้อมูลด้วยการพิมพ์ข้อมูล.....  | 66 |

# บทที่ 1

## บทนำ

### 1.1 หลักการและเหตุผล

การดำเนินการจัดหาหนังสือใหม่ที่มีคุณภาพมาให้บริการของห้องสมุด ภายใต้งบประมาณที่มีอย่างจำกัด และผลกระทบจากราคาหนังสือที่สูงขึ้นทุกปี อีกทั้งปัญหาของการใช้บริการหนังสือในห้องสมุดลดลงเรื่อยๆ เพราะพฤติกรรมการใช้สารสนเทศที่เปลี่ยนแปลงไป ด้วยเทคโนโลยีในยุคนี้ ผู้ใช้บริการมีทางเลือกในการแสวงหาสารสนเทศมากขึ้น จึงเป็นกระบวนการที่ท้าทายและต้องมีการวางแผนงานอย่างรอบคอบ เพราะส่งผลต่อความพึงพอใจของผู้ใช้บริการ และคุณภาพการให้บริการของห้องสมุด

การจัดหาหนังสือเข้ามาให้บริการของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม แม้ว่าจะผ่านขั้นตอนการแนะนำการสั่งซื้อหนังสือจากผู้บริการมาแล้วก็ตาม แต่ก็ยังเกิดข้อคำถามในการจัดซื้อหนังสือเข้ามาให้บริการจากผู้บริหารของมหาวิทยาลัยว่าหนังสือที่จัดซื้อเข้ามามีปริมาณการใช้มากน้อยเพียงใด การเลือกหนังสือตรงตามความต้องการของผู้ใช้บริการส่วนใหญ่หรือไม่ อีกทั้งปริมาณของหนังสือที่ความเหมาะสมของแต่ละหมวดหนังสือกับอัตราการใช้งานควรเป็นเท่าไร เพราะถ้าหากเลือกไม่ตรงตามความต้องการแล้ว จะเป็นการทุ่มเงินจำนวนมากและไม่สามารถตอบสนองความต้องการได้ตรงจุด ถือเป็นการลงทุนที่สูญเปล่าและสิ้นเปลืองงบประมาณในการดูแลรักษา รวมถึงเสียพื้นที่ในการจัดเก็บอีกด้วย ดังนั้นเพื่อให้ทราบถึงการใช้หนังสือจึงต้องใช้ข้อมูลที่มีในระบบห้องสมุดนั้น คือ ข้อมูลการยืม ซึ่งเปรียบเสมือนเป็นตัวชี้วัดว่าการจัดหาหนังสือประสบความสำเร็จหรือล้มเหลว

สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม เป็นแหล่งเรียนรู้ที่สำคัญของมหาวิทยาลัยราชภัฏมหาสารคาม ในฐานะหอสมุดกลาง มีบทบาทสำคัญในการส่งเสริมการเรียน การสอนและการวิจัยของมหาวิทยาลัย โดยมีหน้าที่ให้บริการทรัพยากรสารสนเทศทุกประเภทและครอบคลุมทุกสาขาวิชาที่เปิดการเรียนการสอน ได้จัดให้มีบริการแนะนำสั่งซื้อหนังสือใหม่จากคณาจารย์ นักศึกษา นักเรียนและนักวิจัย เป็นประจำทุกปี และอีกส่วนหนึ่ง



ห้องสมุดเป็นผู้จัดหาและสั่งซื้อหนังสือมาให้บริการ เพราะหากห้องสมุดสามารถจัดซื้อหนังสือให้ตรงกับความต้องการของผู้ใช้บริการมากเท่าใด จะชี้ให้เห็นถึงความสำเร็จในการจัดหาหนังสือมากขึ้นไปด้วย

สำหรับข้อมูลของการใช้หนังสือได้ถูกบันทึกตามลำดับเวลาก่อนหลังในรูปแบบของอนุกรมเวลา (Time Series) การวิเคราะห์ข้อมูลอนุกรมเวลาเข้ามามีบทบาทในการพยากรณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคต ในอดีตหลักการที่นักวิจัยส่วนใหญ่ได้ใช้หลักการคำนวณ Moving average ปัจจุบันการวิเคราะห์ขั้นสูงเข้ามามีบทบาทในการวิเคราะห์ข้อมูลอนุกรมเวลา [1] เช่น การนำเทคนิค Data mining และ Machine Learning มาใช้ในการวิเคราะห์อนุกรมเวลา ซึ่งมีความถูกต้องสูง ความผิดพลาดต่ำ จึงได้รับการยอมรับและนำไปใช้ในหลายๆ งาน เช่นเดียวกับงานวิจัย ของ Igor Aizenberg et al [2] ได้นำเอาเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย มาใช้ในการสร้างแบบจำลอง เพื่อการพยากรณ์การผลิตน้ำมันในแต่ละวัน ผลการวิจัยพบว่า สามารถบรรลุตามเป้าหมายได้อย่างมีประสิทธิภาพ โดยมีผลการพยากรณ์ที่แม่นยำ Fan Zhang et al [3] ได้ศึกษาและพัฒนาแบบจำลองด้วยเทคนิคสนับสนุนเส้นสมมติแบบถดถอย มาใช้ในการสร้างแบบจำลองเพื่อพยากรณ์ถึงการใช้พลังงานในการสร้างอาคาร โดยผลการศึกษาแสดงให้เห็นว่า รูปแบบที่นำเสนอมีความแม่นยำสูงกว่าวิธีอื่น ๆ และมีข้อผิดพลาดร้อยละสัมบูรณ์เฉลี่ยที่ลดลง Edwin Baidoo and Jennifer L. Priestley [4] ได้นำเอาเทคนิคการวิเคราะห์การถดถอยโลจิสติก มาใช้ในการสร้างแบบจำลองเพื่อวิเคราะห์อัตราความถูกต้องในการพยากรณ์การผลิตชำระหนี้ของลูกค้าและตรวจสอบดัชนีประสิทธิภาพการทำการกำไรสูงสุด พบว่า การวิเคราะห์มีประสิทธิภาพสูง Geoffrey and Kelvin [5] ได้นำเอาเทคนิคการทำเหมืองข้อมูลต้นไม้ตัดสินใจแบบถดถอยและโครงข่ายประสาทเทียม มาใช้ในการสร้างแบบจำลอง เพื่อการเปรียบเทียบการคาดการณ์การใช้พลังงานไฟฟ้าในฮ่องกง พบว่า แบบจำลองต้นไม้ตัดสินใจและแบบจำลองโครงข่ายประสาทเทียมมีประสิทธิภาพดีมาก แต่มีนักวิจัยส่วนน้อยได้นำเทคนิคดังกล่าวมาใช้ในการพยากรณ์ข้อมูลห้องสมุด ซึ่งจากงานวิจัยของ Lorena Siguenza Guzman et al. [6] ได้ทบทวนวรรณกรรมในการประยุกต์ใช้การทำเหมืองข้อมูลกับห้องสมุด จากปี พ.ศ. 2541 - 2557 พบว่ามีเพียง 9 บทความเท่านั้น ที่กล่าวถึงเทคนิคการวิเคราะห์ข้อมูล

ดังนั้นงานวิจัยนี้จึงนำข้อมูลอนุกรมเวลา มาสร้างแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคเหมืองข้อมูลต่างๆ โดยการนำข้อมูลการใช้หนังสือจริง ของแต่ละวัน จาก 10 หมวดหนังสือ

ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ เริ่มจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง ธันวาคม พ.ศ. 2560 มาสร้างแบบจำลองที่มีประสิทธิภาพ ประกอบด้วย 1) เทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย 2) เทคนิคโครงข่ายประสาทเทียมแบบถดถอย 3) เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย 4) เทคนิคการวิเคราะห์การถดถอยโลจิสติก และ 5) เทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ จากนั้นนำมาเปรียบเทียบประสิทธิภาพแบบจำลอง ด้วยหลักการหน้าต่างเคลื่อนที่ แล้ววัดค่าด้วยการหาค่าเฉลี่ยคลาดเคลื่อนสมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ที่นิยมใช้กันมากที่สุด เพื่อนำแบบจำลองที่มีประสิทธิภาพสูงสุดไปใช้พยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

## 1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อพัฒนาแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
2. เพื่อพัฒนาเว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ เพื่อเป็นแนวทางในการจัดหาหนังสือมาให้บริการของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

## 1.3 ความสำคัญของงานวิจัย

1. ได้ผลการพยากรณ์แนวโน้มการใช้หนังสือ เพื่อทราบถึงปริมาณการใช้หนังสือในแต่ละหมวดของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
2. ได้เว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ เพื่อเป็นแนวทางในการจัดหาหนังสือมาให้บริการของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

## 1.4 ขอบเขตของงานวิจัย

งานวิจัยนี้ได้เก็บรวบรวมข้อมูลปริมาณการใช้หนังสือแบบรายวัน จาก 10 หมวดหนังสือ ตามการจัดหมวดหนังสือด้วยระบบทศนิยมดิวอี้ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม เพื่อให้ทราบปริมาณการใช้ในแต่ละหมวด โดยเริ่มจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง ธันวาคม พ.ศ. 2560 มีข้อมูล จำนวน 1,457 ชุด

### 1.5 นิยามศัพท์เฉพาะ

1. การใช้หนังสือ หมายถึง ปริมาณการยืมหนังสือในระบบห้องสมุด ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
2. แบบจำลองการพยากรณ์ หมายถึง แบบจำลองที่สร้างขึ้น เพื่อพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
3. อนุกรมเวลา หมายถึง ข้อมูลการใช้หนังสือที่ถูกบันทึกตามลำดับเวลาก่อนหลัง ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม



## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยได้ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อนำมาเป็นกรอบและแนวคิดในการศึกษาค้นคว้า การพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม โดยใช้เทคนิคอนุกรมเวลา ตามหัวข้อดังต่อไปนี้ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ความหมายของการจัดหมู่หนังสือ ข้อมูลอนุกรมเวลา การทำเหมืองข้อมูล การวิเคราะห์ข้อมูลอนุกรมเวลาด้วยการทำเหมืองข้อมูล เทคนิคในการสร้างแบบจำลองข้อมูลเหมืองข้อมูล ประกอบด้วย 1. เทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย 2. เทคนิคโครงข่ายประสาทเทียมแบบถดถอย 3. เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย 4. เทคนิคการวิเคราะห์การถดถอยโลจิสติก และ 5. เทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ การเปรียบเทียบประสิทธิภาพแบบจำลองด้วยหลักการหน้าต่างเคลื่อนที่ และวัดค่าด้วยการหาค่าเฉลี่ยคลาดเคลื่อนสมบูรณ์ (MAE) และค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (RMSE) งานวิจัยที่เกี่ยวข้อง ดังนี้

#### 2.1 สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม [7] เป็นแหล่งเรียนรู้ที่สำคัญของมหาวิทยาลัยราชภัฏมหาสารคาม ในฐานะหอสมุดกลาง มีบทบาทสำคัญในการส่งเสริม การเรียน การสอนและการวิจัยของมหาวิทยาลัย โดยมีหน้าที่ให้บริการทรัพยากรสารสนเทศทุกประเภทและครอบคลุมทุกสาขาวิชาที่เปิดการเรียนการสอน มีการจัดหาหนังสือมาให้บริการและมีการจัดหมวดหมู่หนังสือ โดยแบ่งเป็น 10 หมวดหมู่ใหญ่ ตามระบบทศนิยมของดิวอี้ ใช้ตัวเลขสามหลัก 000 – 900 แบ่งเนื้อหาวิชา และระบบห้องสมุดอัตโนมัติใช้ ระบบ VTLS ซึ่งมีการเก็บข้อมูลการใช้หนังสือในรูปแบบอนุกรมเวลา

สถานที่ตั้ง – ที่ติดต่อ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม 80 ถนนนครสวรรค์ ตำบลตลาด อำเภอเมือง จังหวัดมหาสารคาม โทร. 0-43713618, 043-722118-9 ต่อ 160 โทรสาร 0-4372-5433 อีเมล lib@rmu.ac.th ดังภาพประกอบที่ 2.1



ภาพประกอบที่ 2.1 สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม [7]

## 2.2 ความหมายของการจัดหมู่หนังสือ

การจัดหมวดหมู่หนังสือ (Classification) [8-9] คือ การจัดหนังสือในห้องสมุด นิยมแยกตามเนื้อหาวิชา โดยใช้ระบบการจัดหมวดหมู่ระบบใดระบบหนึ่ง โดยนิยมใช้หมวดเป็นตัวเลข หากมีเนื้อหาเดียวกันจะมีเลขหมู่เดียวกันหรือใกล้เคียงกัน เพื่อให้ผู้ใช้บริการสามารถค้นหาหนังสือที่มีเนื้อหาเกี่ยวข้องกันได้อย่างสะดวก

### 2.2.1 ความสำคัญของการจัดหมู่หนังสือ มีดังนี้

2.2.1.1 มีการจัดหมู่หนังสืออย่างเป็นระบบ หนังสือทุกเล่มมีสัญลักษณ์แทนเนื้อหาของหนังสือ

2.2.1.2 หนังสือที่มีเนื้อหาวิชาเดียวกันหรือคล้ายคลึงกันจะอยู่ในหมวดหมู่เดียวกัน ผู้ใช้บริการจึงมีโอกาสเลือกเนื้อเรื่องที่ต้องการจากหนังสือได้หลายเล่ม

2.2.1.3 หนังสือที่มีเนื้อเรื่องเกี่ยวเนื่องกันหรือสัมพันธ์กันจะอยู่ใกล้ๆ กัน ผู้ใช้บริการจึงหาหนังสือที่มีเรื่องราวเหมือนกันได้ง่ายขึ้น

2.2.1.4 ช่วยให้ผู้ใช้บริการค้นหาหนังสือได้อย่างสะดวกและรวดเร็ว และช่วยให้เจ้าหน้าที่จัดชั้นชั้นได้ถูกต้อง รวดเร็ว

2.2.1.5 ช่วยให้ทราบปริมาณหนังสือในแต่ละหมู่

2.2.2 ระบบการจัดหมู่หนังสือ ในปัจจุบันมีการจัดในระบบต่างๆ ดังนี้

2.2.2.1 ระบบทศนิยมของดิวอี้ (Dewey Decimal Classification : DC, DDC) ของ เมลวิลล์ ดิวอี้ (Melvil Dewey)

2.2.2.2 ระบบหอสมุดรัฐสภาอเมริกัน (Library of Congress Classification : LC) ของ เฮอ์เบิร์ต พูทน์ม (Derbert Putnam) และคณะบรรณารักษ์หอสมุดรัฐสภาอเมริกัน

2.2.2.3 ระบบเอ็กซ์แพนซีฟ (Expansive Classification) ของ ชาร์ลส์ แอมมี คัตเตอร์ (Chartes Ammi Cutter)

2.2.2.4 ระบบทศนิยมสากล (Universal Decimal Classification : UDC) ของ พอล อ็อตเล็ต (Paul Otlet) และองรี ลา ฟงแตน (Henri La Fontaine)

2.2.2.5 ระบบซัฟเจค (Subject Classification : SC) ของ เจมส์ ดัฟฟ์ บราวน์ (James Duff Brown)

2.2.2.6 ระบบโคลอน (Colon Classification : CC) ของ เอส. อาร์. รังกานาธาน (S.R. Ranganathan)

2.2.2.7 ระบบบรรณานุกรม (Bibliographic Classification : BC) ของ เฮนรี เอฟเวลิน บลิสส์ (Henry Evelyn Bliss)

สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ได้นำระบบการจัดหมวดหมู่หนังสือ ระบบทศนิยมของดิวอี้ มาใช้ในการจัดหมวดหมู่หนังสือ ซึ่งนับว่าเป็นระบบที่นิยมใช้กันอย่างแพร่หลาย โดยลักษณะการจัดหมวดหมู่หนังสือระบบทศนิยมของดิวอี้ มีดังต่อไปนี้

การจัดหมวดหมู่หนังสือระบบทศนิยมของดิวอี้ (Dewey Decimal Classification : DC, DDC) [10] คิดค้นโดย นายเมลวิลล์ ดิวอี้ (Melvil Dewey) บรรณารักษ์ชาวอเมริกัน ได้นำเสนอต่อคณะกรรมการห้องสมุดของวิทยาลัยและจัดพิมพ์เป็นรูปเล่มครั้งแรกเมื่อปี พ.ศ. 2419 และได้มีการปรับปรุงให้ทันสมัยอยู่เสมอและจัดพิมพ์ครั้งสุดท้ายเมื่อปี พ.ศ. 2551 เป็นการพิมพ์ครั้งที่ 22 ดังภาพประกอบที่ 2.2



ภาพประกอบที่ 2.2 นายเมลวิลล์ ดิวอี้ ผู้คิดค้นระบบทศนิยมของดิวอี้ [11]

ระบบนี้ใช้สัญลักษณ์เป็นตัวเลขแทนชนิดของหนังสือ โดยใช้ตัวเลขสามหลักและใช้จุดทศนิยมหลังเลขหลักร้อยในการแบ่งย่อยเนื้อหาวิชา ระบบนี้ง่ายต่อการใช้งาน รวมถึงสามารถเข้าใจและจดจำได้ง่าย จึงเป็นที่นิยมใช้กันแพร่หลายทั่วโลก รวมถึงประเทศไทยก็นำระบบนี้ทศนิยมของดิวอี้มาใช้งาน ทั้งในห้องสมุดโรงเรียน ห้องสมุดประชาชนและห้องสมุดวิทยาลัย ด้วยลักษณะของระบบนี้ มีความง่ายต่อการใช้งานและสะดวกในการให้บริการ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ จึงนำระบบดังกล่าว มาใช้ในการบริการและให้ดำเนินงาน ระบบทศนิยมของดิวอี้แบ่งหนังสือเป็นหมวดหมู่ใหญ่ ๆ ดังนี้

หมวด 000 เบ็ดเตล็ด ความรู้ทั่วไป บรรณารักษศาสตร์

หมวด 100 ปรัชญา จิตวิทยา

หมวด 200 ศาสนา

หมวด 300 สังคมศาสตร์

หมวด 400 ภาษาศาสตร์

หมวด 500 วิทยาศาสตร์ คณิตศาสตร์

หมวด 600 เทคโนโลยี หรือวิทยาศาสตร์ประยุกต์

หมวด 700 ศิลปกรรม และนันทนาการ

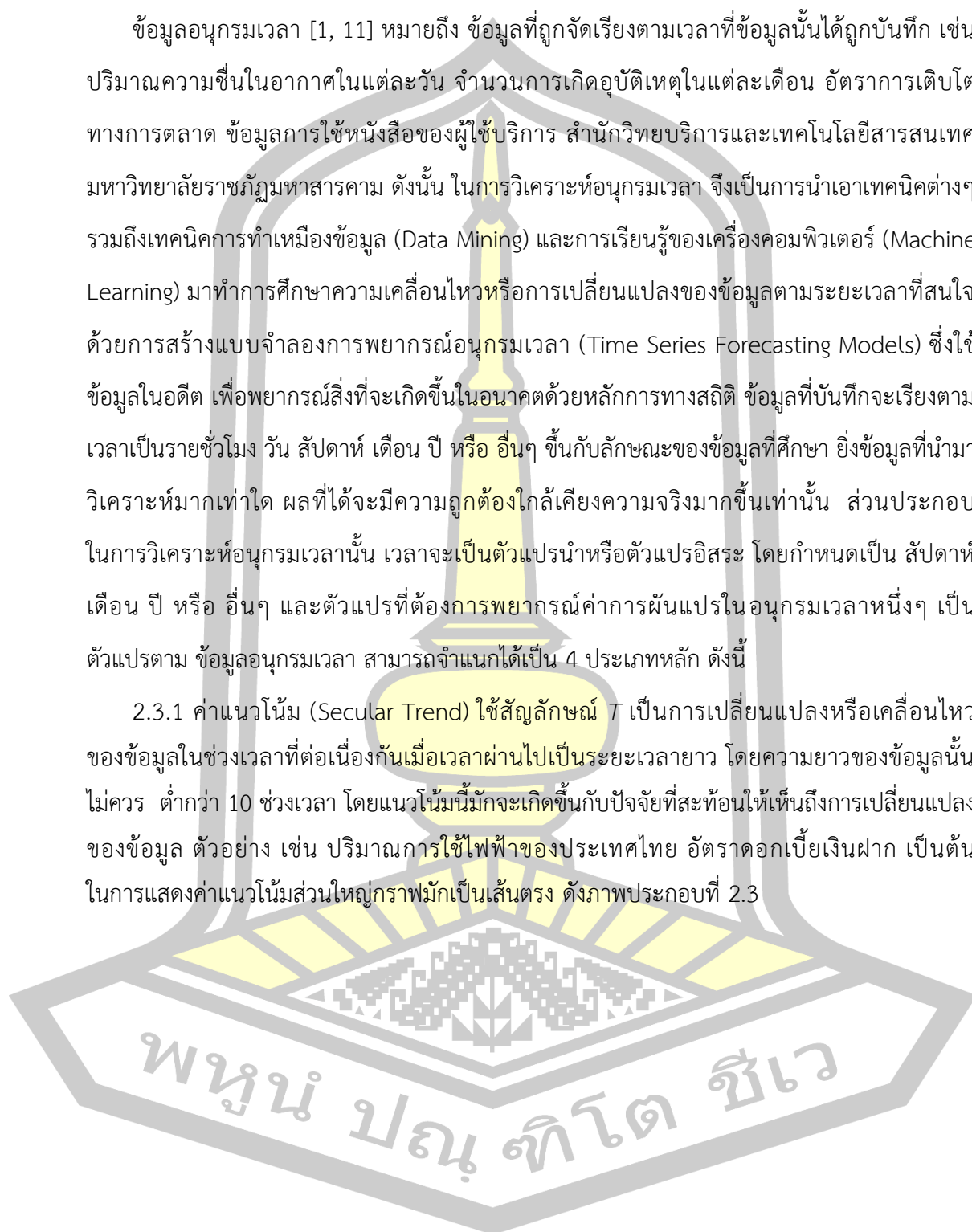
หมวด 800 วรรณคดี

หมวด 900 ภูมิศาสตร์และประวัติศาสตร์

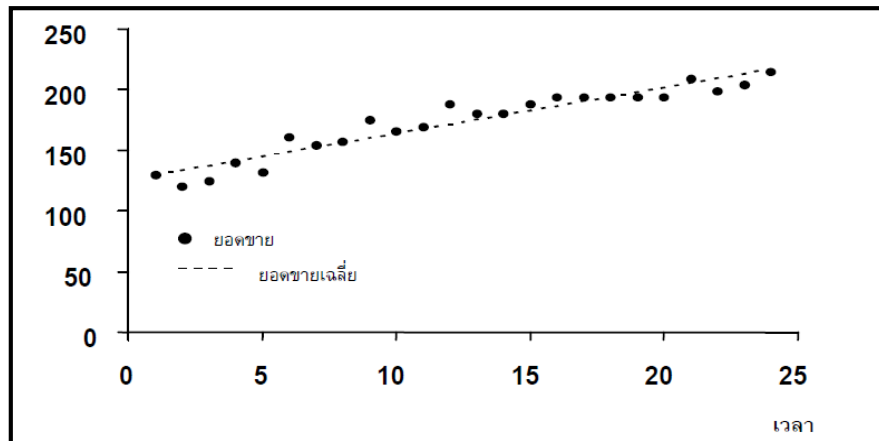
## 2.3 ข้อมูลอนุกรมเวลา

ข้อมูลอนุกรมเวลา [1, 11] หมายถึง ข้อมูลที่ถูกจัดเรียงตามเวลาที่ข้อมูลนั้นได้ถูกบันทึก เช่น ปริมาณความชื้นในอากาศในแต่ละวัน จำนวนการเกิดอุบัติเหตุในแต่ละเดือน อัตราการเติบโตทางการตลาด ข้อมูลการใช้หนังสือของผู้ใช้บริการ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ดังนั้น ในการวิเคราะห์อนุกรมเวลา จึงเป็นการนำเอาเทคนิคต่างๆ รวมถึงเทคนิคการทำเหมืองข้อมูล (Data Mining) และการเรียนรู้ของเครื่องคอมพิวเตอร์ (Machine Learning) มาทำการศึกษาความเคลื่อนไหวหรือการเปลี่ยนแปลงของข้อมูลตามระยะเวลาที่สนใจ ด้วยการสร้างแบบจำลองการพยากรณ์อนุกรมเวลา (Time Series Forecasting Models) ซึ่งใช้ข้อมูลในอดีต เพื่อพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตด้วยหลักการทางสถิติ ข้อมูลที่บันทึกจะเรียงตามเวลาเป็นรายชั่วโมง วัน สัปดาห์ เดือน ปี หรืออื่นๆ ขึ้นกับลักษณะของข้อมูลที่ศึกษา ยิ่งข้อมูลที่น่ามาวิเคราะห์มากเท่าใด ผลที่ได้จะมีความถูกต้องใกล้เคียงความจริงมากขึ้นเท่านั้น ส่วนประกอบในการวิเคราะห์อนุกรมเวลานั้น เวลาจะเป็นตัวแปรนำหรือตัวแปรอิสระ โดยกำหนดเป็น สัปดาห์ เดือน ปี หรืออื่นๆ และตัวแปรที่ต้องการพยากรณ์ค่าการผันแปรในอนุกรมเวลาหนึ่งๆ เป็นตัวแปรตาม ข้อมูลอนุกรมเวลา สามารถจำแนกได้เป็น 4 ประเภทหลัก ดังนี้

2.3.1 ค่าแนวโน้ม (Secular Trend) ใช้สัญลักษณ์  $T$  เป็นการเปลี่ยนแปลงหรือเคลื่อนไหวของข้อมูลในช่วงเวลาที่ต่อเนื่องกันเมื่อเวลาผ่านไปเป็นระยะเวลายาว โดยความยาวของข้อมูลนั้นไม่ควรต่ำกว่า 10 ช่วงเวลา โดยแนวโน้มนี้มักจะเกิดขึ้นกับปัจจัยที่สะท้อนให้เห็นถึงการเปลี่ยนแปลงของข้อมูล ตัวอย่าง เช่น ปริมาณการใช้ไฟฟ้าของประเทศไทย อัตราดอกเบี้ยเงินฝาก เป็นต้น ในการแสดงค่าแนวโน้มส่วนใหญ่กราฟมักเป็นเส้นตรง ดังภาพประกอบที่ 2.3

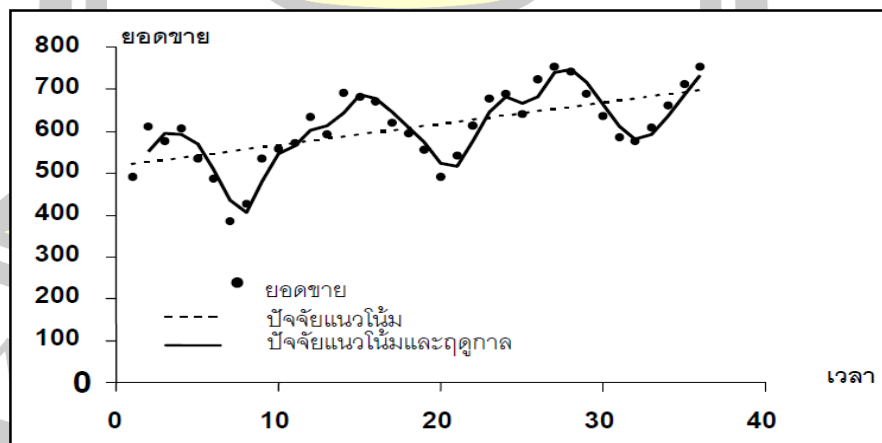






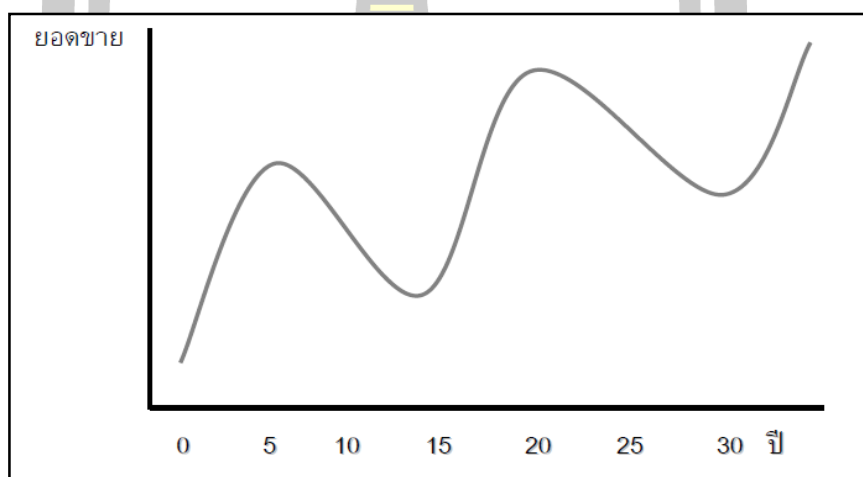
ภาพประกอบที่ 2.3 กราฟอนุกรมเวลาที่มีค่าแนวโน้มเป็นส่วนประกอบ [13]

2.3.2 การเปลี่ยนแปลงตามฤดูกาล (Seasonal Variation) ใช้สัญลักษณ์  $S$  เป็นการเคลื่อนไหวหรือเปลี่ยนแปลงของอนุกรมเวลาที่เกิดขึ้นซ้ำ ๆ กันในช่วงเวลาหนึ่ง โดยจะคล้ายกันในช่วงเวลาเดียวกัน จะเห็นได้อย่างเด่นชัดในอนุกรมเวลาของราคาผลผลิตทางการเกษตร เนื่องจากสิ่งเหล่านี้กระทบกระเทือนง่ายจากสภาวะทางธรรมชาติ เช่น ฤดูกาล ราคาผลไม้มักจะตกต่ำลงในฤดูเก็บเกี่ยว และจะมีราคาสูงขึ้นเมื่อผ่านพ้นฤดูเก็บเกี่ยวไปแล้ว และจะเกิดซ้ำ ๆ กันตามฤดูกาลในแต่ละปี ดังภาพประกอบที่ 2.4



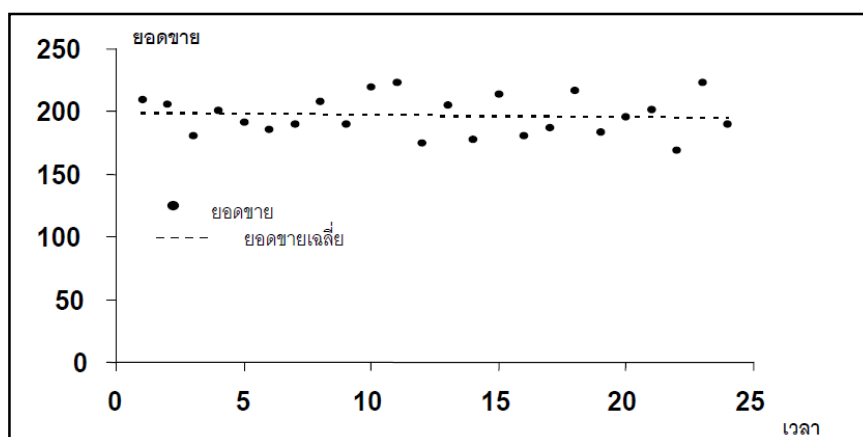
ภาพประกอบที่ 2.4 กราฟอนุกรมเวลาที่มีความผันแปรตามฤดูกาล [13]

2.3.3 การเปลี่ยนแปลงตามวัฏจักร (Cyclical Variation) ใช้สัญลักษณ์ C เป็นการเปลี่ยนแปลงของข้อมูลที่มีลักษณะการเคลื่อนที่ซ้ำๆ กัน คล้ายกับการเปลี่ยนแปลงตามฤดูกาล เพียงแต่การเปลี่ยนแปลงตามวัฏจักรใช้ระยะเวลายาวกว่า โดยแบบแผนของวัฏจักรนั้น ข้อมูลในแต่ละช่วงเวลาจะสั้นยาวไม่เท่ากัน ตามธรรมชาติของข้อมูลที่ได้ทำการสำรวจ ส่วนมากที่พบเสมอในการพยากรณ์ด้วยอนุกรมเวลา เช่น วัฏจักรเศรษฐกิจ (Economic Cycle) ปัจจัยที่ก่อให้เกิดการเปลี่ยนแปลงของเศรษฐกิจ คือ ผลกระทบจากเหตุการณ์ต่างๆ ทั้งในประเทศและนอกประเทศ รวมถึงเหตุการณ์ทางการเมืองต่างๆ โดยวัฏจักรหนึ่งๆอาจจะครอบคลุมเวลาตั้งแต่ 5 – 10 ปีขึ้นไป ดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.5 กราฟอนุกรมเวลาที่มีความผันแปรของวัฏจักรเป็นส่วนประกอบ [13]

2.3.4 การเปลี่ยนแปลงที่ผิดปกติ (Irregular Variation) ใช้สัญลักษณ์ I เป็นการเปลี่ยนแปลงของข้อมูลที่เกิดจากปัจจัยอื่นๆ นอกเหนือจากค่าแนวโน้ม การเปลี่ยนแปลงตามฤดูกาล หรือการเปลี่ยนแปลงตามวัฏจักร เป็นเหตุการณ์ไม่ได้เกิดขึ้นบ่อยและที่ไม่สามารถคาดเดาล่วงหน้าหรือพยากรณ์ได้ เช่น ภัยธรรมชาติ ภัยจากสงคราม ส่งผลให้การเปลี่ยนแปลงของข้อมูลอนุกรมเวลาไม่แน่นอนและมีความแปรปรวน ดังภาพประกอบที่ 2.6



ภาพประกอบที่ 2.6 กราฟอนุกรมเวลาที่มีความผันแปรที่ผิดปกติเป็นส่วนประกอบ [14]

## 2.4 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) [12–14] เป็นกระบวนการค้นหาวิธีการสร้างแบบจำลองมาทำการวิเคราะห์และสกัดความรู้จากความสัมพันธ์ในชุดข้อมูลจำนวนมาก โดยอัตโนมัติด้วยวิธีทางสถิติ เทคนิคการเรียนรู้ของเครื่องคอมพิวเตอร์ และการรู้จำแบบ เพื่อให้ได้องค์ความรู้หรือแบบจำลองที่ต้องการและนำองค์ความรู้ที่ได้มาใช้วางแผนการตัดสินใจในการบริหารหรือแก้ปัญหาต่าง ๆ ปัจจุบันการทำเหมืองข้อมูลมีบทบาทเป็นอย่างยิ่งในการวิเคราะห์การจัดซื้อหนังสือ เพื่อการบริหารงบประมาณที่มีประสิทธิภาพและเกิดประโยชน์สูงสุดแก่ผู้ใช้บริการ จำเป็นอย่างยิ่งที่ต้องทราบความต้องการ รวมถึงปริมาณที่เหมาะสมกับจำนวนหนังสือที่ต้องจัดหามาให้บริการ ซึ่งทราบได้จากการหาแนวโน้มการใช้หนังสือจากสถิติการใช้หนังสือด้วยการสร้างแบบจำลองการพยากรณ์การใช้หนังสือ เป็นต้น ทำให้เห็นว่าเหมืองข้อมูลมีความสำคัญต่อการนำมาวิเคราะห์ข้อมูลและการสร้างแบบจำลองการพยากรณ์ ซึ่งการทำเหมืองข้อมูลเพื่อสร้างแบบจำลองสามารถแบ่งออกได้ 2 ประเภทใหญ่ๆ ประกอบด้วย

### 2.4.1 แบบจำลองเพื่อการพยากรณ์ (Predictive Modeling หรือ Supervised Modeling)

เป็นการพยากรณ์อนาคต โดยนำข้อมูลในอดีตมาสร้างแบบจำลอง ด้วยการใช้ชุดข้อมูลสำหรับการสอน (Training Data) โดยทุกข้อมูลจะมีคุณสมบัติที่ใช้ในการทำนายผลของข้อมูล เรียกว่า ฉลาก (Label) โดยอัลกอริทึมจะแบ่งแยกข้อมูลเป็นกลุ่มตามฉลาก หากค่าฉลากมีค่าไม่ต่อเนื่อง จะเรียกกระบวนการที่ใช้ในการแบ่งแยกว่า การแยกแยะ (Classification) หากค่าฉลากมีค่าต่อเนื่อง จะเรียกกระบวนการนี้ว่า การถดถอย (Regression)

2.4.2 การสร้างแบบจำลองในการบรรยาย (Descriptive Modeling หรือ Unsupervised Modeling) ใช้ในการหาความสัมพันธ์ (Association) หรือการจัดกลุ่มข้อมูล (Clustering) โดยไม่มีวัตถุประสงค์ เพื่อการทำนาย

การทำเหมืองข้อมูลเพื่อวิเคราะห์อนุกรมเวลา [18] มี 4 ขั้นตอน ดังนี้

1. ขั้นตอนกระบวนการก่อนการสร้างแบบจำลอง (Pre-processing) ประกอบด้วย ขั้นตอน ดังต่อไปนี้

1.1 การศึกษารายละเอียดข้อมูล (Descriptive Data Summarization) เป็นการรวบรวมข้อมูลที่เกี่ยวข้อง โดยต้องศึกษาถึงแหล่งที่มาและการจัดเก็บข้อมูลที่มีความถูกต้อง เชื่อถือได้ และต้องเป็นข้อมูลที่มีรายละเอียดเพียงพอต่อการนำไปใช้

1.2 การกลั่นกรองข้อมูล (Data Cleaning) เป็นกระบวนการเตรียมข้อมูลให้สมบูรณ์ ที่สุดก่อนนำไปใช้ ซึ่งข้อมูลอาจเกิดจากความเสียหายได้ เช่น ข้อมูลขาดหาย (Missing Data) ข้อมูลรบกวน (Noisy Data) เป็นต้น

1.3 การแปลงข้อมูล (Data Transformation) เป็นการแทนค่าปัจจัย เพื่อให้ข้อมูลสามารถนำไปใช้ในกระบวนการสร้างแบบจำลอง

2. การสร้างแบบจำลอง (Modeling) เป็นการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์ข้อมูล เพื่อให้ได้แบบจำลองที่มีผลลัพธ์ที่ดีหรือใกล้เคียงค่าความจริงที่สุด

3. การวัดประสิทธิภาพ (Evaluation) เป็นการทดสอบความสามารถของแบบจำลอง ซึ่งสามารถวัดได้หลายค่า เช่น ค่าความถูกต้อง (Correctly) ค่าความแม่นยำ (Accuracy) และค่าความไว (Sensitivity) เป็นต้น

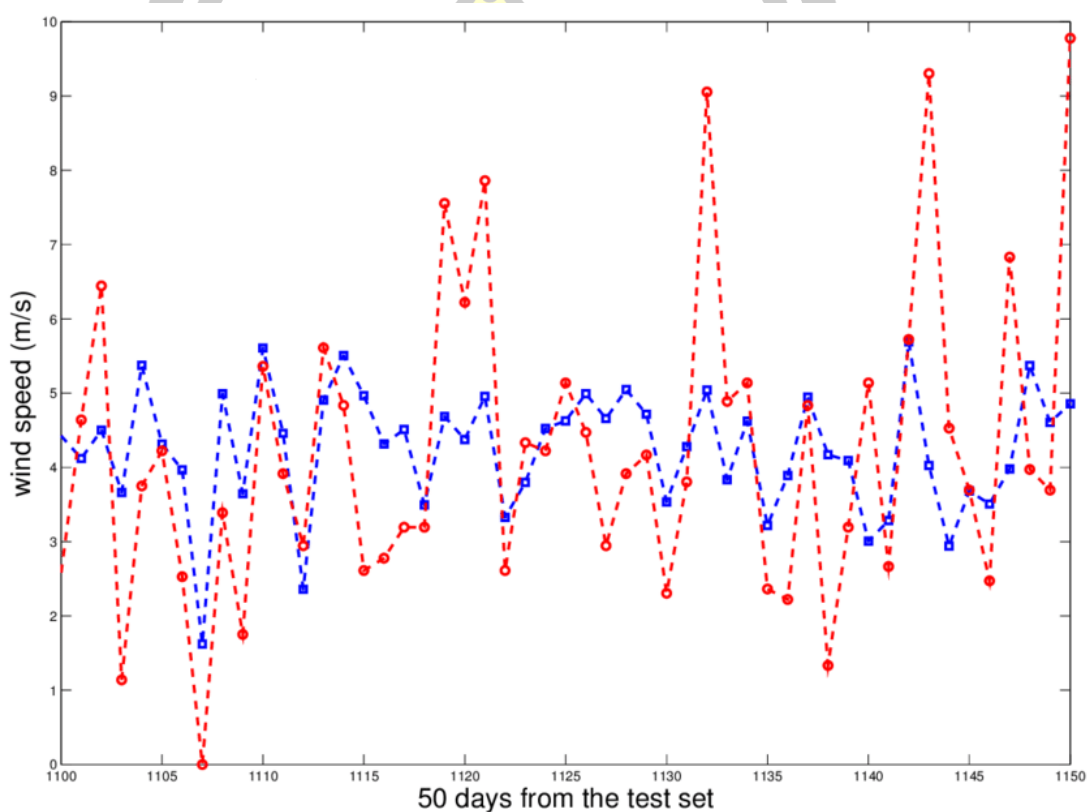
4. การแปลความหมายแบบจำลอง (Interpretation model) เป็นการนำแบบจำลองที่ได้ไปใช้งาน จากแบบจำลองที่มีผลลัพธ์ที่ดีหรือใกล้เคียงค่าความจริงที่สุด

## 2.5 เทคนิคในการสร้างแบบจำลองข้อมูลเหมืองข้อมูล

งานวิจัยนี้ใช้เทคนิคการทำเหมืองข้อมูล เพื่อสร้างแบบจำลองในการพยากรณ์ (Predictive Modeling, Supervised Modeling) โดยผู้วิจัยได้ทำการศึกษาค้นคว้าถึงข้อดีและประสิทธิภาพของแต่ละเทคนิค เพื่อนำมาสร้างแบบจำลอง ประกอบด้วย 1. เทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย 2. เทคนิคโครงข่ายประสาทเทียมแบบถดถอย 3. เทคนิคสนับสนุนเส้นสมมุติ

แบบถดถอย 4. เทคนิคการวิเคราะห์การถดถอยโลจิสติก และ 5. เทคนิคการเรียนรู้ต้นไม้ตัดสินใจ แบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ ซึ่งสามารถอธิบายได้ดังต่อไปนี้

2.5.1 เทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย (Multi-Layer Perceptron Regression) เป็นโครงข่ายประสาทเทียมแบบหลายชั้น ใช้กับงานที่มีความซับซ้อน จึงเพิ่มจำนวน โหนดการคำนวณขึ้นเพื่อให้ได้ผลลัพธ์ที่ดี ดังภาพประกอบที่ 2.7



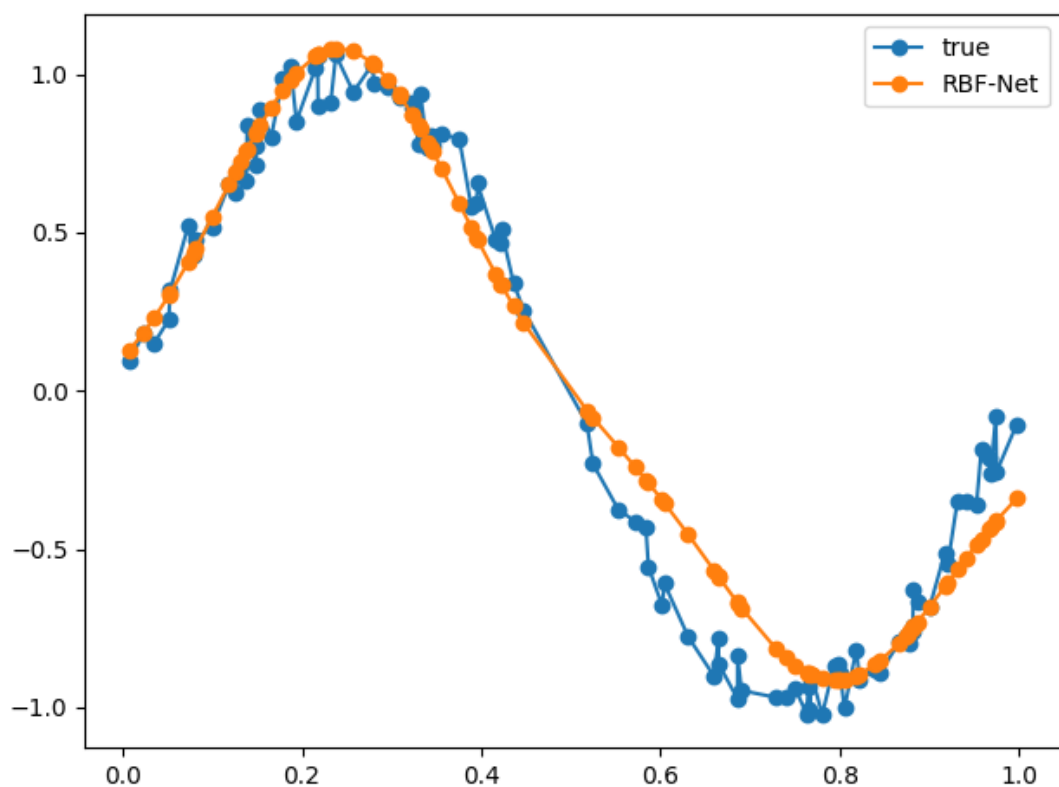
ภาพประกอบที่ 2.7 แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย [17]

จากคุณสมบัติดังกล่าว จึงมีนักวิจัยนำเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย มาใช้สร้างแบบจำลองในการพยากรณ์หลายท่าน อาทิ Igor Aizenberg et al. [2] ได้นำเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย (Multi-Layer Perceptron Regression) มาใช้สร้างแบบจำลองการถดถอยของอนุกรมเวลา เพื่อการพยากรณ์การผลิตน้ำมันในแต่ละวัน โดยใช้โครงข่ายประสาทเทียมแบบหลายชั้นและมีชุดข้อมูลเกี่ยวกับเซลล์ประสาทหลายค่า ได้ประเมินวิธีการที่เสนอ โดยใช้ชุดข้อมูลจริง ในการอธิบายลักษณะการทำงานแบบไดนามิกของสินทรัพย์ในบ่อน้ำมันซึ่งตั้งอยู่ในบริเวณหนองน้ำชายฝั่งของอ่าวเม็กซิโก โดยแสดงให้เห็นว่าเทคนิคนี้สามารถนำไปประยุกต์

ใช้กับตัวแปรเดียวและหลายตัวแปรได้ในแบบชั้นเดียวและหลายชั้นได้ ในการพยากรณ์ การเปลี่ยนแปลงของอ่างเก็บน้ำมันล่องหนานี้ ใช้การสร้างแบบจำลองเพื่อการพยากรณ์ โดยแบ่งออกเป็นการถดถอยและการจำแนก ซึ่งลักษณะเด่นของวิธีการที่นำเสนอในเอกสารฉบับนี้ คือ การพยากรณ์ตามเทคนิคนี้รวมคุณสมบัติทั้งสองวิธีไว้

ผลการวิจัย พบว่า ทั้งสองแบบสามารถใช้งานและบรรลุตามเป้าหมายได้อย่างมีประสิทธิภาพ โดยมีผลการพยากรณ์ที่แม่นยำในระยะเวลาที่ได้ดีกว่าที่คาดไว้

2.5.2 เทคนิคโครงข่ายประสาทเทียมแบบถดถอย (Artificial Neural Network Regression : ANNR) [15–19] จากคุณสมบัติเด่นของเทคนิคโครงข่ายประสาทเทียม จึงนำมาใช้ในการวิเคราะห์ และพัฒนาแบบจำลองกับข้อมูลอนุกรมเวลา เพื่อพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตได้อย่างแม่นยำ ซึ่งมีหลักการทำงาน ดังภาพประกอบที่ 2.8



ภาพประกอบที่ 2.8 แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยโครงข่ายประสาทเทียมแบบถดถอย [17]

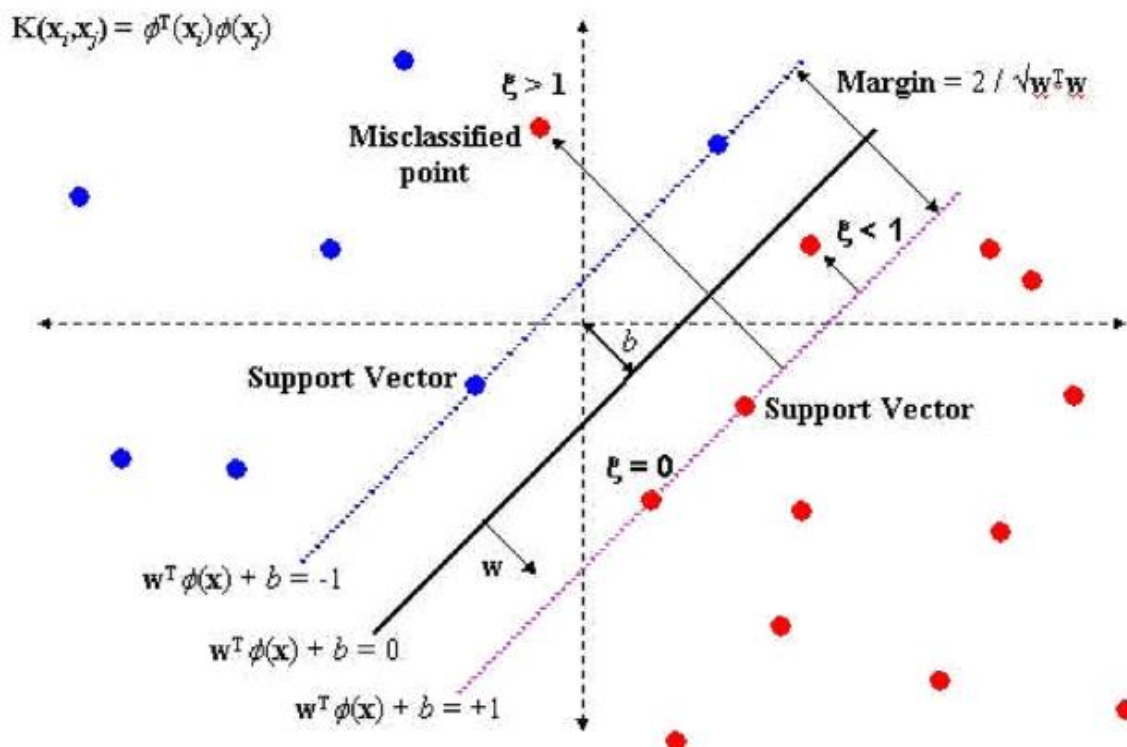
เขียนสมการ (17) ดังสมการ 2.1

$$If (sum (input * weight) > threshold) then output \quad (2.1)$$

สำหรับการพยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาล ได้มีนักวิจัยนำเทคนิคโครงข่ายประสาทเทียมแบบถดถอย มาใช้สร้างแบบจำลองหลายท่าน อาทิ G. Peter Zhang and Min Qi b [22] ได้พยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาลด้วยเทคนิคโครงข่ายประสาทเทียมแบบถดถอยที่ได้ผลลัพธ์ที่หลากหลาย ดังนั้น จึงได้ทำการตรวจสอบปัญหาเกี่ยวกับวิธีการสร้างชุดข้อมูลอนุกรมเวลาอย่างมีประสิทธิภาพ ตามรูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาลและแนวโน้ม พบว่า การนำข้อมูลดิบมาใช้ในการพยากรณ์ลดข้อผิดพลาดได้อย่างมากและเป็นวิธีการประมวลผลข้อมูลที่มีประสิทธิภาพมากที่สุด

2.5.3 เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (Support Vector Machine for Regression : SVMR) [17-18], [20-24] เป็นเทคนิคที่คล้ายกับโครงข่ายประสาทเทียม ด้วยการใช้ทฤษฎีการเรียนรู้จากสถิติเป็นพื้นฐานในการจำแนกข้อมูลเช่นเดียวกัน แต่เทคนิคสนับสนุนเส้นสมมุติมีความแตกต่าง คือ ใช้หลักการลดค่าความเสี่ยงเชิงโครงสร้างให้น้อยที่สุด (Structural Risk Minimization) เพื่อให้การพยากรณ์มีความผิดพลาดลดลง (Minimized Error) ด้วยการเพิ่มระยะการแบ่งข้อมูลให้มากที่สุด (Maximized Margin) โดยเทคนิคนี้ ใช้การนำข้อมูลในอดีตและปัจจุบัน มาทำการเรียนรู้ (Training) แล้วย้ายข้อมูลไปยัง Feature Space แล้วหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกข้อมูล หรือไฮเปอร์เพลน (Hyper-Plane) แล้วเพิ่มเส้นขอบ (Margin) ให้กับเส้นแบ่งทั้งสองด้าน ยิ่งเส้นขอบนี้มีความกว้างมากเท่าไรการแยกข้อมูลจะชัดเจนมากตามไปด้วย จึงเป็นเส้นแบ่งที่ดีที่สุด โดยเส้นขอบที่แบ่งข้อมูลนี้ จะแสดงเป็นเส้นประ และเรียกตำแหน่งสัมพัทธ์ข้อมูลที่ใกล้ที่สุดว่า ตำแหน่งสนับสนุนเวกเตอร์ (Support Vector) โดยการนำเทคนิคนี้ มาประยุกต์ใช้ในการแก้ไขปัญหการวิเคราะห์การถดถอยและการวิเคราะห์ข้อมูลอนุกรมเวลา เพื่อทำให้เกิดความผิดพลาดต่ำสุดที่ใช้สำหรับการพยากรณ์ เพื่อให้ได้รูปแบบที่ใช้ในการพยากรณ์ผลที่จะเกิดขึ้นในอนาคต ดังภาพประกอบที่ 2.9

พูน ปณ ทิโต ชีเว



ภาพประกอบที่ 2.9 แสดง Feature Space และการแบ่งข้อมูล [23]

จากคุณสมบัติดังกล่าว จึงมีนักวิจัยนำเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมาใช้สร้างแบบจำลองในการพยากรณ์ อาทิ Fan Zhang et al. [3] ได้ศึกษาและมีการพัฒนาแบบจำลองด้วยเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (Support Vector Regression : SVR) มาใช้ในการสร้างแบบจำลองการถดถอยของอนุกรมเวลาที่มีการถ่วงน้ำหนัก เพื่อพยากรณ์ถึงการใช้จ่ายพลังงานในการสร้างอาคาร เนื่องจากเป็นสิ่งสำคัญสำหรับการดำเนินงานและการจัดการอาคารที่มีประสิทธิภาพ โดยมีการเลือกพารามิเตอร์ของตัวประมวลผลวิวัฒนาการ (DE) เพื่อเพิ่มพารามิเตอร์ของรูปแบบและน้ำหนักให้กับ SVR ทั้งสองรุ่น ได้แก่ eps-SVR และ nu-SVR ทำให้มีประสิทธิภาพสำหรับงานนี้ โดยพยากรณ์มีความถูกต้องสูงมาก เนื่องจากชุดข้อมูลมีความซับซ้อนและไม่เป็นเชิงเส้น ซึ่งได้นำแบบจำลองไปใช้กับชุดข้อมูลสองชุด สำหรับการสร้างแบบเดียวกัน ได้แก่ ปริมาณการใช้พลังงานรายครึ่งชั่วโมงเป็นระยะเวลาสิบวัน และปริมาณการใช้พลังงานรายวันสำหรับการใช้พลังงานต่อปีเป็นเวลา 1 ปี พบว่า แบบจำลองที่เสนอสามารถคาดการณ์การใช้พลังงานสำหรับทั้งชุดข้อมูลที่มีความถูกต้องโดยไม่ต้องเปลี่ยนโครงสร้างแบบจำลอง ตัวอย่าง ข้อมูลการใช้พลังงานจากการสร้างอาคารในสิงคโปร์ เพื่ออธิบายถึงประสิทธิภาพของรูปแบบที่เสนอ ทั้งแบบครึ่งชั่วโมง

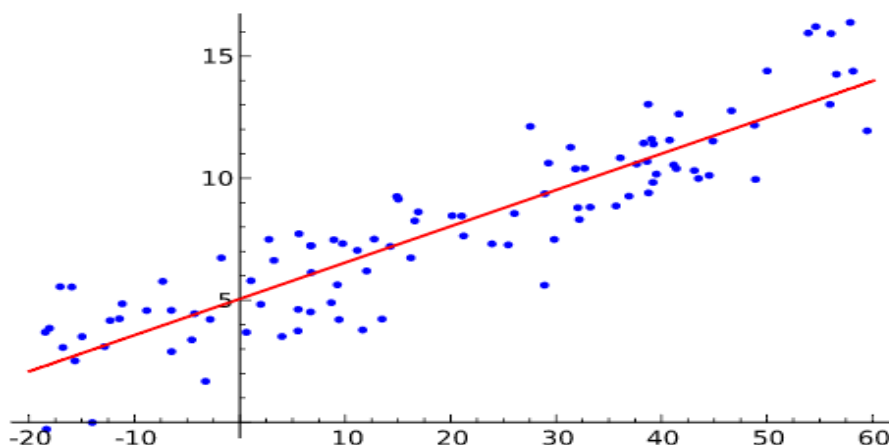


และข้อมูลการใช้พลังงานไฟฟ้าแบบรายวันสำหรับอาคารเดียวกัน ด้วยการใช้วิธีถ่วงน้ำหนักที่อัลกอริทึม DE ดังนั้น สำหรับการวิเคราะห์การใช้พลังงานรายวัน อัลกอริทึม DE จะกำหนดน้ำหนักที่สูงขึ้นให้กับ eps-SVR และสำหรับการวิเคราะห์การใช้พลังงานครึ่งชั่วโมง อัลกอริทึม DE จะกำหนดน้ำหนักที่สูงขึ้นให้กับ nu-SVR ดังนั้น รูปแบบเดียวสามารถคาดการณ์การใช้พลังงานได้ทุกวัน และทุกครึ่งชั่วโมงโดยไม่มีการแทรกแซงใดๆ

โดยผลการศึกษาแสดงให้เห็นว่า ข้อผิดพลาดร้อยละสัมบูรณ์เฉลี่ย (MAPE) สำหรับข้อมูลการใช้พลังงานรายวัน เป็น 5.843 และสำหรับการใช้พลังงานครึ่งชั่วโมงเป็น 3.767 ตามลำดับ และเมื่อการเปรียบเทียบกับพยากรณ์การใช้พลังงานรูปแบบอื่นๆ ซึ่งมักใช้สำหรับการเพิ่มประสิทธิภาพ เช่น GA และ PSO จะแสดงให้เห็นว่ารูปแบบที่นำเสนอมีความแม่นยำสูงกว่าวิธีอื่น ๆ และมี MAPE ที่ลดลงสำหรับทั้งชุดข้อมูลรายวันและชุดข้อมูลรายครึ่งชั่วโมง

ในการพยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาล ได้มีนักวิจัยนำเทคนิคสนับสนุนเส้นสมมติแบบถดถอย มาใช้สร้างแบบจำลองหลายท่าน อาทิ Ervin Ceperic and Vladimir Ceperic [25] ได้ใช้เทคนิคการพยากรณ์ระยะสั้น (STLF) และเทคนิค Support Vector Regression (SVR) โดยการสร้างแบบจำลองและการใช้อัลกอริทึมสำหรับการเลือกคุณลักษณะในการพยากรณ์นั้น ซึ่งได้ใช้เทคนิค SVR hyper – parameters เพิ่มประสิทธิภาพในการทำงาน และใช้เทคนิคการพยากรณ์ระยะสั้น (STLF) ในการพยากรณ์และเปรียบเทียบอัลกอริทึม ผลปรากฏว่า ได้ค่าความถูกต้องดีขึ้น

2.5.4 การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) [17-18], [20-25] เป็นการวิเคราะห์ข้อมูล เพื่อหาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ นำมาใช้ในการพยากรณ์ว่าจะเกิดเหตุการณ์หนึ่งขึ้นหรือไม่ หรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด โดยใช้สมการโลจิสติกที่สร้างขึ้นจากชุดตัวแปรทำนาย ทำให้ทราบถึงความสัมพันธ์ ทิศทางความสัมพันธ์ และลักษณะความสัมพันธ์ระหว่างตัวแปร ในการวิเคราะห์การถดถอย (Linear Regression) ที่ตัวแปรอิสระส่วนมากจะเป็นตัวแปรเชิงปริมาณ ส่วนตัวแปรตามจะต้องเป็นตัวแปรเชิงปริมาณเท่านั้น ดังภาพประกอบที่ 2.10



ภาพประกอบที่ 2.10 แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์การถดถอยโลจิสติก [27]

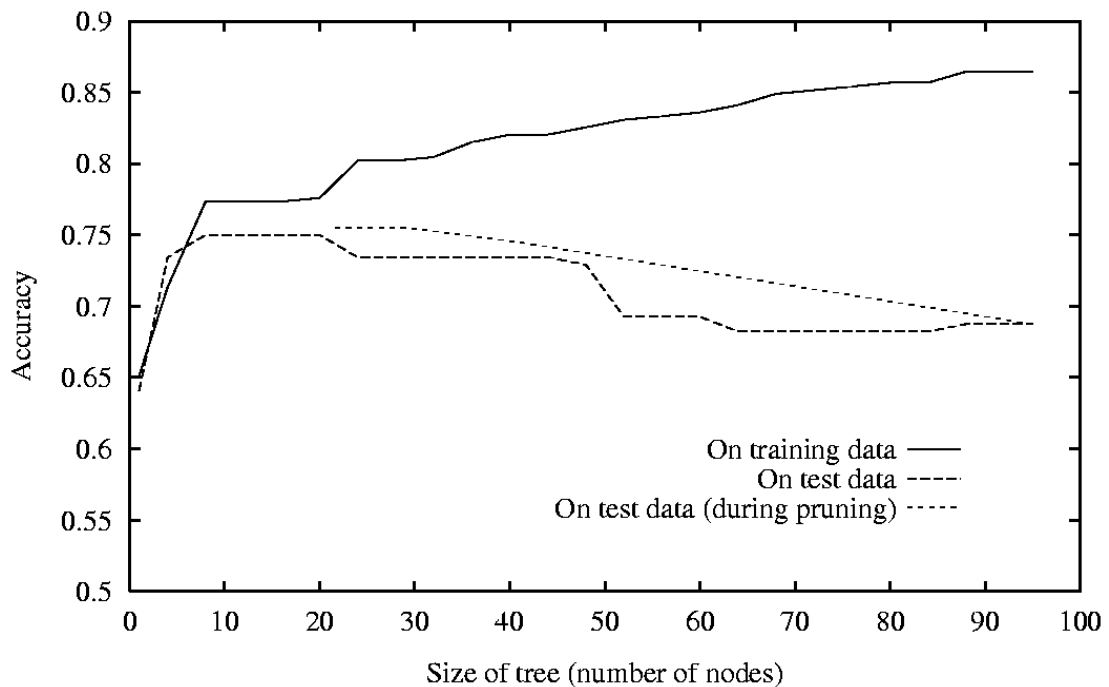
จากคุณสมบัติดังกล่าว จึงมีนักวิจัยนำเทคนิคการวิเคราะห์การถดถอยโลจิสติก มาใช้สร้างแบบจำลองในการพยากรณ์ เช่น งานวิจัยของ Edwin Baidoo and Jennifer L. Priestley [4] ได้นำเอาเทคนิคการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) มาใช้สร้างแบบจำลองการถดถอยของอนุกรมเวลา เพื่อวิเคราะห์อัตราความถูกต้องในการพยากรณ์การผิנדชำระหนี้ของลูกค้าหนี้และตรวจสอบดัชนีประสิทธิภาพการทำกำไรสูงสุด (EMP) ของบริษัทบนข้อสมมติฐานทางธุรกิจของตลาดสินเชื่อ ซึ่งมีแนวคิดเกี่ยวกับความล้มเหลวทางธุรกิจและการผิנדชำระหนี้ของลูกค้าผ่านทางการใช้อัตราส่วนทางการเงิน แม้ว่าการวิจัยก่อนหน้านี้ได้มุ่งเน้นไปที่เมตริกการประเมินผล เช่น ดัชนี AUC และ Gini แต่การวัดใหม่ที่ประยุกต์ใช้กับสร้างแบบจำลอง Logistic และ Time Series นี้ มีการตีความง่ายขึ้นสำหรับผู้บริหารและผู้มีอำนาจในการตัดสินใจต่างๆ พบว่า การวิเคราะห์อนุกรมเวลาด้วยเทคนิคการวิเคราะห์การถดถอยมีประสิทธิภาพดีกว่า ซึ่งมีความถูกต้อง เป็น 92.31%

ในการพยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาล ได้มีนักวิจัยนำเทคนิคการวิเคราะห์การถดถอยโลจิสติก มาใช้สร้างแบบจำลองหลายท่าน อาทิ Mohamed Addel-Aty et al. [28] ได้ทำการศึกษาวิจัยเกี่ยวกับการพยากรณ์ เพื่อลดความผิดพลาดบนทางด่วน ในระบบการจัดการจราจรขั้นสูงและระบบสารสนเทศ โดยใช้ตัวแปรการจราจรตามเวลาจริง ซึ่งได้ใช้เทคนิคการวิเคราะห์การถดถอยโลจิสติกในการพยากรณ์ เพื่อลดความผิดพลาดที่เกิดขึ้นจริงบนทางด่วน ในระบบการจัดการจราจร ที่เกิดความผิดพลาดเป็นอย่างมากบนทางด่วน ดังนั้น จึงได้นำความผิดพลาดที่เกิดขึ้นมาใช้

ในการคำนวณอัตราส่วน log-odds ซึ่งได้ใช้เกณฑ์สำหรับการกำหนดอัตราส่วนในการคำนวณอัตราส่วน log-odds อยู่ที่ 1.0 ผลปรากฏว่า ได้อัตราส่วน log-odds มากที่สุด 69%

2.5.5 การเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย (Decision Tree Regression) [16–18], [25–32] อัลกอริทึม (REP Tree) เป็นเทคนิคที่นิยมใช้ในเทคนิคอนุกรมเวลาสำหรับการพยากรณ์ข้อมูล มีรูปแบบโครงสร้างคล้ายต้นไม้ที่ประกอบไปด้วย โหนดตัดสินใจ (Decision Nodes) ใช้แสดง ค่าคุณสมบัติ (Attribute) ที่ใช้ทดสอบข้อมูล แล้วเชื่อมต่อกันด้วยกิ่ง (Branches) แต่ละกิ่ง จะแสดงผลในการทดสอบและในแต่ละกิ่งจะนำไปสู่โหนดตัดสินใจอื่นหรือโหนดใบ ซึ่งโหนดใบ จะเก็บผลลัพธ์ของการจำแนกไว้ โดยเลือกปัจจัยที่มีค่า Information Gain สูงที่สุดเป็นโหนดแรก เรียกว่า โหนดราก (Root Node) ในโหนดถัดลงมาจะถูกเรียกว่า โหนดลูก (Children Node) และ โหนดสุดท้ายเรียกว่า โหนดใบ (Leaf Node) เป็นโหนดตัดสินใจ

การเรียนรู้ต้นไม้ตัดสินใจด้วยอัลกอริทึม REP Tree เป็นอัลกอริทึมที่สามารถใช้ได้ ทั้งงาน Classification เช่น ทำนายว่าใครจะเป็นแชมป์ฟุตบอลโลก หรือใช้ในงาน Regression เช่น การพยากรณ์ราคาหุ้น โดยอัลกอริทึม REP Tree ได้รับการพัฒนามาจากอัลกอริทึม C4.5 การตัดกิ่งแบบความผิดพลาดลดลง (Reduced-error pruning) เป็นการแก้ปัญหาอันเกิดจากการสร้างต้นไม้ตัดสินใจ ซึ่งอาจมีการแตกกิ่งมากเกินไปหรือในแต่ละกิ่งอาจเกิดการสร้างอย่างผิดพลาด รวมถึงปัญหาการเจาะจงโมเดลกับข้อมูลมากเกินไป (Over Fitting) ได้ ดังนั้น การทำการเรียนรู้ต้นไม้ตัดสินใจที่ดีจึงต้องแก้ปัญหาเหล่านี้ด้วยเทคนิคการตัดกิ่งต้น ซึ่งการตัดกิ่งแบบความผิดพลาดลดลงนี้เป็นเทคนิคที่มีแนวคิดง่ายที่สุด โดยใช้ชุดข้อมูลการตัดกิ่ง (Pruning set) ที่แยกออกจากชุดข้อมูลฝึก ในการประเมินความถูกต้องของโหนดและใบของต้นไม้ โดยจะตรวจสอบโหนดจากล่างสุดขึ้นไปยังรากของต้นไม้ (Bottom-up Strategy) โดยใช้วิธีการเข้าถึงโหนด แบบ post-Order Traversal ด้วยการเปลี่ยนโหนดที่อยู่ข้างล่างโหนดลูกที่จะทดสอบให้เป็นโหนดใบ แล้วทดสอบด้วยชุดตัดกิ่ง โดยเปรียบเทียบกับจำนวนตัวอย่างที่ไม่ถูกต้องของโหนดลูกกับผลรวมของโหนดใบ ถ้าได้ค่าเปรียบเทียบซึ่งเป็นจำนวนตัวอย่างที่ไม่ถูกต้องของโหนดลูกน้อยกว่าผลรวมของโหนดใบ ให้ตัดโหนดลูกทิ้งแล้วเปลี่ยนเป็นโหนดใบของโหนดลูกชั้นต่อไป ผลที่ได้จะทำให้ต้นไม้ตัดสินใจมีขนาดเล็ก และมีค่าความผิดพลาดต่ำ การตัดกิ่งต้นไม้ตัดสินใจด้วยวิธีการตัดกิ่ง แบบความผิดพลาดลดลง ดังภาพประกอบที่ 2.11



ภาพประกอบที่ 2.11 แสดงรูปแบบกราฟที่เกิดจากการวิเคราะห์ข้อมูลด้วยการตัดกิ่งต้นไม้ตัดสินใจ ด้วยวิธี Reduced-error pruning [37]

Geoffrey and Kelvin [5] ได้นำเอาเทคนิคการทำเหมืองข้อมูล ต้นไม้ตัดสินใจแบบถดถอย (decision tree Regression) และโครงข่ายประสาทเทียม มาใช้ในสร้างแบบจำลองการถดถอยของอนุกรมเวลา เพื่อการเปรียบเทียบการคาดการณ์การใช้พลังงานไฟฟ้าในฮ่องกงกับการวิเคราะห์การถดถอยแบบเดิม แล้วนำมาประเมินประสิทธิภาพของแบบจำลองที่แตกต่างกันและเลือกรูปแบบที่เหมาะสมที่สุดสำหรับการคาดการณ์ในอนาคต เมื่อเปรียบเทียบความถูกต้องในการคาดการณ์การใช้พลังงานไฟฟ้า แบบจำลองต้นไม้ตัดสินใจและแบบจำลองโครงข่ายประสาทเทียมมีประสิทธิภาพดีกว่ารุ่นอื่นๆ ในช่วงฤดูร้อนและฤดูหนาวตามลำดับ

ในการพยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาล ได้มีนักวิจัยนำเทคนิคต้นไม้ตัดสินใจแบบถดถอย มาใช้สร้างแบบจำลองหลายท่าน อาทิ Xiuying Zhang et al. [38] ได้ทำการศึกษาวิจัยเกี่ยวกับการพยากรณ์การเกิดฝนกรด โดยใช้ตำแหน่งลักษณะภูมิศาสตร์ และลักษณะภูมิประเทศในการวิเคราะห์ปัจจัยทางด้านอุตุนิยมวิทยาความหนาแน่นของชั้นบรรยากาศ SO<sub>2</sub> และ NO<sub>2</sub> ซึ่งในงานวิจัยใช้เทคนิคต้นไม้ตัดสินใจแบบถดถอยในการพยากรณ์ ผลการวิจัยพบว่า ฝนกรดที่เกิดขึ้นใน 17 สถานี ในกลุ่ม 81 สถานี พยากรณ์ในจังหวัดเหลียวหนิงนั้น มีฝนกรด

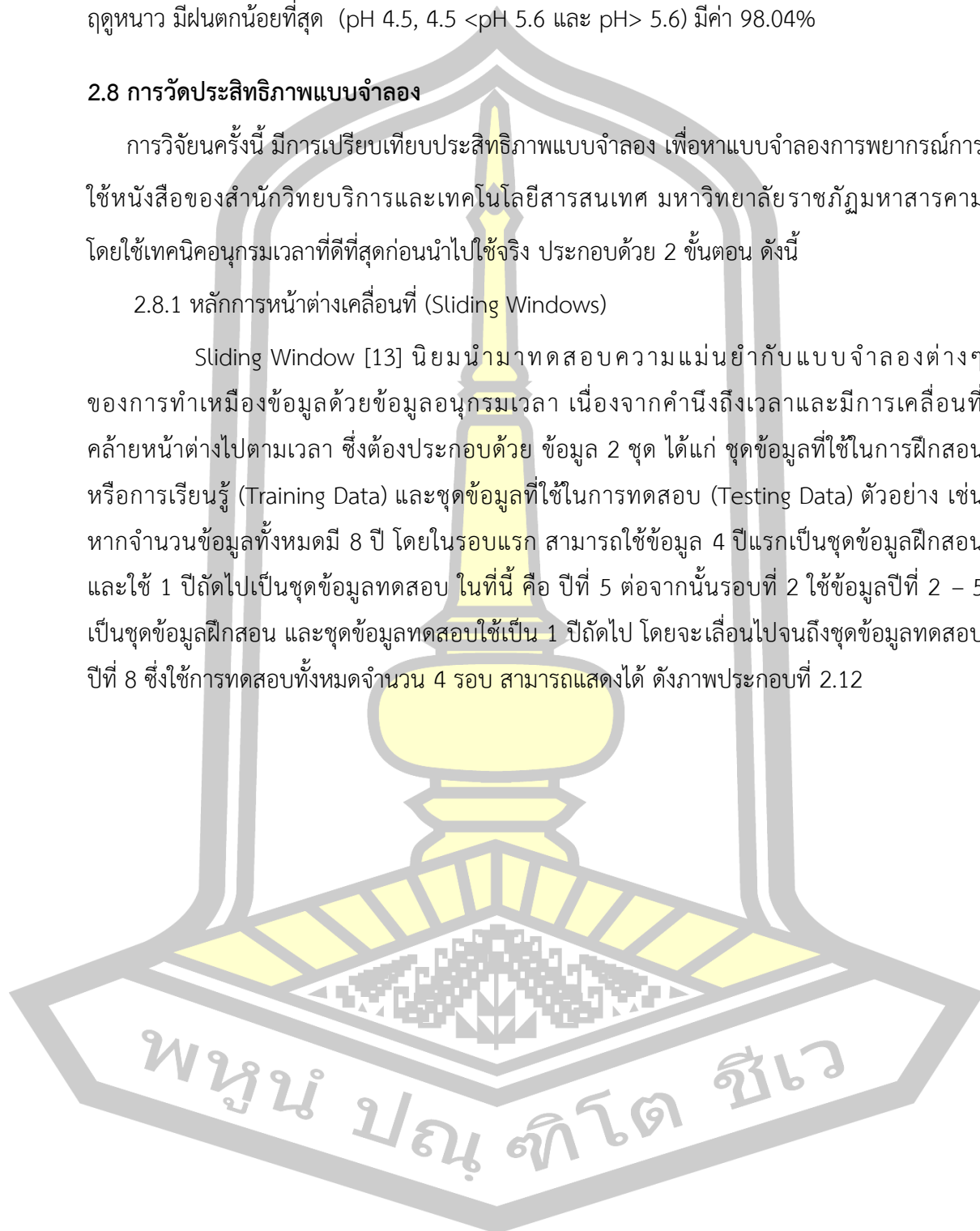
ตั้งแต่ 0 ถึง 84.38% ในฤดูร้อนมีฝนตกมากที่สุด และในฤดูฝนตามด้วยฤดูใบไม้ผลิและฤดูใบไม้ร่วง ฤดูหนาว มีฝนตกน้อยที่สุด (pH 4.5,  $4.5 < \text{pH} < 5.6$  และ  $\text{pH} > 5.6$ ) มีค่า 98.04%

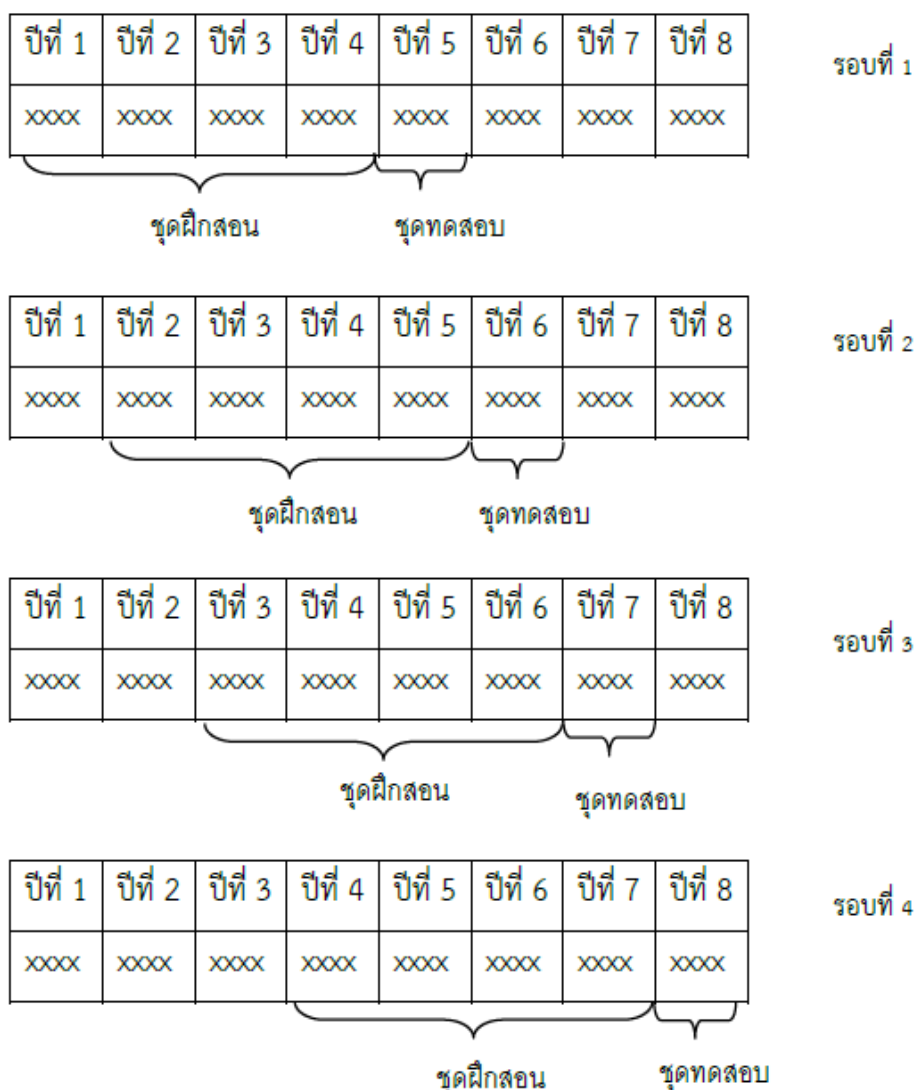
## 2.8 การวัดประสิทธิภาพแบบจำลอง

การวิจัยครั้งนี้ มีการเปรียบเทียบประสิทธิภาพแบบจำลอง เพื่อหาแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม โดยใช้เทคนิคอนุกรมเวลาที่ดีที่สุดก่อนนำไปใช้จริง ประกอบด้วย 2 ขั้นตอน ดังนี้

### 2.8.1 หลักการหน้าต่างเคลื่อนที่ (Sliding Windows)

Sliding Window [13] นิยมนำมาทดสอบความแม่นยำกับแบบจำลองต่างๆ ของการทำเหมืองข้อมูลด้วยข้อมูลอนุกรมเวลา เนื่องจากคำนึงถึงเวลาและมีการเคลื่อนที่คล้ายหน้าต่างไปตามเวลา ซึ่งต้องประกอบด้วย ข้อมูล 2 ชุด ได้แก่ ชุดข้อมูลที่ใช้ในการฝึกสอนหรือการเรียนรู้ (Training Data) และชุดข้อมูลที่ใช้ในการทดสอบ (Testing Data) ตัวอย่าง เช่น หากจำนวนข้อมูลทั้งหมดมี 8 ปี โดยในรอบแรก สามารถใช้ข้อมูล 4 ปีแรกเป็นชุดข้อมูลฝึกสอน และใช้ 1 ปีถัดไปเป็นชุดข้อมูลทดสอบ ในที่นี้ คือ ปีที่ 5 ต่อจากนั้นรอบที่ 2 ใช้ข้อมูลปีที่ 2 - 5 เป็นชุดข้อมูลฝึกสอน และชุดข้อมูลทดสอบใช้เป็น 1 ปีถัดไป โดยจะเลื่อนไปจนถึงชุดข้อมูลทดสอบปีที่ 8 ซึ่งใช้การทดสอบทั้งหมดจำนวน 4 รอบ สามารถแสดงได้ ดังภาพประกอบที่ 2.12





ภาพประกอบที่ 2.12 แสดงการเคลื่อนที่คล้ายหน้าต่างไปตามเวลาของ Sliding Window

Kapoor and Bedi [39] ได้พยากรณ์สภาวะอากาศในอนาคต ที่ต้องใช้ความผันแปรของสภาวะต่างๆ ในรอบหลายปีที่ผ่านมา ความน่าจะเป็นที่สภาพอากาศของวันที่พิจารณาจะตรงกับวันเดียวกันในปีที่แล้วมีน้อยมาก แต่ความน่าจะเป็นที่การจับคู่ภายในช่วงของสัปดาห์ถัดไปของปีก่อนสูงมาก ดังนั้น การ Sliding Windows ถูกเลือกให้มีขนาดเท่ากับหนึ่งสัปดาห์ทุกสัปดาห์ของหน้าต่างบานเลื่อนจะถูกจับคู่กับสัปดาห์ของปีปัจจุบัน การพิจารณาหน้าต่างที่จับคู่กันดีที่สุดจะทำเพื่อเข้าร่วมในกระบวนการพยากรณ์สภาพอากาศทำตามขั้นตอนวิธีการเลื่อนหน้าต่าง ผลการคำนวณเป็นรายเดือนจะคำนวณเป็นเวลาสามปีเพื่อตรวจสอบความถูกต้อง

ผลของวิธีการ ชี้ให้เห็นว่าวิธีการที่ใช้ในการทำนายสภาพอากาศค่อนข้างมีประสิทธิภาพ และมีความแม่นยำเฉลี่ยถึง 92.2%

## 2.8.2 การวัดค่าประสิทธิภาพของแบบจำลอง

งานวิจัยนี้ผู้วิจัยใช้การวัดค่าประสิทธิภาพของแบบจำลองด้วยการหาค่าทางสถิติ ซึ่งประกอบด้วย Mean Absolute Error และ Root Mean Square Error เนื่องจากนิยมใช้กันมากที่สุด โดยมีการแสดงผลเป็นอัตราของความคลาดเคลื่อนทำให้เข้าใจได้ ดังนี้

2.8.2.1 ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Error: MAE) [13] เป็นการวัดค่าความแตกต่างระหว่างค่าที่ได้จากการพยากรณ์ของแบบจำลองกับค่าจริง หากค่าที่ได้จากการพยากรณ์ของแบบจำลองนั้นใกล้เคียงกับค่าจริงมากผลที่ได้จะยิ่งมีค่าน้อย และหากมีค่าเท่ากับ 0 แล้วนั้น หมายความว่า ไม่มีความคลาดเคลื่อนในแบบจำลอง โดยการคำนวณ MAE ดังสมการที่ 2.2

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i| \quad (2.2)$$

2.8.2.2 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) [13] คือ ค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน (RMSD) เป็นการวัดทางสถิติของปริมาณที่มีการเปลี่ยนแปลงอยู่ตลอดเวลา ใช้คำนวณหาอนุกรมของค่าใดๆ ที่แปรผันต่อเนื่อง ซึ่งการวัดความคลาดเคลื่อนจากค่าที่พยากรณ์จากแบบจำลองกับค่าจริงที่เกิดขึ้นนี้ หากมีค่าน้อยแสดงว่าการพยากรณ์ของแบบจำลองนั้นใกล้เคียงกับค่าจริง และหากค่านี้เท่ากับ 0 แล้วนั้น หมายความว่า ไม่มีความคลาดเคลื่อนในแบบจำลอง โดยการคำนวณ RMSE ดังสมการที่ 2.3

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2} \quad (2.3)$$

พหุ ประถมศึกษา

## 2.9 งานวิจัยที่เกี่ยวข้อง

Igor Aizenberg et al. [2] ได้นำเอาเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย (Multi-Layer Perceptron Regression) มาใช้ในสร้างแบบจำลองการถดถอยของอนุกรมเวลาเพื่อการพยากรณ์การผลิตน้ำมันในแต่ละวัน โดยใช้โครงข่ายประสาทเทียมแบบหลายชั้นและมีชุดข้อมูลเกี่ยวกับเซลล์ประสาทหลายค่า ได้ประเมินวิธีการที่เสนอโดยใช้ชุดข้อมูลจริง ในการอธิบายลักษณะการทำงานแบบไดนามิกส์ของสินทรัพย์ในบ่อน้ำมัน ซึ่งตั้งอยู่ในบริเวณหนองน้ำชายฝั่งของอ่าวเม็กซิโก โดยแสดงให้เห็นว่าเทคนิคนี้สามารถนำไปประยุกต์ใช้กับตัวแปรเดียวและหลายตัวแปรได้ในแบบชั้นเดียวและหลายชั้นได้ ในการพยากรณ์การเปลี่ยนแปลงของอ่างเก็บน้ำมันล่องหนานี้ ใช้การสร้างแบบจำลองเพื่อการพยากรณ์ โดยแบ่งออกเป็นการถดถอยและการจำแนก ซึ่งลักษณะเด่นของวิธีการที่นำเสนอในเอกสารฉบับนี้ คือ การพยากรณ์ตามเทคนิคนี้ รวมคุณสมบัติทั้งสองวิธีไว้

ผลการวิจัย พบว่า ทั้งสองแบบสามารถใช้งานและบรรลุตามเป้าหมายได้อย่างมีประสิทธิภาพ โดยมีผลการพยากรณ์ที่แม่นยำในระยะเวลาที่ได้ดีกว่าที่คาดไว้และช่วยให้การพยากรณ์ระยะยาว

G. Peter Zhang and Min Qi b [22] ได้พยากรณ์รูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาลด้วยเทคนิคโครงข่ายประสาทเทียมแบบถดถอยที่ได้ผลลัพธ์ที่หลากหลาย ดังนั้น จึงได้ทำการตรวจสอบปัญหาเกี่ยวกับวิธีการสร้างชุดข้อมูลอนุกรมเวลาอย่างมีประสิทธิภาพ ตามรูปแบบการเปลี่ยนแปลงอนุกรมเวลาตามฤดูกาลและแนวโน้มรูปแบบ พบว่า การนำข้อมูลดิบมาใช้ในการพยากรณ์ลดข้อผิดพลาดได้อย่างมากและเป็นวิธีการประมวลผลข้อมูลที่มีประสิทธิภาพมากที่สุด

Fan Zhang et al. [3] ได้ศึกษาและมีการพัฒนาแบบจำลองด้วยเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (Support Vector Regression : SVR) มาใช้ในสร้างแบบจำลองการถดถอยของอนุกรมเวลาที่มีการถ่วงน้ำหนัก เพื่อพยากรณ์ถึงการใช้พลังงานในการสร้างอาคาร เนื่องจากเป็นสิ่งสำคัญสำหรับการดำเนินงานและการจัดการอาคารที่มีประสิทธิภาพ โดยมีการเลือกพารามิเตอร์ของตัวประมวลผลวิวัฒนาการ (DE) เพื่อเพิ่มพารามิเตอร์ของรูปแบบและน้ำหนักให้กับ SVR ทั้งสองรุ่น ได้แก่ eps-SVR และ nu-SVR ทำให้มีประสิทธิภาพสำหรับงานนี้ โดยการพยากรณ์มีความถูกต้องสูงมาก เนื่องจากชุดข้อมูลมีความซับซ้อนและไม่เป็นเชิงเส้น ซึ่งได้นำแบบจำลองไปใช้กับชุดข้อมูลสองชุดสำหรับการสร้างแบบเดียวกัน ได้แก่ ปริมาณการใช้พลังงานรายครึ่งชั่วโมง เป็นระยะเวลาสิบวัน และปริมาณการใช้พลังงานรายวันสำหรับการใช้พลังงานต่อปีเป็นเวลา 1 ปี พบว่า แบบจำลองที่เสนอสามารถคาดการณ์การใช้พลังงานสำหรับทั้งชุดข้อมูลที่มีความถูกต้องดี



โดยไม่ต้องเปลี่ยนโครงสร้างแบบจำลอง ตัวอย่าง ข้อมูลการใช้พลังงานจากการสร้างอาคารในสิงคโปร์ เพื่ออธิบายถึงประสิทธิภาพของรูปแบบที่เสนอ ทั้งแบบครึ่งชั่วโมงและข้อมูลการใช้พลังงานไฟฟ้า แบบรายวันสำหรับอาคารเดียวกัน ด้วยการใช้วิธีถ่วงน้ำหนักที่อัลกอริธึม DE ดังนี้ สำหรับการวิเคราะห์การใช้พลังงานรายวัน อัลกอริธึม DE จะกำหนดน้ำหนักที่สูงขึ้นให้กับ eps-SVR และ สำหรับการวิเคราะห์การใช้พลังงานครึ่งชั่วโมง อัลกอริธึม DE จะกำหนดน้ำหนักที่สูงขึ้นให้กับ nu-SVR ดังนั้น รูปแบบเดียวสามารถคาดการณ์การใช้พลังงานได้ทุกวันและทุกครึ่งชั่วโมง โดยไม่มีการแทรกแซงใด ๆ

โดยผลการศึกษาแสดงให้เห็นว่า ข้อผิดพลาดร้อยละสัมบูรณ์เฉลี่ย (MAPE) สำหรับข้อมูลการใช้พลังงานรายวัน เป็น 5.843 และสำหรับการใช้พลังงานครึ่งชั่วโมงเป็น 3.767 ตามลำดับ และเมื่อการเปรียบเทียบกับพยากรณ์การใช้พลังงานรูปแบบ อื่นๆ ซึ่งมักใช้สำหรับการเพิ่มประสิทธิภาพ เช่น GA และ PSO จะแสดงให้เห็นว่ารูปแบบที่นำเสนอมีความแม่นยำสูงกว่าวิธีอื่นๆ และมี MAPE ที่ลดลงสำหรับทั้งชุดข้อมูลรายวันและชุดข้อมูลรายครึ่งชั่วโมง

Ceperic Baric A [25] ได้ใช้เทคนิคการพยากรณ์ระยะสั้น (STLF) และเทคนิค Support Vector Machine (SVR) โดยการสร้างแบบจำลองและการใช้อัลกอริธึมสำหรับการเลือกคุณลักษณะ ในการพยากรณ์นั้น ซึ่งได้ใช้เทคนิค SVR hyper - parameters เพิ่มประสิทธิภาพในการทำงาน และใช้เทคนิคการพยากรณ์ระยะสั้น (STLF) ในการพยากรณ์และเปรียบเทียบอัลกอริธึม ผลปรากฏว่า ได้ค่าความถูกต้องดีขึ้น

Edwin Baidoo and Jennifer L. Priestley [4] ได้นำเอาเทคนิคการวิเคราะห์การถดถอย โลจิสติก (logistic regression analysis) มาใช้ในการสร้างแบบจำลองการถดถอยของอนุกรมเวลา เพื่อวิเคราะห์อัตราความถูกต้องในการพยากรณ์การผิตนัดชำระหนี้ของลูกค้าและตรวจสอบดัชนี ประสิทธิภาพการทำการกำไรสูงสุด (EMP) ของบริษัท บนข้อสมมติฐานทางธุรกิจของตลาดสินเชื่อ ซึ่งมีแนวคิดเกี่ยวกับความล้มเหลวทางธุรกิจและการผิตนัดชำระหนี้ของลูกค้าผ่านทาง การใช้อัตราส่วนทางการเงิน แม้ว่าการวิจัยก่อนหน้านี้ได้มุ่งเน้นไปที่เมตริกการประเมินผล เช่น ดัชนี AUC และ Gini แต่การวัดใหม่ที่ประยุกต์ใช้กับสร้างแบบจำลอง Logistic และ Time Series นี้ มีการตีความง่ายขึ้นสำหรับผู้บริหารและผู้มีอำนาจในการตัดสินใจต่างๆ

พบว่า การวิเคราะห์อนุกรมเวลาด้วยเทคนิคการวิเคราะห์การถดถอยมีประสิทธิภาพดีกว่า ซึ่งมีความถูกต้อง เป็น 92.31%

Mohamed Addel-Aty et al. [28] ได้ทำการศึกษาวิจัยเกี่ยวกับการพยากรณ์ เพื่อลดความผิดพลาดบนทางด่วนในระบบการจัดการจราจรชั้นสูงและระบบสารสนเทศ โดยใช้ตัวแปรการจราจรตามเวลาจริง ซึ่งได้ใช้เทคนิคการวิเคราะห์การถดถอยโลจิสติกในการพยากรณ์ เพื่อลดความผิดพลาดที่เกิดขึ้นจริงบนทางด่วนในระบบการจัดการจราจร ที่เกิดความผิดพลาดเป็นอย่างมากบนทางด่วน ดังนั้น จึงได้นำความผิดพลาดที่เกิดขึ้นมาใช้ในการคำนวณอัตราส่วน log-odds ซึ่งได้ใช้เกณฑ์สำหรับการกำหนดอัตราส่วนในการคำนวณอัตราส่วน log-odds อยู่ที่ 1.0 ผลปรากฏว่า ได้อัตราส่วน log-odds มากที่สุด 69%

Geoffrey and Kelvin [5] ได้นำเอาเทคนิคการทำเหมืองข้อมูล ต้นไม้ตัดสินใจแบบถดถอย (decision tree Regression) และโครงข่ายประสาทเทียม มาใช้ในสร้างแบบจำลองการถดถอยของอนุกรมเวลา เพื่อการเปรียบเทียบการคาดการณ์การใช้พลังงานไฟฟ้าในฮ่องกงกับการวิเคราะห์การถดถอยแบบเดิม แล้วนำมาประเมินประสิทธิภาพของแบบจำลองที่แตกต่างกันและเลือกรูปแบบที่เหมาะสมที่สุดสำหรับการคาดการณ์ในอนาคต เมื่อเปรียบเทียบความถูกต้องในการคาดการณ์การใช้พลังงานไฟฟ้า แบบจำลองต้นไม้ตัดสินใจและแบบจำลองโครงข่ายประสาทเทียม มีประสิทธิภาพดีกว่ารุ่นอื่นๆ ในช่วงฤดูร้อนและฤดูหนาวตามลำดับ

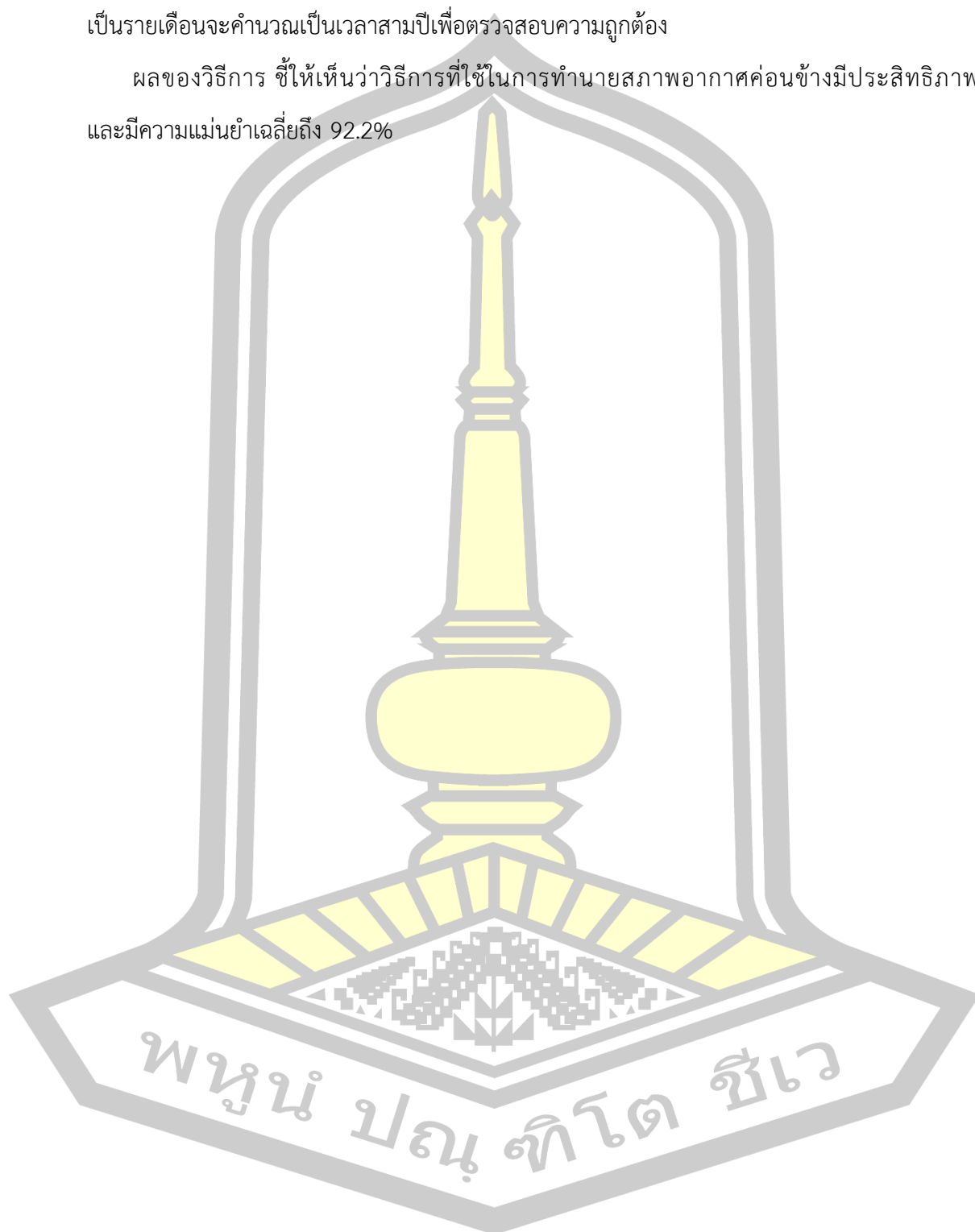
Xiuying Zhang et al. [38] ได้ทำการศึกษาวิจัยเกี่ยวกับการพยากรณ์การเกิดฝนกรด โดยใช้ตำแหน่งลักษณะภูมิศาสตร์และลักษณะภูมิประเทศในการวิเคราะห์ปัจจัยทางด้านอุตุนิยมวิทยา ความหนาแน่นของชั้นบรรยากาศ SO<sub>2</sub> และ NO<sub>2</sub> ซึ่งในการงานวิจัยใช้เทคนิคต้นไม้ตัดสินใจแบบถดถอยในการพยากรณ์ ผลการวิจัย พบว่า ฝนกรดที่เกิดขึ้นใน 17 สถานี ในกลุ่ม 81 สถานี พยากรณ์ในจังหวัดเหลียวหนิงนั้น มีฝนกรดตั้งแต่ 0 ถึง 84.38% ในฤดูร้อนมีฝนตกมากที่สุด และในฤดูฝนตามด้วยฤดูใบไม้ผลิและฤดูใบไม้ร่วงฤดูหนาวมีฝนตกน้อยที่สุด (pH 4.5, 4.5 < pH 5.6 และ pH > 5.6) มีค่า 98.04%

Kapoor and Bedi [39] ได้พยากรณ์สภาวะอากาศในอนาคต ที่ต้องใช้ความผันแปรของสภาวะต่างๆ ในรอบหลายปีที่ผ่านมา ความน่าจะเป็นที่สภาพอากาศของวันที่พิจารณาจะตรงกับวันเดียวกันในปีที่แล้วมีน้อยมาก แต่ความน่าจะเป็นที่การจับคู่ภายในช่วงของสัปดาห์ถัดไปของปีก่อนสูงมาก

ดังนั้น การ sliding windows ถูกเลือกให้มีขนาดเท่ากับหนึ่งสัปดาห์ทุกสัปดาห์ของหน้าต่างบานเลื่อนจะถูกจับคู่กับสัปดาห์ของปีปัจจุบัน การพิจารณาหน้าต่างที่จับคู่กันดีที่สุดจะทำเพื่อเข้าร่วม

ในกระบวนการพยากรณ์สภาพอากาศทำตามขั้นตอนวิธีการเลื่อนหน้าต่าง ผลการคำนวณเป็นรายเดือนจะคำนวณเป็นเวลาสามปีเพื่อตรวจสอบความถูกต้อง

ผลของวิธีการ ชี้ให้เห็นว่าวิธีการที่ใช้ในการทำนายสภาพอากาศค่อนข้างมีประสิทธิภาพ และมีความแม่นยำเฉลี่ยถึง 92.2%



### บทที่ 3

#### วิธีการดำเนินการวิจัย

งานวิจัยนี้ได้ดำเนินการศึกษาค้นคว้าและสร้างแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ตามขั้นตอนการวิเคราะห์อนุกรมเวลา [18] 4 ขั้นตอน ประกอบด้วย การเตรียมข้อมูลจากระบบห้องสมุดอัตโนมัติ VTLS การทำกระบวนการก่อนการสร้างแบบจำลอง ด้วยการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลการใช้หนังสือแต่ละหมวดแบบรายวัน การสร้างแบบจำลองได้นำเอาโปรแกรม WEKA เวอร์ชัน 3.9.2 มาใช้สร้างแบบจำลองด้วยเทคนิคแบบต่างๆ และการวัดประสิทธิภาพของแบบจำลองดังภาพประกอบที่ 3.1



ภาพประกอบที่ 3.1 ขั้นตอนการวิเคราะห์อนุกรมเวลา

#### 3.1 การเตรียมข้อมูล

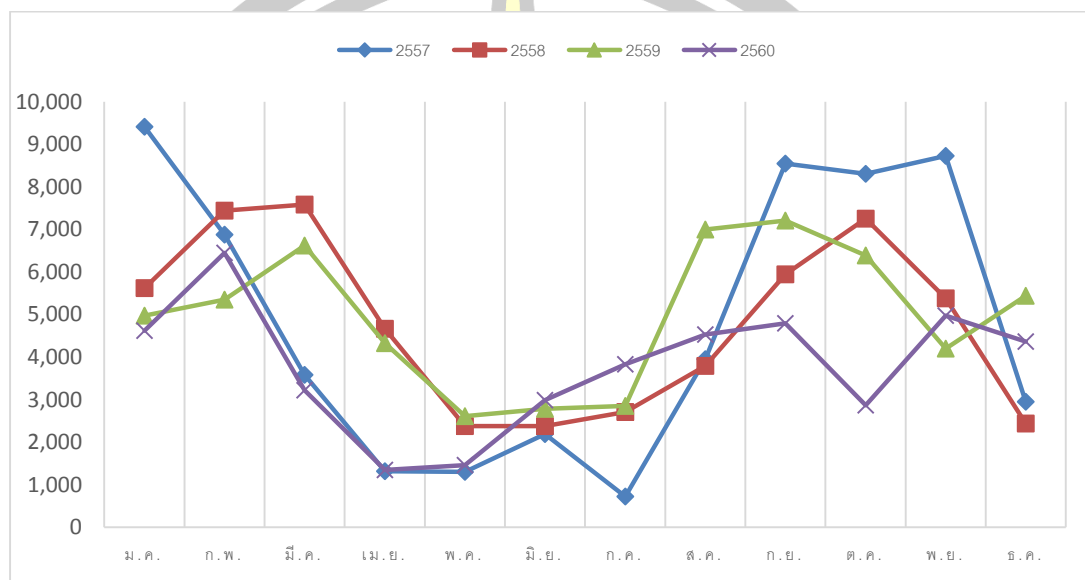
การวิจัยครั้งนี้ ได้ข้อมูลจากระบบห้องสมุดอัตโนมัติ VTLS ด้วยความอนุเคราะห์ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ซึ่งมีการจัดเก็บข้อมูลการใช้หนังสือ ของแต่ละวัน จาก 10 หมวดหนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ เริ่มจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง ธันวาคม พ.ศ. 2560 ดังตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูลการใช้หนังสือเป็นรายวัน

| ปี-เดือน-วัน | เลขเรียก          | ชื่อ   |
|--------------|-------------------|--|
| 14-01-02     | 531.1 ส47ก        | กลศาสตร์เบื้องต้น / สุรสิงห์ ไชยคุณ                |
| 14-01-02     | 531.1 ส47ก        | กลศาสตร์เบื้องต้น / สุรสิงห์ ไชยคุณ                |
| 14-01-02     | 531.1 ส47ก        | กลศาสตร์เบื้องต้น / สุรสิงห์ ไชยคุณ                |
| 14-01-02     | 379.111 ส42ก 2554 | การพัฒนาคุณภาพตามมาตรฐานคุณภาพการศึกษา             |
| 14-01-02     | 507 ม164ก 2549    | การพัฒนาแบบฝึกการทำโครงงานวิทยาศาสตร์              |
| 14-01-02     | 375 จ372ผ 2554    | ผลการจัดการเรียนรู้โดยใช้เทคนิค KWDL               |
| 14-01-02     | 530 ส424ฟ 2550    | ฟิสิกส์ 1 ระดับมหาวิทยาลัย / สุชาติ สุภาพ          |
| 14-01-02     | 530 ส424ฟ 2550    | ฟิสิกส์ 1 ระดับมหาวิทยาลัย เล่ม 2 / สุชาติ สุภาพ   |
| 14-01-02     | 530 ส424ฟ 2550    | ฟิสิกส์ 1 ระดับมหาวิทยาลัย เล่ม 2 / สุชาติ สุภาพ   |
| 14-01-02     | 531 ส25ก          | กลศาสตร์ / สานิตย์ ชัยพัฒน์.                       |
| 14-01-02     | 510 บ276ป 2551    | ปรัชญาคณิตศาสตร์ : Philosophy of Mathematics       |
| 14-01-02     | 531.1 ส47ก        | กลศาสตร์เบื้องต้น / สุรสิงห์ ไชยคุณ                |
| 14-01-02     | 515 ส4111ค 2547   | แคลคูลัส 2 ฉบับเสริมประสบการณ์ / สุกัญญา สนิท      |
| 14-01-02     | 515 ส4111ค 2546   | แคลคูลัส 1 / สุกัญญา สนิทวงศ์                      |
| 14-01-02     | 531 ส25ก          | กลศาสตร์ / สานิตย์ ชัยพัฒน์.                       |
| 14-01-02     | 531.1 ส47ก        | กลศาสตร์เบื้องต้น / สุรสิงห์ ไชยคุณ                |
| 14-01-02     | 530 ย112ฟ 2547    | ฟิสิกส์ระดับอุดมศึกษา เล่ม 1 = University physics  |
| 14-01-02     | 530 ย112ฟ 2547    | ฟิสิกส์ระดับอุดมศึกษา เล่ม 1 = University physics  |
| 14-01-02     | 895.9109 ว177     | วรรณกรรมทักษิณ : วรรณกรรมปริทัศน์ / กลิ่น คงเหมือน |
| 14-01-02     | 745.925 ศ578ศ     | ศิลปะอลังการงานใบตอง / เศรษฐมนตร์ กาญจนกุล.        |
| 14-01-02     | 745.92 อ462บ 2550 | บายศรีปากชามประกอบพิธีกรรม / อูมาวดี ทรัพย์สิน.    |
| 14-01-02     | 745.5 ม           | งานใบตอง   |
| 14-01-02     | 745.5 ม           | งานใบตอง   |
| 14-01-02     | 634.4602 ก15ค     | คู่มือมะขามหวาน                                    |
| 14-01-02     | 613.71 ย462ส 2553 | สภาพและปัญหาการเรียนการสอน ศึกษาศาสตร์และพลศึกษา   |

จากตารางที่ 3.1 แสดงถึงตัวอย่างข้อมูลการใช้หนังสือที่ได้จากระบบห้องสมุดอัตโนมัติ VTLS โดยข้อมูลจะถูกบันทึกตามเวลาที่หนังสือได้ถูกใช้ ตามรูปแบบของอนุกรมเวลา ซึ่งมีการจัดเก็บข้อมูลการใช้

หนังสือเป็นรายวัน แต่ยังไม่ได้แยกตามหมวดหนังสือ ด้วยการจัดเก็บข้อมูลการใช้หนังสือของแต่ละวัน เริ่มจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง ธันวาคม พ.ศ. 2560 ซึ่งแสดงการเคลื่อนไหวของข้อมูลการใช้หนังสือของแต่ละวันสามารถสรุปเป็นรายเดือนได้ ดังภาพประกอบที่ 3.2



ภาพประกอบที่ 3.2 กราฟแสดงปริมาณการใช้หนังสือแบบรายเดือน พ.ศ. 2557 - 2560

จากภาพประกอบที่ 3.2 พบว่า กราฟได้แสดงข้อมูลการใช้หนังสือที่มีการจัดเก็บเรียงตามเวลาที่ข้อมูลได้ถูกบันทึกจากระบบห้องสมุดอัตโนมัติ VTLS นั้น ปี พ.ศ. 2557 มีแนวโน้มปริมาณการใช้หนังสือสูงในเดือน มกราคม และ มิถุนายน แล้วขึ้นอีกในช่วงเดือน สิงหาคม – พฤศจิกายน และจะมีแนวโน้มปริมาณการใช้ต่ำในเดือน กุมภาพันธ์ – พฤษภาคม และในเดือน กรกฎาคม แล้วจะลงอีกครั้งในช่วงเดือนธันวาคม ปี พ.ศ. 2558 - 2559 มีแนวโน้มปริมาณการใช้หนังสือสูงในเดือน มกราคม – มีนาคม และ สิงหาคม – พฤศจิกายน และจะมีแนวโน้มปริมาณการใช้ต่ำในเดือน เมษายน - กรกฎาคม และ ธันวาคม และปี พ.ศ. 2560 มีแนวโน้มปริมาณการใช้หนังสือสูงในเดือน กุมภาพันธ์ และ มิถุนายน - กันยายน แล้วจะขึ้นอีกครั้งในเดือนพฤศจิกายน และจะมีแนวโน้มปริมาณการใช้ต่ำในเดือน เมษายน - พฤษภาคม และในเดือน ตุลาคม แล้วจะลงอีกครั้งในช่วงเดือน ธันวาคม - มกราคม จะเห็นว่า ปริมาณการใช้หนังสือมีการเคลื่อนไหวของข้อมูลที่เกิดขึ้นโดยส่วนใหญ่จะเกิดขึ้นซ้ำรอบกันในแต่ละปี มีเพียงบางปีที่มีรายละเอียดแตกต่างกันไปเล็กน้อย ซึ่งตรงกับรูปแบบการ

เปลี่ยนแปลงอนุกรมเวลาตามฤดูกาล (Seasonal Variation) เมื่อนำข้อมูลการใช้หนังสือมาวิเคราะห์ตามหมวดหมู่ในระบบ โดยค่าทางสถิติจะได้ดังตารางที่ 3.2

ตารางที่ 3.2 ค่าทางสถิติตามเลขหมวดหมู่ในระบบทศนิยมของดิวอี้

| หมวด  | ค่าต่ำสุด | ค่าสูงสุด | ค่าเฉลี่ย | ค่าเบี่ยงเบนมาตรฐาน |
|-------|-----------|-----------|-----------|---------------------|
| Call0 | 0         | 151       | 10.21     | 10.03               |
| Call1 | 0         | 75        | 8.78      | 8.85                |
| Call2 | 0         | 64        | 3.22      | 5.25                |
| Call3 | 0         | 219       | 51.81     | 34.10               |
| Call4 | 0         | 144       | 13.41     | 12.42               |
| Call5 | 0         | 123       | 21.98     | 18.76               |
| Call6 | 0         | 238       | 28.99     | 30.88               |
| Call7 | 0         | 81        | 6.65      | 8.84                |
| Call8 | 0         | 93        | 8.90      | 9.92                |
| Call9 | 0         | 75        | 7.77      | 9.99                |

จากตารางที่ 3.2 ได้แสดงค่าทางสถิติจากข้อมูลจริงในแต่ละวันตามระยะเวลาที่กำหนด โดยแยกตามเลขหมวดหมู่ในระบบทศนิยมของดิวอี้ ประกอบด้วย ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของปริมาณการใช้หนังสือ พบว่า มีการใช้หนังสือหมวด 300 (ค่าเฉลี่ย 51.81) มากที่สุด รองลงมา ได้แก่ หนังสือหมวด 600 (ค่าเฉลี่ย 28.99) หนังสือหมวด 500 (ค่าเฉลี่ย 21.98) ตามลำดับ และหนังสือหมวด 200 (ค่าเฉลี่ย 3.22) มีการใช้น้อยที่สุด

### 3.2 การทำกระบวนการก่อนการสร้างแบบจำลอง

การทำกระบวนการก่อนการสร้างแบบจำลองโดยการแปลงข้อมูล (Data transformation) จากระบบห้องสมุดอัตโนมัติ VTLS โดยแบ่งข้อมูลแต่ละหมวดหนังสือตามเนื้อหาวิชา ดังตารางที่ 3.3

ตารางที่ 3.3 Class เลขหมวดหมู่ตัวเลขแทนชนิดของหนังสือ

| หมวด  | เนื้อหาวิชา                             |
|-------|---|
| Call0 | เบ็ดเตล็ด ความรู้ทั่วไป บรรณารักษศาสตร์ |
| Call1 | ปรัชญา จิตวิทยา                         |
| Call2 | ศาสนา                                   |
| Call3 | สังคมศาสตร์                             |
| Call4 | ภาษาศาสตร์                              |
| Call5 | วิทยาศาสตร์ คณิตศาสตร์                  |
| Call6 | เทคโนโลยี หรือวิทยาศาสตร์ประยุกต์       |
| Call7 | ศิลปกรรม และนันทนาการ                   |
| Call8 | วรรณคดี                                 |
| Call9 | ภูมิศาสตร์และประวัติศาสตร์              |

จากตารางที่ 3.3 แสดงถึงเลขเรียกหมวดหมู่แทนชนิดของหนังสือ เพื่อแบ่งหนังสือเป็นหมวดหมู่ใหญ่ๆ ตามระบบทศนิยมของดิวอี้ โดยรวบรวมข้อมูลจากแต่ละหมวดหนังสือเข้าด้วยกันเป็นแบบรายวัน ดังตารางที่ 3.4

ตารางที่ 3.4 คำอธิบายตัวแปรและชนิดของตัวแปร

| หมวด  | ค่า          | ประเภทตัวแปร |
|-------|--------------|--------------|
| Date  | ปี-เดือน-วัน | Date         |
| Call0 | จำนวนการใช้  | Nominal      |
| Call1 | จำนวนการใช้  | Nominal      |
| Call2 | จำนวนการใช้  | Nominal      |
| Call3 | จำนวนการใช้  | Nominal      |
| Call4 | จำนวนการใช้  | Nominal      |
| Call5 | จำนวนการใช้  | Nominal      |
| Call6 | จำนวนการใช้  | Nominal      |
| Call7 | จำนวนการใช้  | Nominal      |
| Call8 | จำนวนการใช้  | Nominal      |
| Call9 | จำนวนการใช้  | Nominal      |



จากตารางที่ 3.4 ได้อธิบายถึงตัวแปรและชนิดของตัวแปร สำหรับรวบรวมข้อมูลจากแต่ละหมวดหนังสือเข้าด้วยกันเป็นแบบรายวัน เมื่อแปลงข้อมูลแล้ว จะได้ดังตารางที่ 3.5

ตารางที่ 3.5 การแปลงข้อมูลการใช้หนังสือแบบรายวัน

| date       | call0 | call1 | call2 | call3 | call4 | call5 | call6 | call7 | call8 | call9 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2014-01-02 | 11    | 15    | 4     | 38    | 7     | 79    | 58    | 6     | 6     | 0     |
| 2014-01-03 | 11    | 0     | 2     | 36    | 9     | 48    | 54    | 5     | 7     | 5     |
| 2014-01-04 | 26    | 1     | 4     | 84    | 12    | 68    | 75    | 6     | 20    | 0     |
| 2014-01-05 | 37    | 6     | 0     | 70    | 4     | 37    | 46    | 0     | 14    | 0     |
| 2014-01-06 | 2     | 26    | 4     | 58    | 9     | 35    | 25    | 2     | 8     | 7     |
| 2014-01-07 | 25    | 28    | 10    | 129   | 6     | 39    | 36    | 22    | 25    | 5     |
| 2014-01-08 | 10    | 15    | 5     | 74    | 7     | 64    | 94    | 1     | 27    | 7     |
| 2014-01-09 | 5     | 3     | 0     | 42    | 18    | 84    | 77    | 3     | 6     | 11    |
| 2014-01-10 | 24    | 19    | 12    | 60    | 16    | 71    | 82    | 17    | 3     | 14    |
| 2014-01-11 | 19    | 12    | 0     | 71    | 4     | 49    | 67    | 12    | 8     | 14    |
| 2014-01-12 | 43    | 7     | 0     | 99    | 11    | 30    | 65    | 3     | 12    | 1     |
| 2014-01-13 | 25    | 16    | 20    | 57    | 9     | 64    | 99    | 18    | 27    | 22    |
| 2014-01-14 | 10    | 27    | 7     | 113   | 15    | 43    | 39    | 3     | 7     | 39    |
| 2014-01-15 | 17    | 10    | 6     | 80    | 13    | 76    | 87    | 16    | 15    | 39    |
| 2014-01-16 | 12    | 34    | 8     | 55    | 15    | 70    | 148   | 10    | 18    | 26    |
| 2014-01-17 | 6     | 11    | 1     | 68    | 15    | 69    | 132   | 23    | 93    | 19    |
| 2014-01-18 | 9     | 5     | 4     | 110   | 4     | 23    | 64    | 13    | 1     | 3     |
| 2014-01-19 | 9     | 3     | 0     | 59    | 5     | 28    | 98    | 17    | 22    | 4     |
| 2014-01-20 | 20    | 7     | 15    | 44    | 12    | 123   | 190   | 62    | 9     | 10    |
| 2014-01-21 | 14    | 19    | 7     | 104   | 13    | 50    | 132   | 14    | 25    | 8     |
| 2014-01-22 | 22    | 17    | 9     | 81    | 9     | 81    | 129   | 9     | 14    | 8     |
| 2014-01-23 | 16    | 26    | 11    | 71    | 11    | 38    | 156   | 19    | 37    | 14    |
| 2014-01-24 | 12    | 8     | 8     | 89    | 7     | 60    | 104   | 7     | 67    | 17    |
| 2014-01-25 | 15    | 0     | 1     | 71    | 5     | 36    | 47    | 4     | 3     | 2     |

จากตารางที่ 3.5 แสดงการแปลงข้อมูลจากระบบห้องสมุดอัตโนมัติ VTLS โดยรวบรวมข้อมูลจากแต่ละหมวดหนังสือเข้าด้วยกันเป็นแบบรายวัน

### 3.3 การสร้างแบบจำลอง

งานวิจัยนี้ได้นำข้อมูลอนุกรมเวลา โดยการนำข้อมูลการใช้หนังสือที่บันทึกไว้ของแต่ละวันตามลำดับเวลา จาก 10 หมวดหนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มาสร้างแบบจำลองที่มีประสิทธิภาพ เพื่อการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม ด้วยเทคนิคเหมืองข้อมูลต่าง ๆ ดังนี้

3.3.1 เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นแบบถดถอย (Multi-Layer Perceptron Regression) ใช้กับกรณีที่ปัญหาที่มีความซับซ้อน จึงเพิ่มจำนวนโหนดที่มีการคำนวณหรือชั้นแอบแฝงให้กับโครงข่าย เป็นเทคนิคที่มีรูปแบบโครงข่ายประสาทเทียมแบบ Feedforward โดยสามารถวิเคราะห์ข้อมูลการถดถอยเชิงเส้นได้

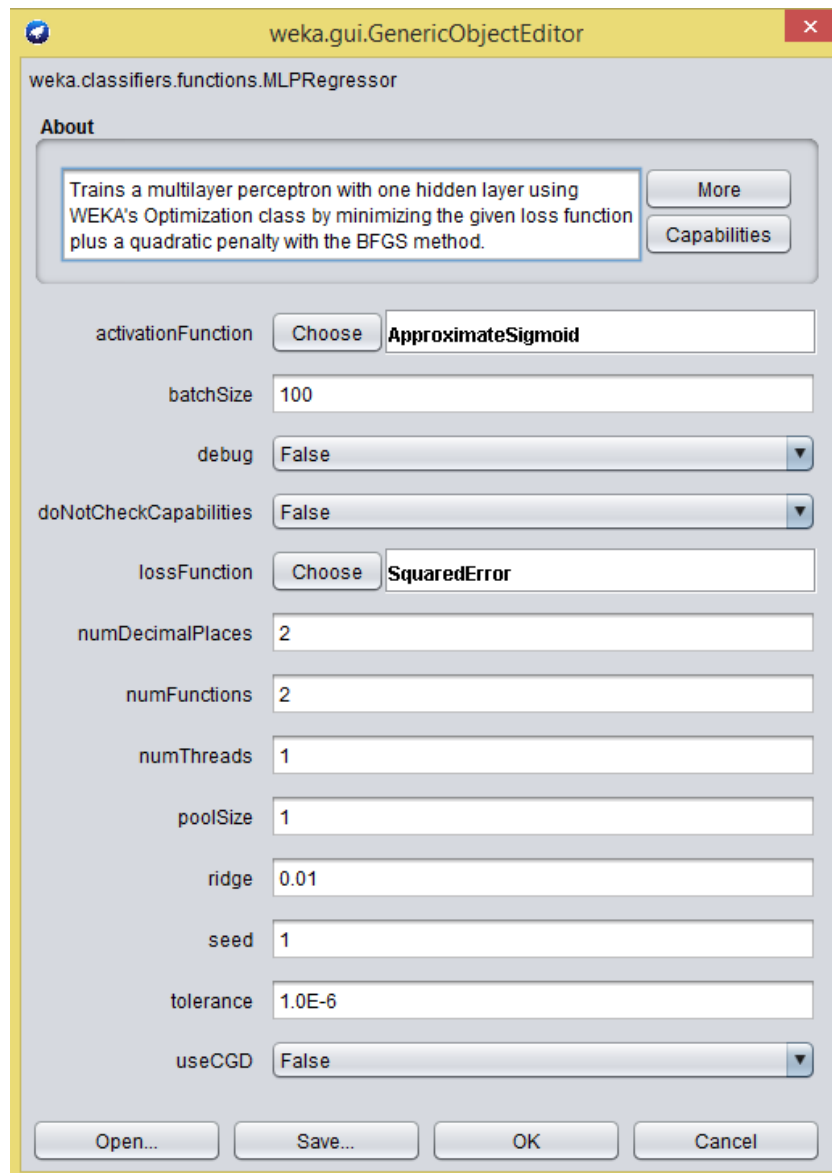
3.3.2 เทคนิคโครงข่ายประสาทเทียมแบบถดถอย (Artificial Neural Network Regression) สามารถพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคตได้อย่างแม่นยำ เป็นการวิเคราะห์และพัฒนาแบบจำลองเพื่อใช้กับการวิเคราะห์ข้อมูลการถดถอยเชิงเส้น

3.3.3 เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (Support Vector Machine for Regression) เป็นการวิเคราะห์ข้อมูลอนุกรมเวลา เพื่อการพยากรณ์ที่เกิดความผิดพลาดต่ำสุด มีการใช้ทฤษฎีการเรียนรู้จากสถิติเป็นพื้นฐานในการจำแนกข้อมูลเช่นเดียวกับโครงข่ายประสาทเทียม แต่มีความแตกต่าง คือ เทคนิคสนับสนุนเส้นสมมุติแบบถดถอยใช้หลักการลดค่าความเสี่ยงเชิงโครงสร้างให้น้อยที่สุด

3.3.4 การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) เป็นเทคนิคถดถอยที่ใช้กันมากในทางสถิติ โดยใช้เป็นพื้นฐานในอนุกรมเวลา เพื่อหาแนวโน้มของข้อมูลด้วยสมการเชิงเส้น

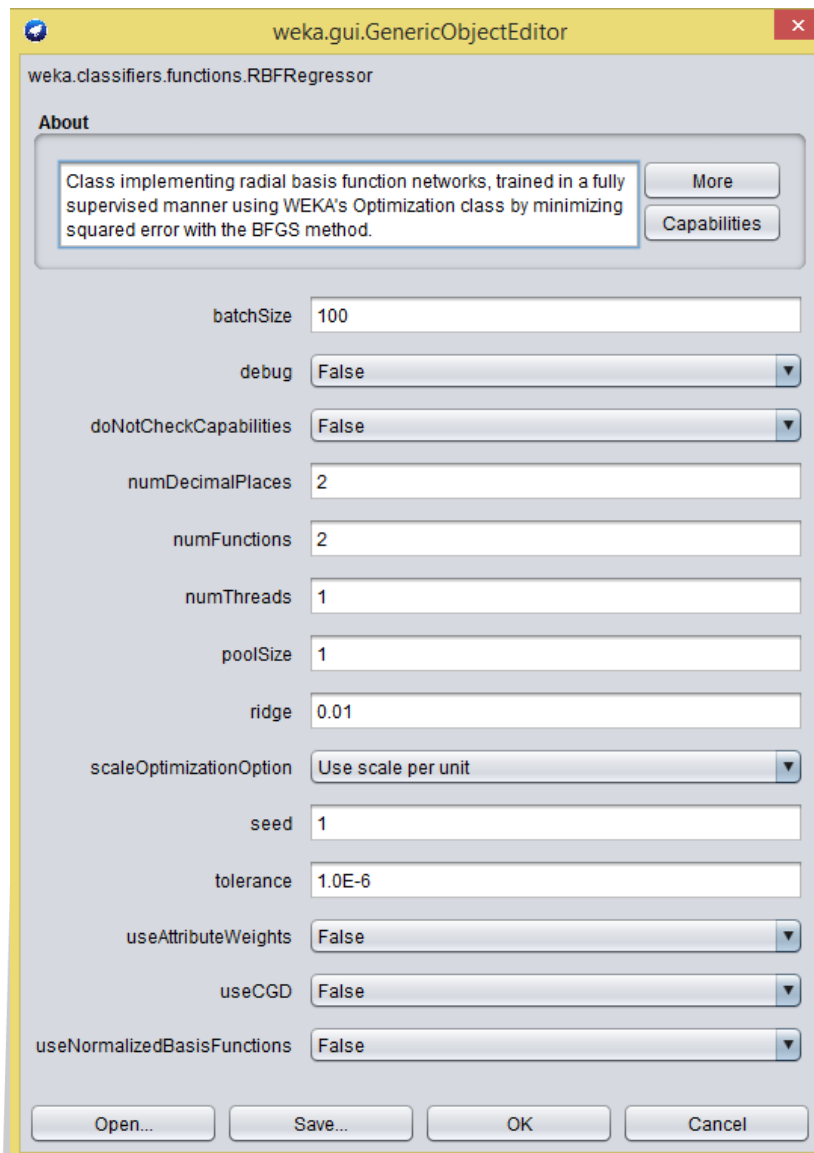
3.3.5 การเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย (Decision Tree Regression) อัลกอริทึมการตัดกิ่งต้นไม้ (Reduced Error Pruning Tree: REPT) เป็นเทคนิคที่นิยมใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลา ได้รับการพัฒนาจากอัลกอริทึมการเรียนรู้ต้นไม้ตัดสินใจ โดยการแก้ปัญหาอันเกิดจากการสร้างต้นไม้ตัดสินใจด้วยการตัดกิ่ง เพื่อให้ความผิดพลาดลดลง ซึ่งอาจมีการแตกกิ่งมากเกินไปทำให้เสียเวลาในการสร้างแบบจำลอง

งานวิจัยนี้จึงได้สร้างแบบจำลอง โดยได้นำเอาโปรแกรม WEKA เวอร์ชัน 3.9.2 มาเป็นเครื่องมือในการสร้างแบบจำลองด้วยเทคนิคแบบต่างๆ และมีค่าพารามิเตอร์ ดังภาพประกอบที่ 3.3 - 3.7



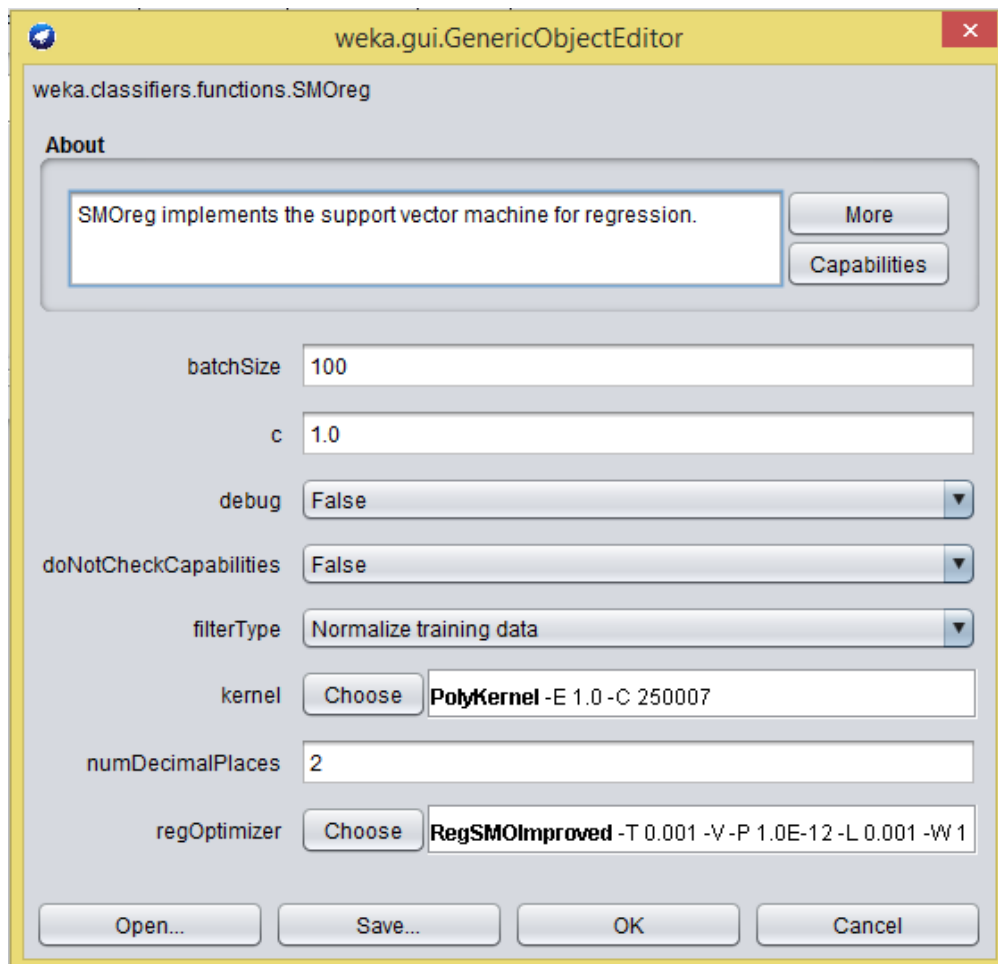
ภาพประกอบที่ 3.3 ค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย

จากภาพประกอบที่ 3.3 แสดงค่าพารามิเตอร์ต่างๆ ของเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย เช่น batchSize เป็นจำนวนตัวอย่างการพยากรณ์ที่ต้องการประมวลผลเบื้องต้น debug เป็นรายงานการแก้ไขข้อบกพร่องของข้อมูล doNotCheckCapabilities เป็นการตั้งค่าความสามารถในการจำแนกประเภทจะไม่ได้รับการตรวจสอบก่อนสร้างตัวจำแนกประเภท numDecimalPlaces เป็นรูปแบบจำนวนทศนิยม ridge คือ ความคุ่มค่าสูงสุดของพารามิเตอร์ seed เป็นผลที่ได้ของผู้ใช้ส่งไปสุ่มจำนวนใหม่ เป็นต้น



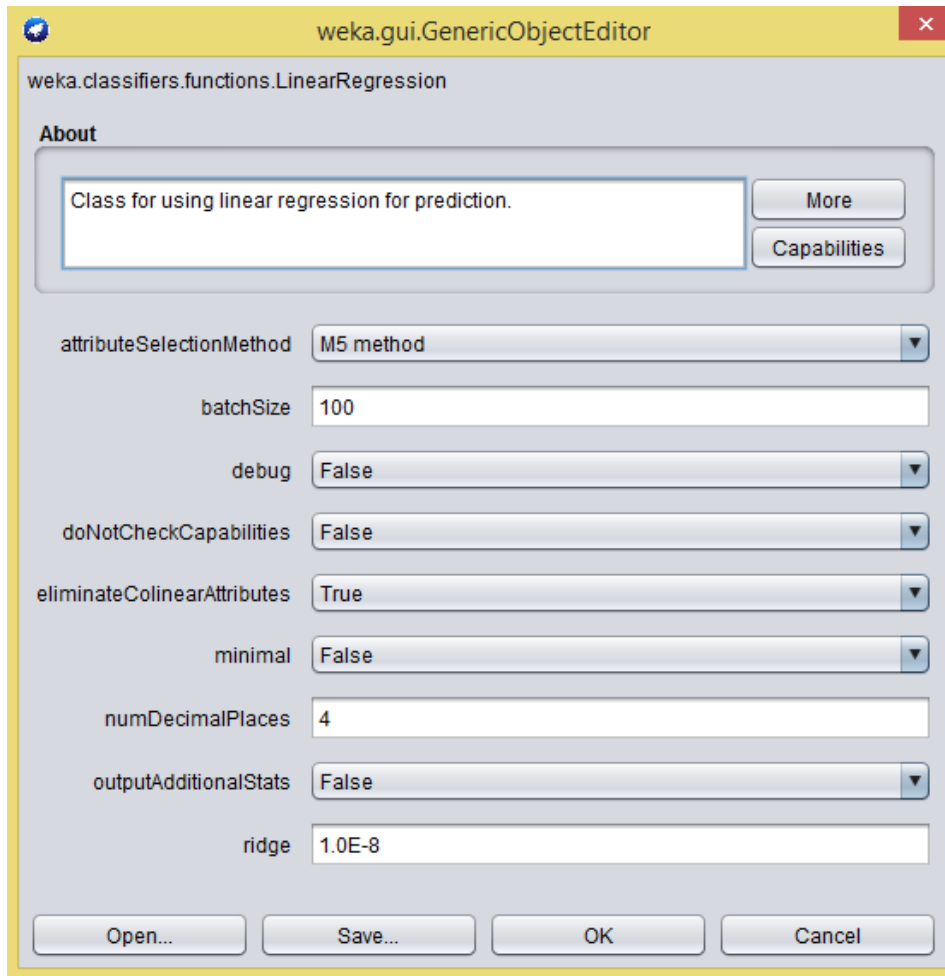
ภาพประกอบที่ 3.4 ค่าพารามิเตอร์ของเทคนิคโครงข่ายประสาทเทียมแบบถดถอย

จากภาพประกอบที่ 3.4 แสดงค่าพารามิเตอร์ต่างๆ ของเทคนิคโครงข่ายประสาทเทียมแบบถดถอย เช่น batchSize เป็นจำนวนตัวอย่างการพยากรณ์ที่ต้องการประมวลผลเบื้องต้น debug เป็นรายงานการแก้ไขข้อบกพร่องของข้อมูล doNotCheckCapabilities เป็นการตั้งค่าความสามารถในการจำแนกประเภทจะไม่ได้รับการตรวจสอบก่อนสร้างตัวจำแนกประเภท numDecimalPlaces เป็นรูปแบบจำนวนทศนิยม ridge เป็นความคุ่มค่าสูงสุดของพารามิเตอร์ seed เป็นผลที่ได้ของผู้ใช้ส่งไปสุ่มจำนวนใหม่ เป็นต้น



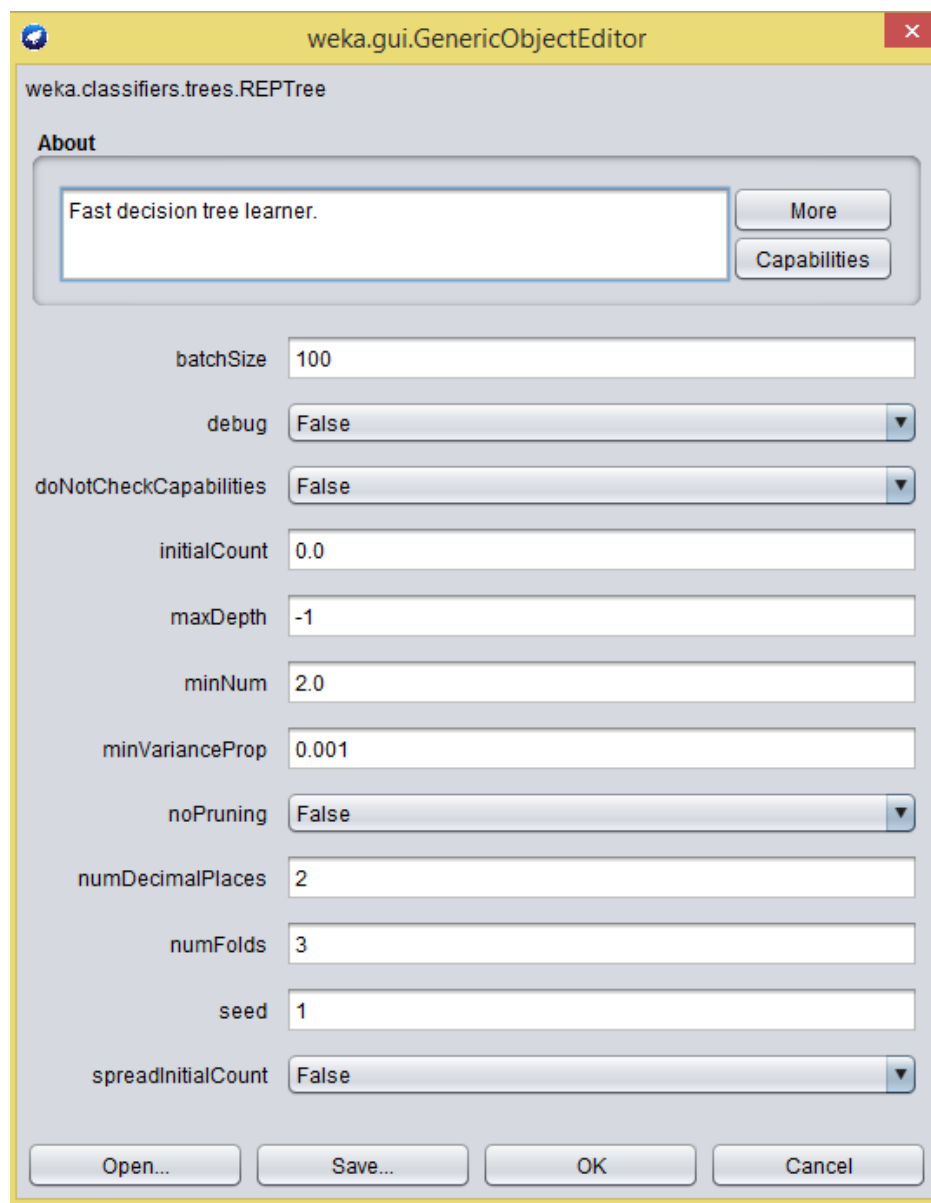
ภาพประกอบที่ 3.5 ค่าพารามิเตอร์ของเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย

จากภาพประกอบที่ 3.5 แสดงค่าพารามิเตอร์ของต่างๆ เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย ดังนี้ batchSize เป็นจำนวนตัวอย่างการพยากรณ์ที่ต้องการประมวลผลเบื้องต้น c คือ ค่ากำหนดของ c (ค่าปกติ = 1) debug เป็นรายงานการแก้ไขจุดข้อพร่องของข้อมูล doNotCheckCapabilities เป็นการตั้งค่าความสามารถในการจำแนกประเภทจะไม่ได้รับการตรวจสอบก่อนสร้างตัวจำแนกประเภท filterType เป็นการกำหนดข้อมูลให้มีการเปลี่ยนแปลง Kernel เป็นประเภทของ kernel numDecimalPlaces เป็นรูปแบบจำนวนทศนิยม regOptimizer คือ ขั้นตอนกระบวนการเรียนรู้



ภาพประกอบที่ 3.6 ค่าพารามิเตอร์ของเทคนิคการวิเคราะห์การถดถอยโลจิสติก

จากภาพประกอบที่ 3.6 แสดงค่าพารามิเตอร์ของต่างๆ ของเทคนิคการวิเคราะห์การถดถอยโลจิสติก ดังนี้ attributeSelectionMethod เป็นการเลือกวิธีการในการทำงานของเทคนิค Linear Regression ในเมนูเลือกเป็นค่าเริ่มต้น คือ M5 method ส่วนของ batchSize เป็นจำนวนตัวอย่างการพยากรณ์ที่ต้องการประมวลผลเบื้องต้น debug เป็นรายงานการแก้ไขข้อบกพร่องของข้อมูล doNotCheckCapabilities เป็นการตั้งค่าความสามารถในการจำแนกประเภทจะไม่ได้รับการตรวจสอบก่อนสร้างตัวจำแนกประเภท eliminate colinear attributes คือ การจำกัดคุณสมบัติในแนวเดียวกัน minimal คือ ค่าที่ต่ำที่สุด numDecimalPlaces เป็นรูปแบบจำนวนทศนิยม outputAdditionalStats คือ ค่าเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์และค่าสถิติ(t-test) ridge คือ ความคุ่มค่าสูงสุดของพารามิเตอร์



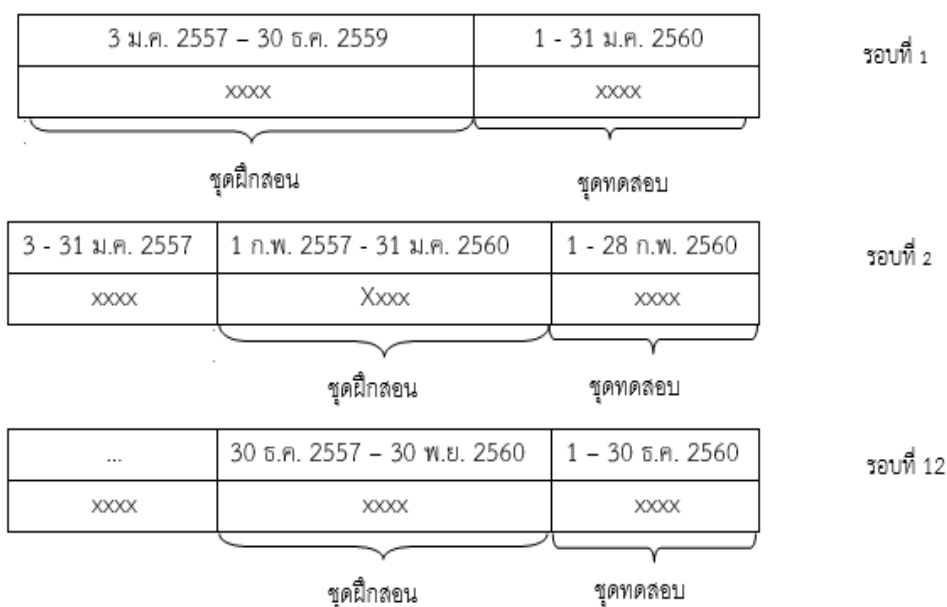
ภาพประกอบที่ 3.7 ค่าพารามิเตอร์ของเทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบลดถอย  
อัลกอริทึมการตัดกิ่งต้นไม้

จากภาพประกอบที่ 3.7 แสดงค่าพารามิเตอร์ของต่างๆของเทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบลดถอย อัลกอริทึมการตัดกิ่งต้นไม้ เช่น batchSize เป็นจำนวนตัวอย่างการพยากรณ์ที่ต้องการประมวลผลเบื้องต้น debug เป็นรายงานการแก้ไขข้อบกพร่องของข้อมูล doNotCheckCapabilities เป็นการตั้งค่าความสามารถในการจำแนกประเภทจะไม่ได้รับการตรวจสอบก่อนสร้างตัวจำแนกประเภท minNum เป็นจำนวนขั้นต่ำของกรณีต่อโหนดใบ numFolds เป็นจำนวนโพลด์ของข้อมูลที่ใช้สำหรับ Reduced-error pruning โดยหนึ่งโพลด์จะถูกใช้ในการตัดกิ่งของต้นไม้ และโพลด์ที่เหลือจะใช้

ในการสร้างต้นไม้ numDecimalPlaces เป็นรูปแบบจำนวนทศนิยม seed เป็นผลที่ได้ของผู้ใช้ส่งไปสู่จำนวนใหม่ เป็นต้น

### 3.4 การวัดประสิทธิภาพของแบบจำลอง

งานวิจัยนี้ เปรียบเทียบประสิทธิภาพการทำงานของแบบจำลองในการพยากรณ์ข้อมูลการใช้หนังสือในรูปแบบของอนุกรมเวลา ด้วยหลักการ Sliding Windows โดยการแบ่งข้อมูลออกเป็น 2 ชุด ประกอบด้วย ชุดข้อมูลฝึกสอน (Training set) โดยเริ่มรอบแรกจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง 30 ธันวาคม พ.ศ. 2559 และ ชุดข้อมูลทดสอบ (Testing set) โดยพยากรณ์ข้อมูลจากวันที่ 1 ถึง 31 มกราคม พ.ศ. 2560 แล้วเลื่อนไปข้างหน้าตามหลักการ Sliding Windows ไป 1 เดือน ดังนั้น ชุดข้อมูลฝึกสอน รอบที่ 2 เป็นระหว่าง วันที่ 1 กุมภาพันธ์ พ.ศ. 2557 ถึง 31 มกราคม พ.ศ. 2560 และ ชุดข้อมูลทดสอบรอบนี้ คือ ระหว่าง วันที่ 1 ถึง 28 กุมภาพันธ์ พ.ศ. 2560 และเลื่อนไปเรื่อยๆ ทั้งสิ้น 12 รอบ คือ ระยะเวลา 1 ปี ดังภาพประกอบที่ 3.8



ภาพประกอบที่ 3.8 การพยากรณ์ข้อมูลการใช้ด้วยหลักการหน้าต่างเคลื่อนที่

โดยทดสอบกับการทำงานของแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ ทีละ 1 แบบจำลอง และวัดค่าประสิทธิภาพด้วย mean absolute error และ root mean square error ที่นิยมใช้กันมากที่สุด เนื่องจากมีการแสดงผลเป็นอัตราของค่าความคลาดเคลื่อนที่เข้าใจได้



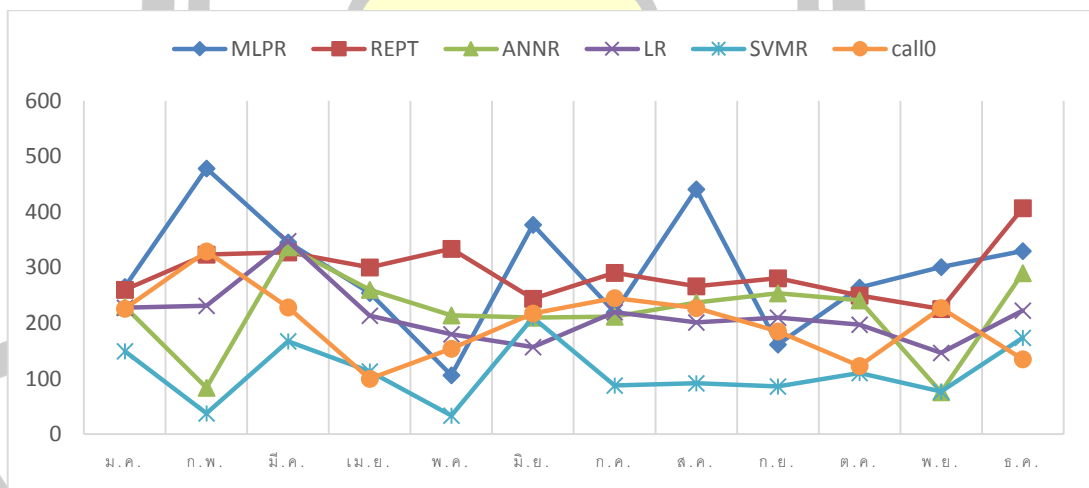
## บทที่ 4

### ผลการวิจัย

งานวิจัยนี้มีจุดมุ่งหมาย เพื่อพัฒนาแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม โดยได้นำข้อมูลจากระบบห้องสมุดอัตโนมัติ VTLS ซึ่งมีการจัดเก็บข้อมูลการใช้หนังสือ จาก 10 หมวดหนังสือ เป็นรายวัน เริ่มจากวันที่ 3 มกราคม พ.ศ. 2557 ถึง ธันวาคม พ.ศ. 2560 และสร้างแบบจำลอง โดยนำเอาโปรแกรม WEKA เวอร์ชัน 3.9.2 มาใช้สร้างแบบจำลองด้วยเทคนิคแบบต่างๆ แล้ววัดประสิทธิภาพของแบบจำลอง ผลการทดลองสามารถแสดงตามหัวข้อ ดังนี้

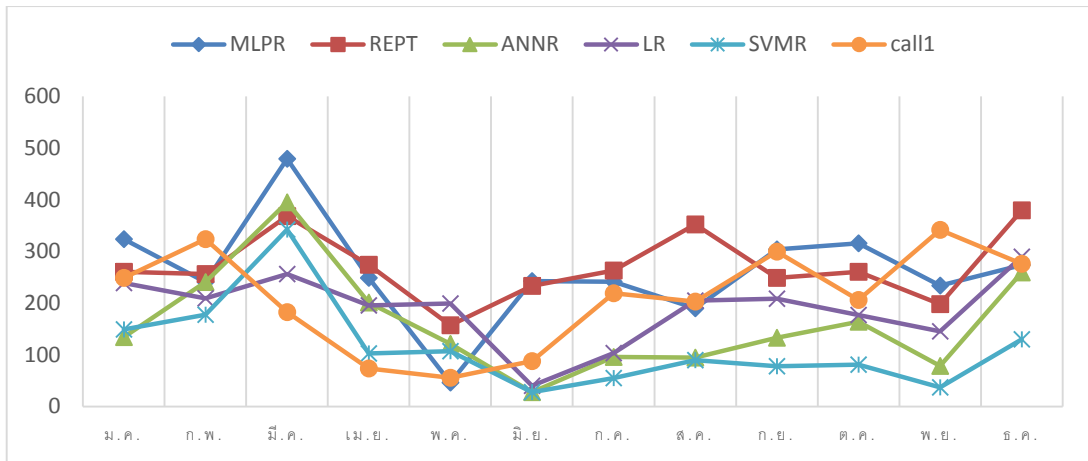
#### 4.1 ผลการพยากรณ์การใช้หนังสือ

ผลการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม โดยนำโปรแกรม WEKA มาสร้างแบบจำลองด้วยเทคนิค MLPR ANNR SVMR LR และ REPT โดยสามารถแสดงปริมาณการใช้จริงกับปริมาณจากการพยากรณ์ของหมวดหนังสือต่างๆ ได้ดังภาพประกอบที่ 4.1 – 4.10



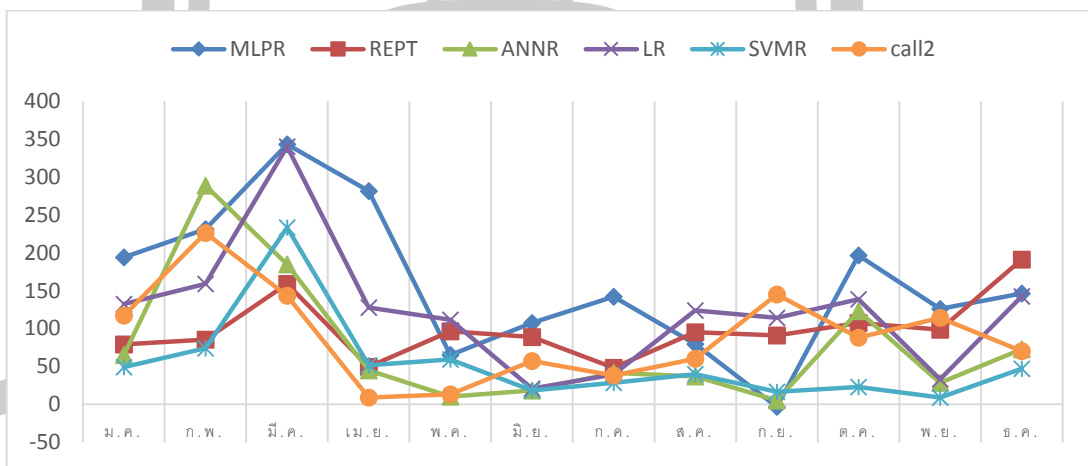
ภาพประกอบที่ 4.1 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 000

จากภาพประกอบที่ 4.1 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 000 พบว่าแบบจำลองที่สร้างด้วยเทคนิค LR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



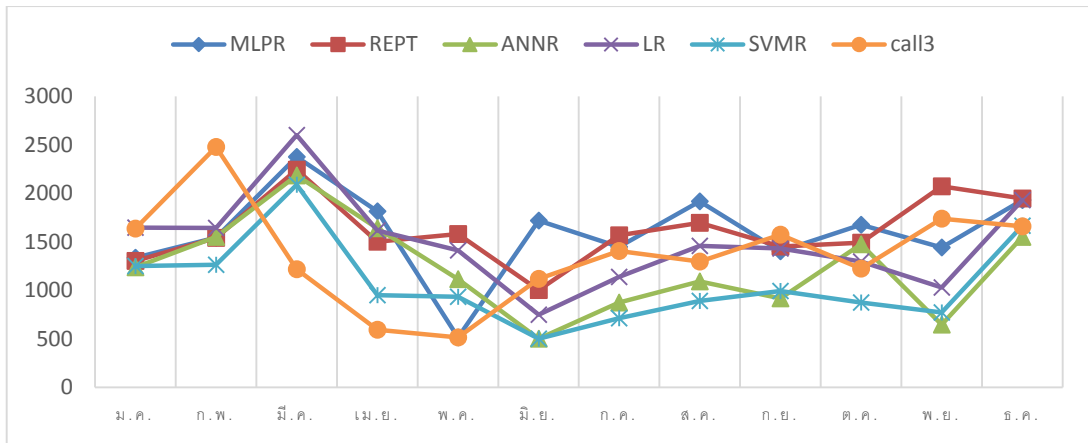
ภาพประกอบที่ 4.2 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 100

จากภาพประกอบที่ 4.2 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 100 พบว่าแบบจำลองที่สร้างด้วยเทคนิค LR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



ภาพประกอบที่ 4.3 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 200

จากภาพประกอบที่ 4.3 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 200 พบว่าแบบจำลองที่สร้างด้วยเทคนิค ANNR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



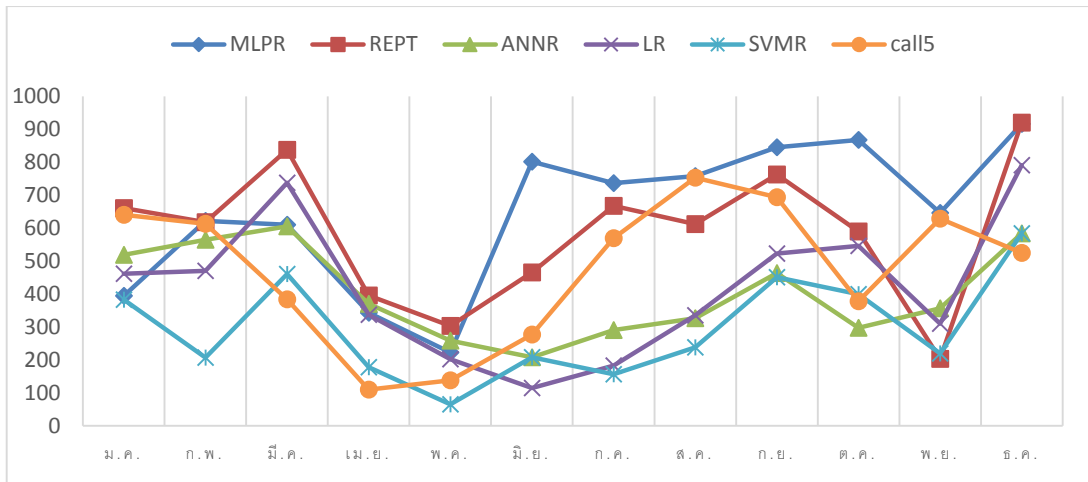
ภาพประกอบที่ 4.4 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 300

จากภาพประกอบที่ 4.4 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 300 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



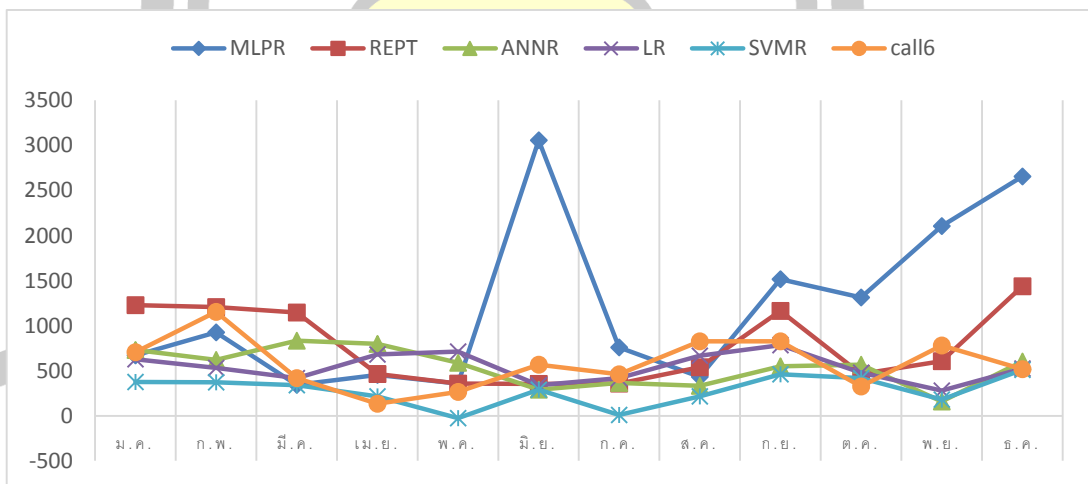
ภาพประกอบที่ 4.5 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 400

จากภาพประกอบที่ 4.5 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 400 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



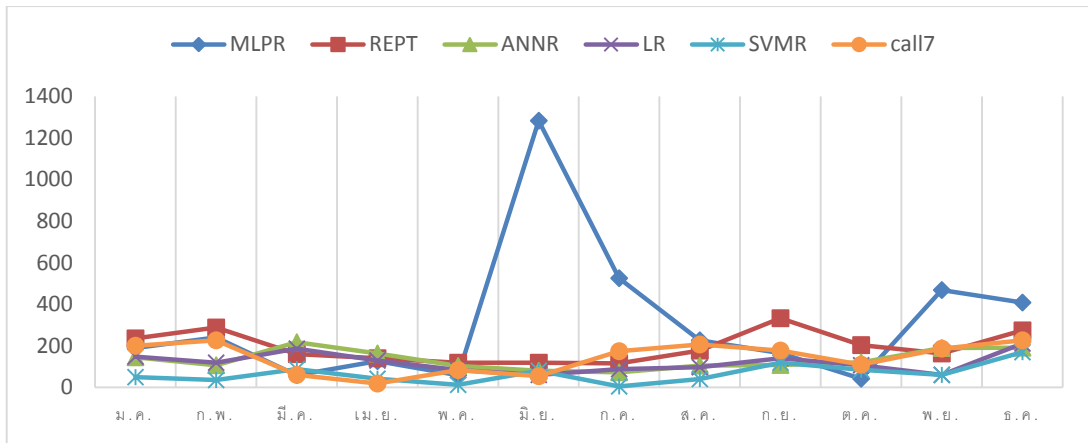
ภาพประกอบที่ 4.6 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 500

จากภาพประกอบที่ 4.6 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 500 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



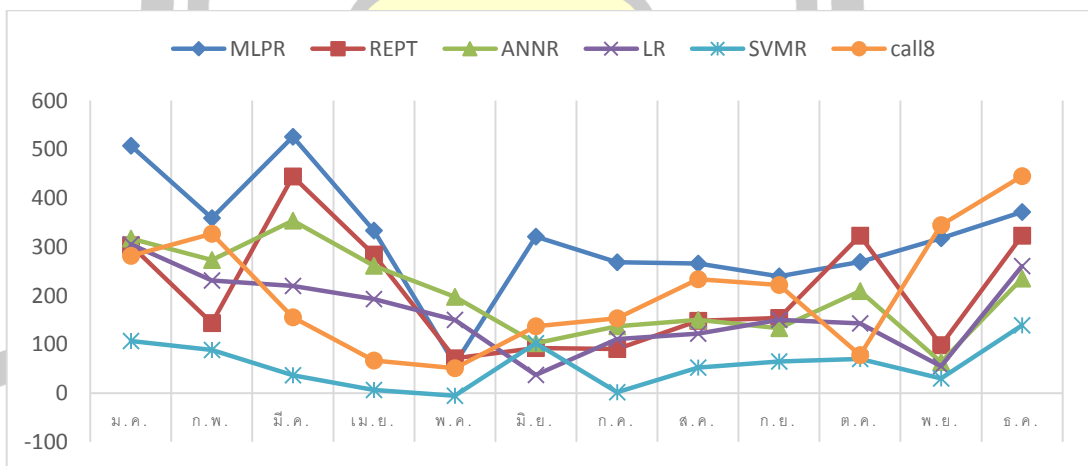
ภาพประกอบที่ 4.7 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 600

จากภาพประกอบที่ 4.7 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 600 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



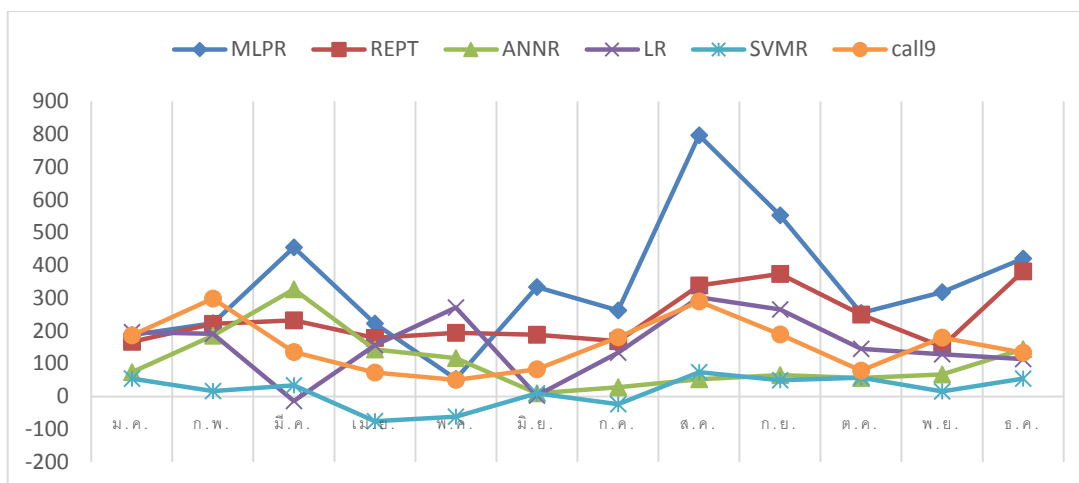
ภาพประกอบที่ 4.8 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 700

จากภาพประกอบที่ 4.8 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 700 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



ภาพประกอบที่ 4.9 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 800

จากภาพประกอบที่ 4.9 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 800 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด



ภาพประกอบที่ 4.10 กราฟแสดงการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ของหมวด 900

จากภาพประกอบที่ 4.10 กราฟแสดงผลการพยากรณ์ด้วยเทคนิคต่างๆ ของหมวด 900 พบว่าแบบจำลองที่สร้างด้วยเทคนิค SVMR มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด

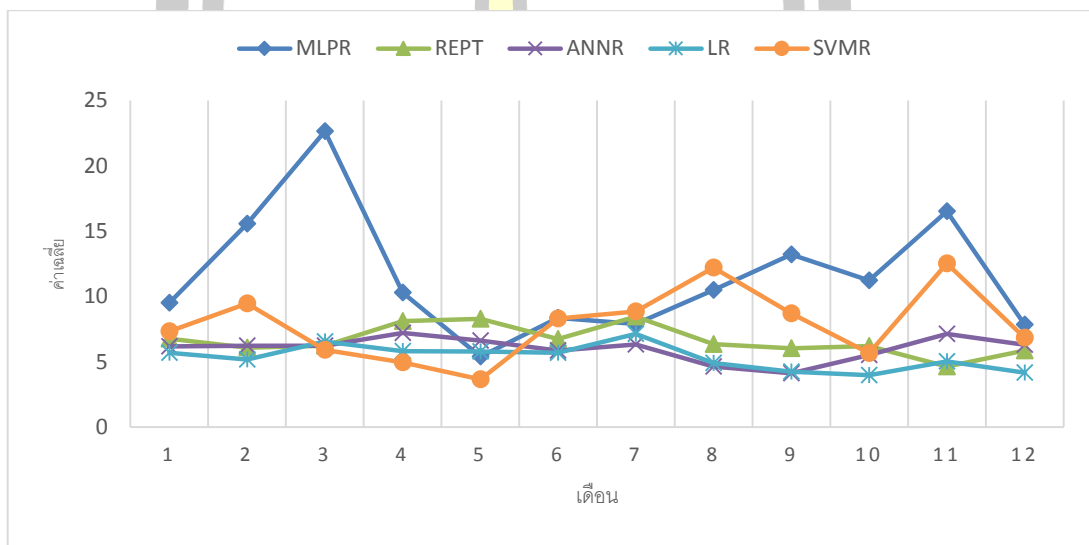
จากภาพประกอบที่ 4.1 – 4.10 แสดงกราฟการเปรียบเทียบระหว่างปริมาณการใช้หนังสือจริงกับการพยากรณ์ด้วยแบบจำลองต่างๆ ในระหว่าง เดือน มกราคม - ธันวาคม 2560 ของหนังสือทั้ง 10 หมวดหมู่ พบว่า กราฟปริมาณการใช้หนังสือจริงเป็นกราฟเส้นที่มีการขึ้นลงตามปริมาณการใช้หนังสือกราฟที่เกิดจากการพยากรณ์ด้วยเทคนิคถดถอยต่างๆ มีลักษณะเป็นเส้นขึ้นลงสอดคล้องกับกราฟปริมาณการใช้หนังสือจริง โดยแบบจำลองที่มีผลการพยากรณ์ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด คือ เทคนิค SVMR (7 หมวด) รองลงมา ได้แก่ เทคนิค LR (2 หมวด) และเทคนิค ANNR (1 หมวด) ตามลำดับ

#### 4.2 เปรียบเทียบประสิทธิภาพแบบจำลอง

การเปรียบเทียบประสิทธิภาพแบบจำลองครั้งนี้ ใช้หลักการหน้าต่างเคลื่อนที่ แล้ววัดค่าด้วยการหาค่าเฉลี่ยคลาดเคลื่อนสมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง จากการศึกษา พบว่า

4.3.1 การวัดค่าด้วยการหาค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (MAE) ของแบบจำลองในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด เป็นค่าความแตกต่างโดยเฉลี่ยระหว่างการใช้จริงกับการพยากรณ์ของแต่ละแบบจำลอง ประกอบด้วย เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้น

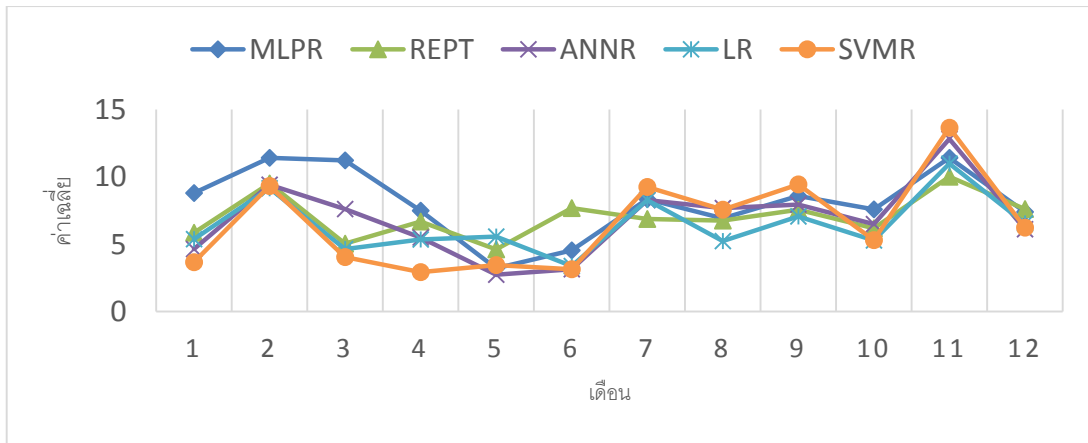
แบบถดถอย เทคนิคโครงข่ายประสาทเทียมแบบถดถอย เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย การวิเคราะห์การถดถอยโลจิสติก และการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ ซึ่งมีค่าความแตกต่างยิ่งน้อยยิ่งดี โดยสามารถแสดงการวัดค่าด้วยการหาค่าเฉลี่ย ความคลาดเคลื่อนสัมบูรณ์ของปริมาณการใช้จริงและการใช้จากการพยากรณ์ตามหมวดหนังสือต่างๆ ได้ดังภาพประกอบที่ 4.11 - 4.20



ภาพประกอบที่ 4.11 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 000

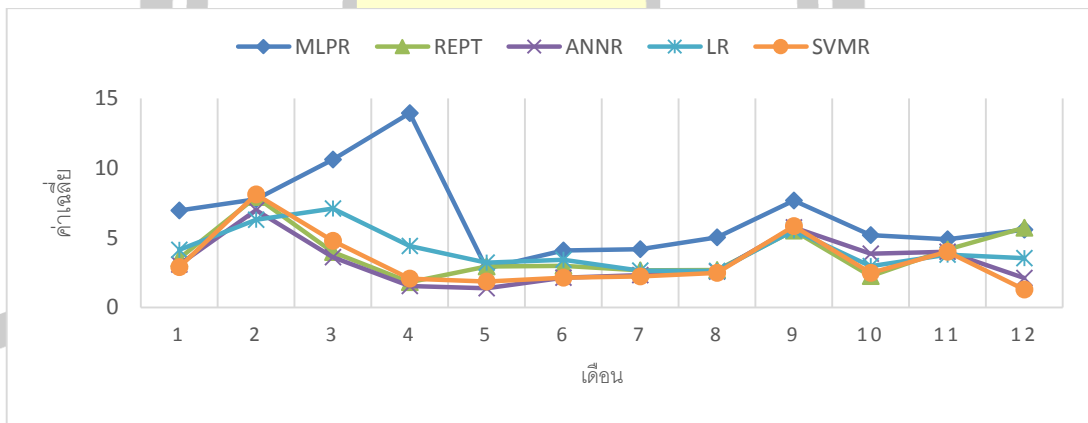
จากภาพประกอบที่ 4.11 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 000 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค LR มีค่าต่ำสุด

พหุ ประถมศึกษา



ภาพประกอบที่ 4.12 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 100

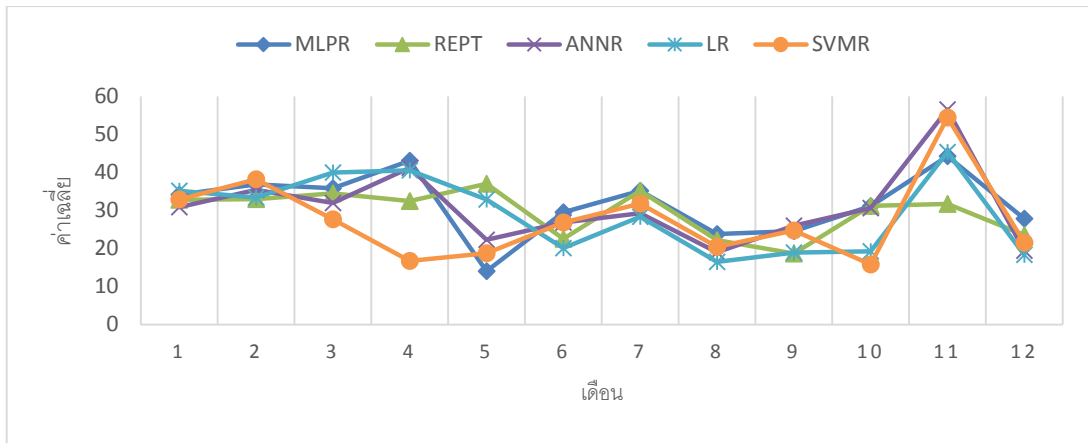
จากภาพประกอบที่ 4.12 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 100 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค LR มีค่าต่ำสุด



ภาพประกอบที่ 4.13 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 200

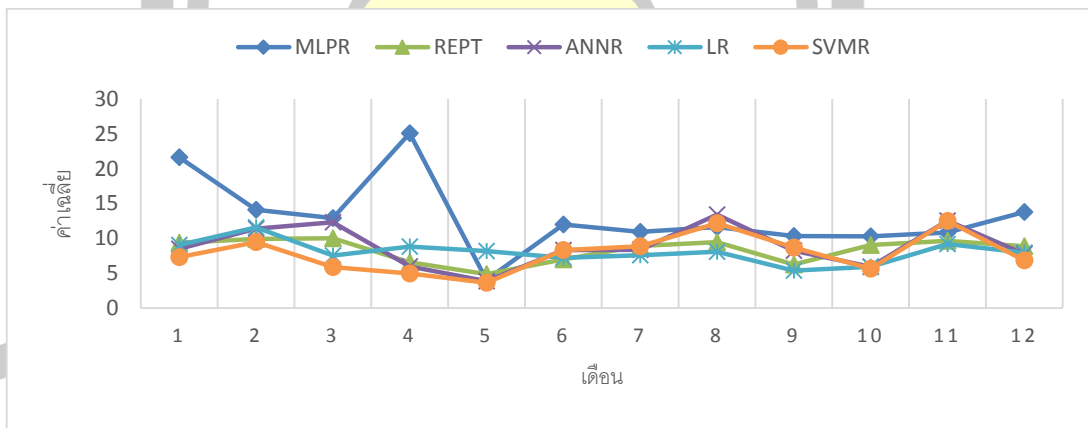
จากภาพประกอบที่ 4.13 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 200 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค ANNR มีค่าต่ำสุด





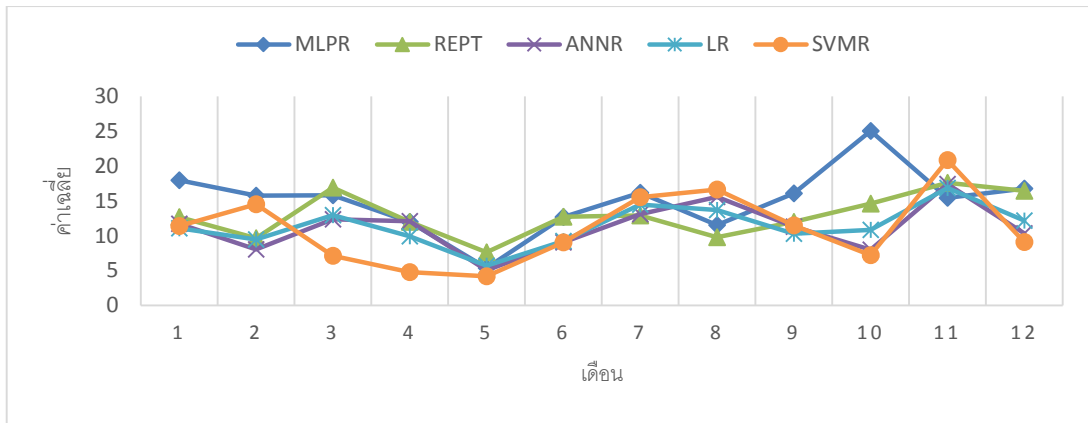
ภาพประกอบที่ 4.14 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ  
ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 300

จากภาพประกอบที่ 4.14 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 300 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



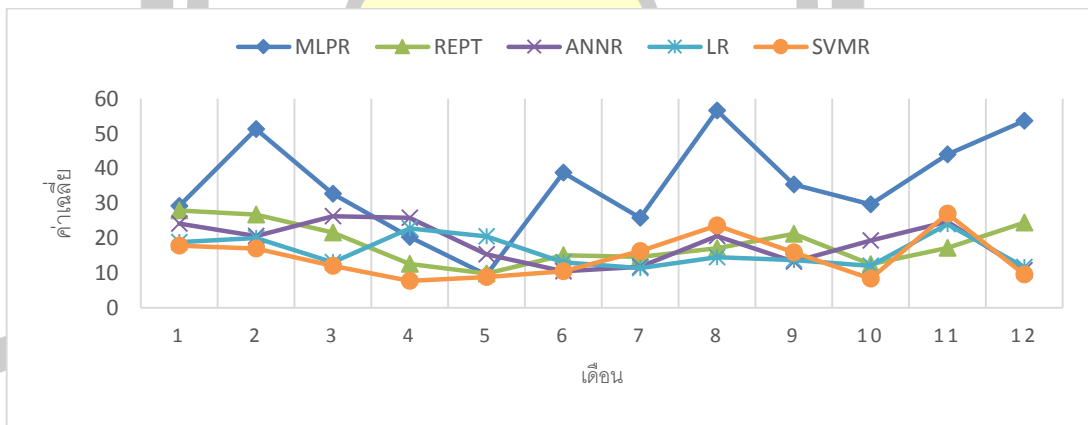
ภาพประกอบที่ 4.15 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ  
ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 400

จากภาพประกอบที่ 4.15 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 400 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



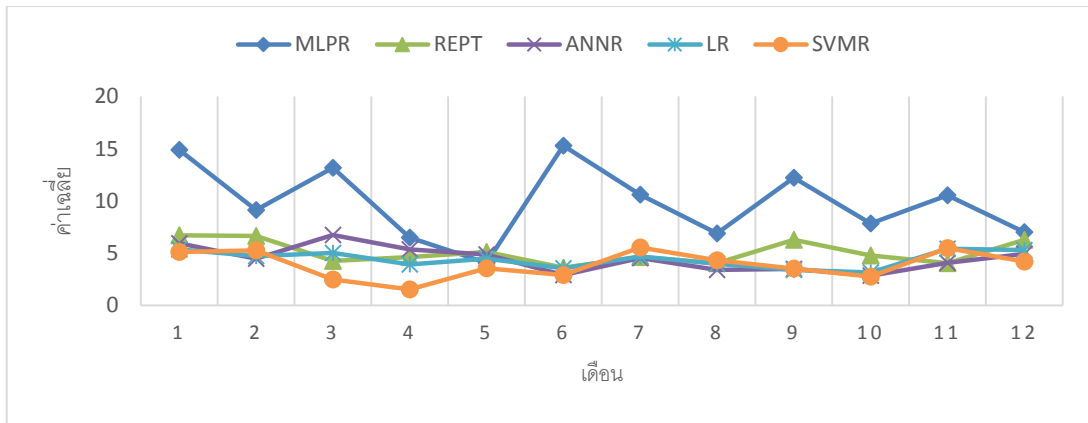
ภาพประกอบที่ 4.16 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 500

จากภาพประกอบที่ 4.16 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 500 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



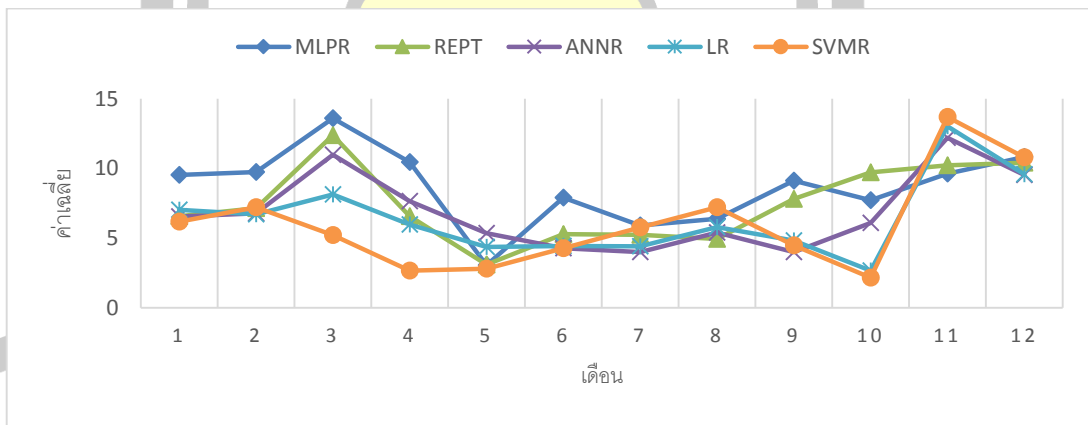
ภาพประกอบที่ 4.17 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 600

จากภาพประกอบที่ 4.17 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 600 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



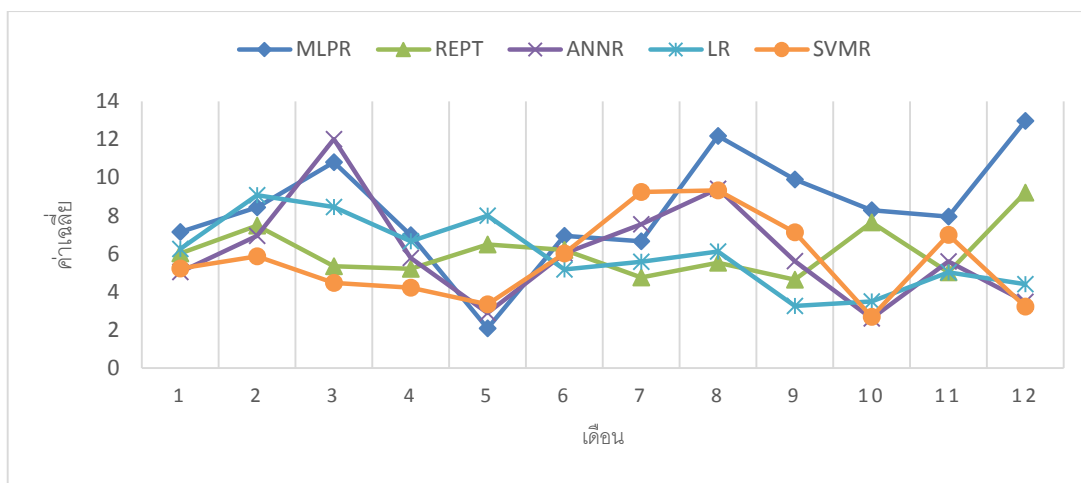
ภาพประกอบที่ 4.18 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 700

จากภาพประกอบที่ 4.18 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 700 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



ภาพประกอบที่ 4.19 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 800

จากภาพประกอบที่ 4.19 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ์ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 800 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



ภาพประกอบที่ 4.20 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณ  
ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 900

จากภาพประกอบที่ 4.20 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนสมบูรณระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 900 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด

จากภาพประกอบที่ 4.11 - 4.20 กราฟแสดงการเปรียบเทียบระหว่างค่าความคลาดเคลื่อนสมบูรณ ของการพยากรณ์ด้วยแบบจำลองต่างๆ ในระหว่าง เดือน มกราคม - ธันวาคม 2560 ของหนังสือทั้ง 10 หมวดหมู่ พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยความคลาดเคลื่อนสมบูรณแล้ว แบบจำลองที่มีค่าต่ำสุด คือ SVMR (7 หมวด) รองลงมา ได้แก่ เทคนิค LR (2 หมวด) และเทคนิค ANNR (1 หมวด) ตามลำดับ

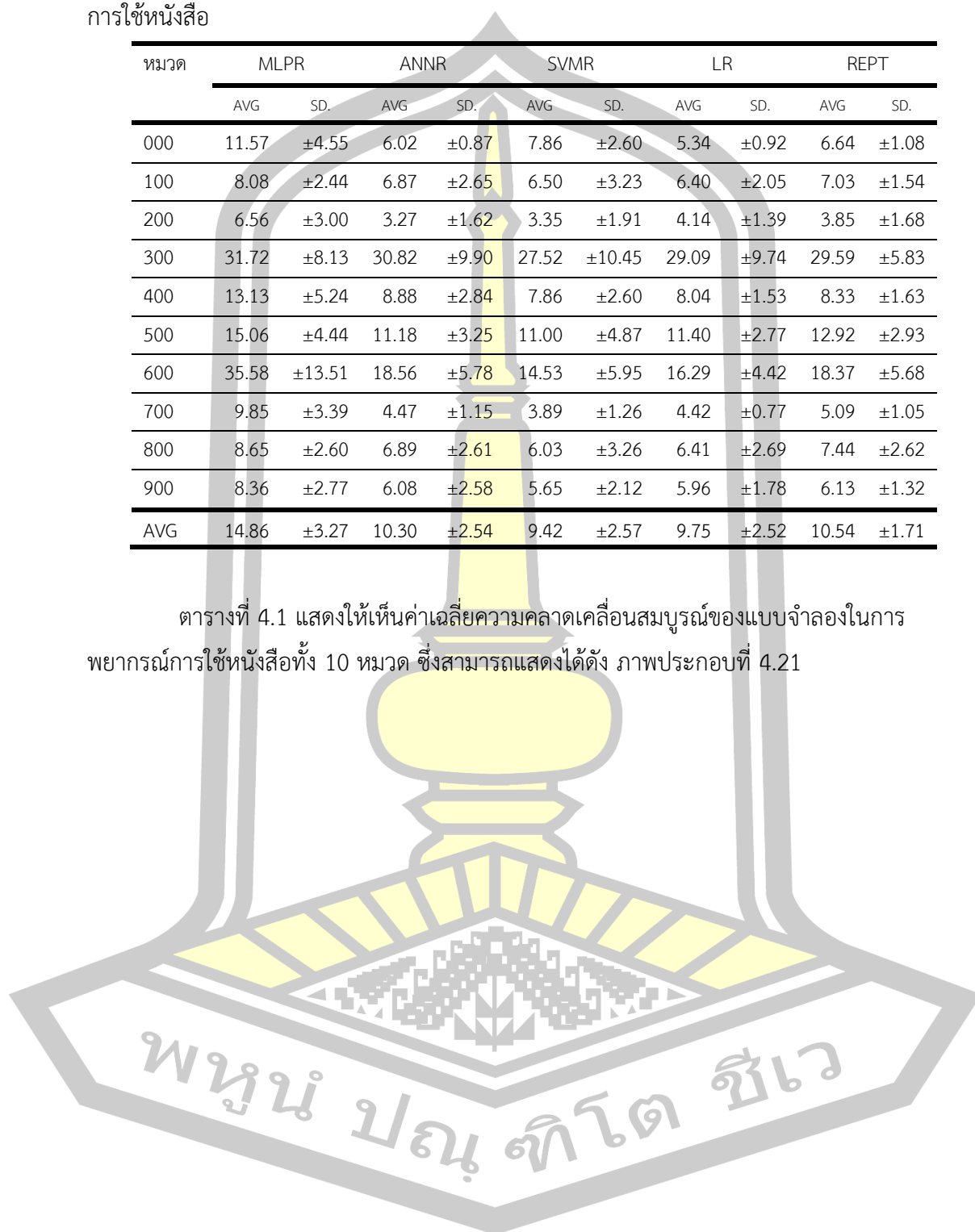
จากค่าเฉลี่ย MAE ของแบบจำลองในการพยากรณ์การใช้หนังสือในแต่ละหมวด 12 รอบ จึงได้นำมาหาค่าเฉลี่ย (Average : AVG) ของ MAE ของแต่ละหมวดหนังสือ และค่าความผันผวน (Standard Deviation : SD.) อีกรอบ ดังตารางที่ 4.1

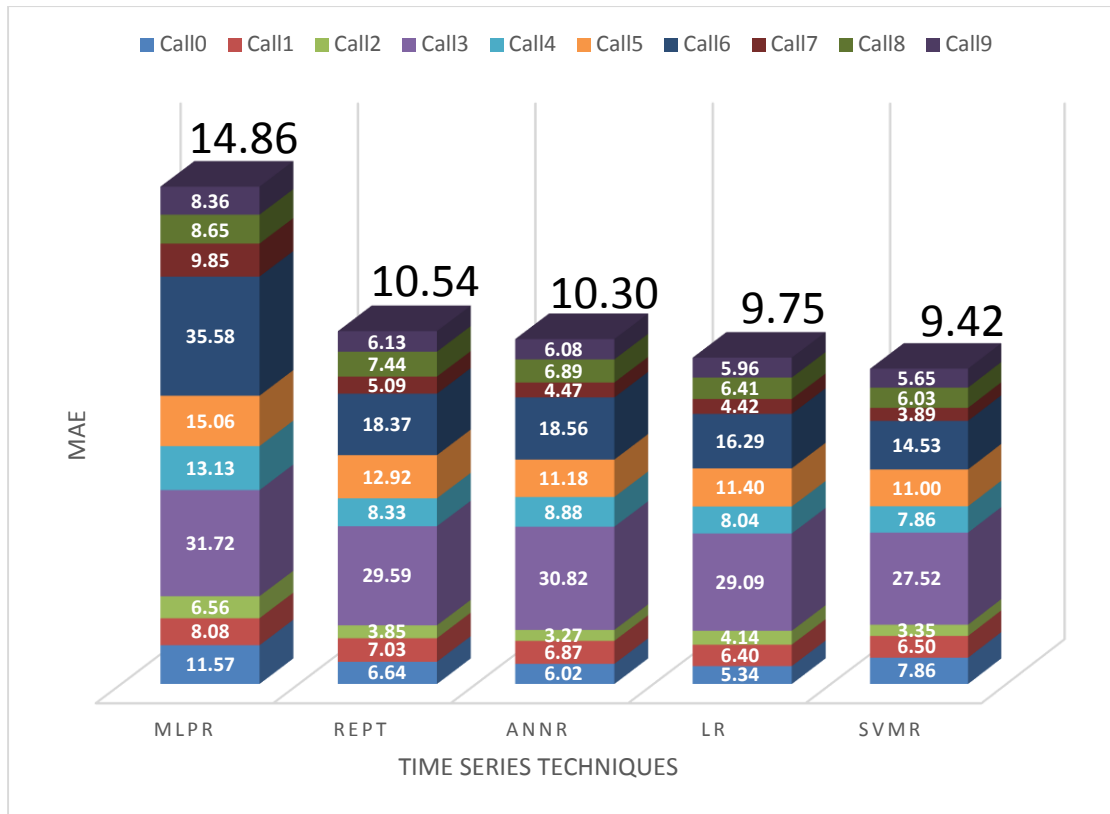
พหุบัณฑิต ชีวะ

ตารางที่ 4.1 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ ของแบบจำลองในการพยากรณ์การใช้หนังสือ

| หมวด | MLPR  |        | ANNR  |       | SVMR  |        | LR    |       | REPT  |       |
|------|-------|--------|-------|-------|-------|--------|-------|-------|-------|-------|
|      | AVG   | SD.    | AVG   | SD.   | AVG   | SD.    | AVG   | SD.   | AVG   | SD.   |
| 000  | 11.57 | ±4.55  | 6.02  | ±0.87 | 7.86  | ±2.60  | 5.34  | ±0.92 | 6.64  | ±1.08 |
| 100  | 8.08  | ±2.44  | 6.87  | ±2.65 | 6.50  | ±3.23  | 6.40  | ±2.05 | 7.03  | ±1.54 |
| 200  | 6.56  | ±3.00  | 3.27  | ±1.62 | 3.35  | ±1.91  | 4.14  | ±1.39 | 3.85  | ±1.68 |
| 300  | 31.72 | ±8.13  | 30.82 | ±9.90 | 27.52 | ±10.45 | 29.09 | ±9.74 | 29.59 | ±5.83 |
| 400  | 13.13 | ±5.24  | 8.88  | ±2.84 | 7.86  | ±2.60  | 8.04  | ±1.53 | 8.33  | ±1.63 |
| 500  | 15.06 | ±4.44  | 11.18 | ±3.25 | 11.00 | ±4.87  | 11.40 | ±2.77 | 12.92 | ±2.93 |
| 600  | 35.58 | ±13.51 | 18.56 | ±5.78 | 14.53 | ±5.95  | 16.29 | ±4.42 | 18.37 | ±5.68 |
| 700  | 9.85  | ±3.39  | 4.47  | ±1.15 | 3.89  | ±1.26  | 4.42  | ±0.77 | 5.09  | ±1.05 |
| 800  | 8.65  | ±2.60  | 6.89  | ±2.61 | 6.03  | ±3.26  | 6.41  | ±2.69 | 7.44  | ±2.62 |
| 900  | 8.36  | ±2.77  | 6.08  | ±2.58 | 5.65  | ±2.12  | 5.96  | ±1.78 | 6.13  | ±1.32 |
| AVG  | 14.86 | ±3.27  | 10.30 | ±2.54 | 9.42  | ±2.57  | 9.75  | ±2.52 | 10.54 | ±1.71 |

ตารางที่ 4.1 แสดงให้เห็นค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ของแบบจำลองในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด ซึ่งสามารถแสดงได้ดัง ภาพประกอบที่ 4.21

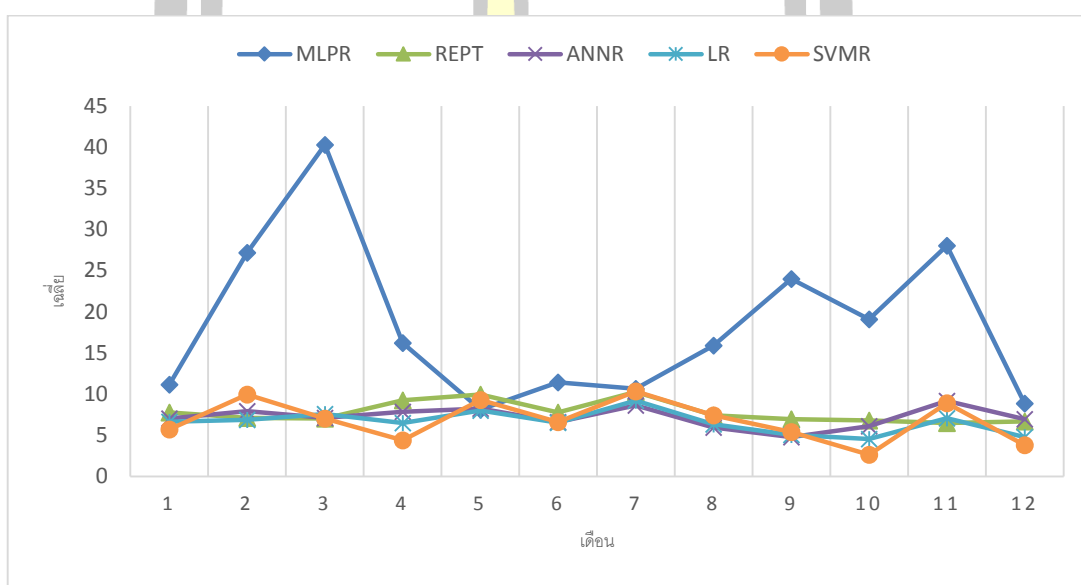




ภาพประกอบที่ 4.21 แสดงค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ของแบบจำลองในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด

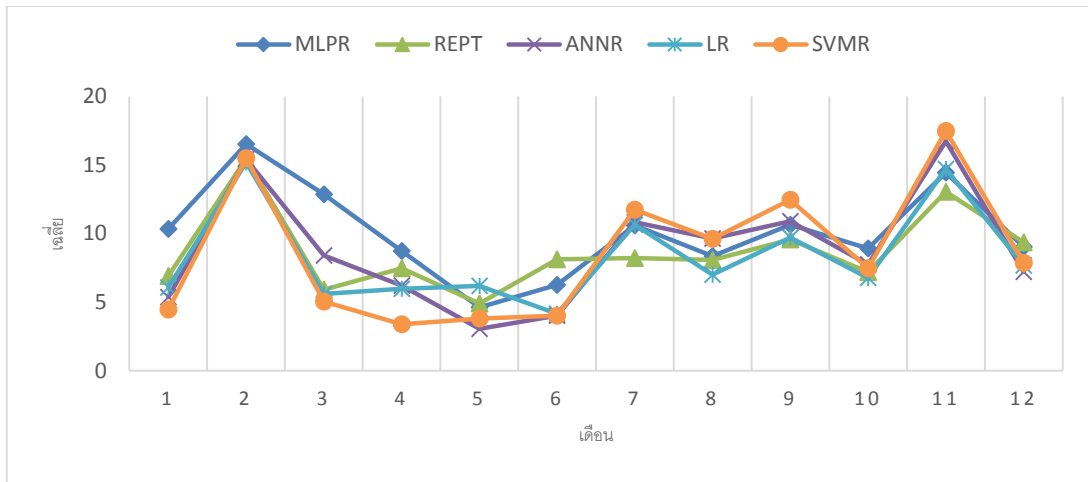
จากภาพประกอบที่ 4.21 พบว่า ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ของแบบจำลองในการพยากรณ์ที่ค่าต่ำสุด อยู่ในระดับดีมาก ได้แก่ เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (ค่าเฉลี่ย 9.42) แต่มีค่าเบี่ยงเบนมาตรฐาน ( $\pm 2.57$ ) ต่ำกว่าเพียงเทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย ( $\pm 3.27$ ) เท่านั้น ค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์รองลงมาที่อยู่ในระดับดีมาก ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจิสติก (ค่าเฉลี่ย 9.75) ความคลาดเคลื่อนของแบบจำลองในการพยากรณ์ที่อยู่ในระดับดี ได้แก่ เทคนิคโครงข่ายประสาทเทียมแบบถดถอย (ค่าเฉลี่ย 10.30) การเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ (ค่าเฉลี่ย 10.54) ตามลำดับ และความคลาดเคลื่อนของแบบจำลองในการพยากรณ์ที่ค่าสูงสุดอยู่ในระดับน้อย คือ เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นแบบถดถอย (ค่าเฉลี่ย 14.86)

4.3.2 การเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (RMSE) ของแบบจำลอง ในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด เป็นค่าความแตกต่างรากที่สองของการพยากรณ์แต่ละแบบจำลอง ประกอบด้วย เทคนิคโครงข่ายประสาทเทียมหลายชั้นแบบถดถอย เทคนิคโครงข่ายประสาทเทียมแบบถดถอย เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย การวิเคราะห์การถดถอยโลจิสติก และการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ โดยสามารถแสดงการวัดค่าด้วยความคลาดเคลื่อนเฉลี่ยรากที่สอง ได้ดังภาพประกอบที่ 4.22 – 4.31



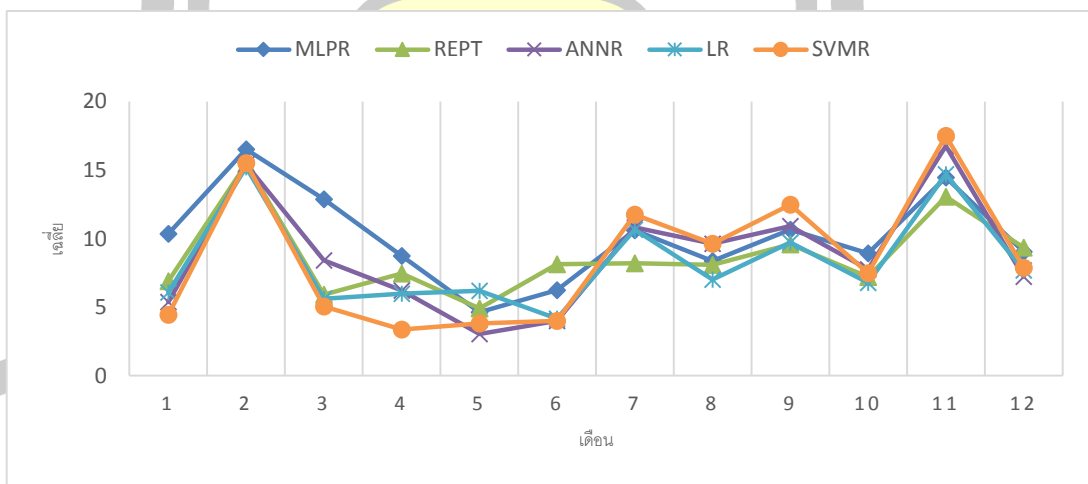
ภาพประกอบที่ 4.22 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 000

จากภาพประกอบที่ 4.22 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 000 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค LR มีค่าต่ำสุด



ภาพประกอบที่ 4.23 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 100

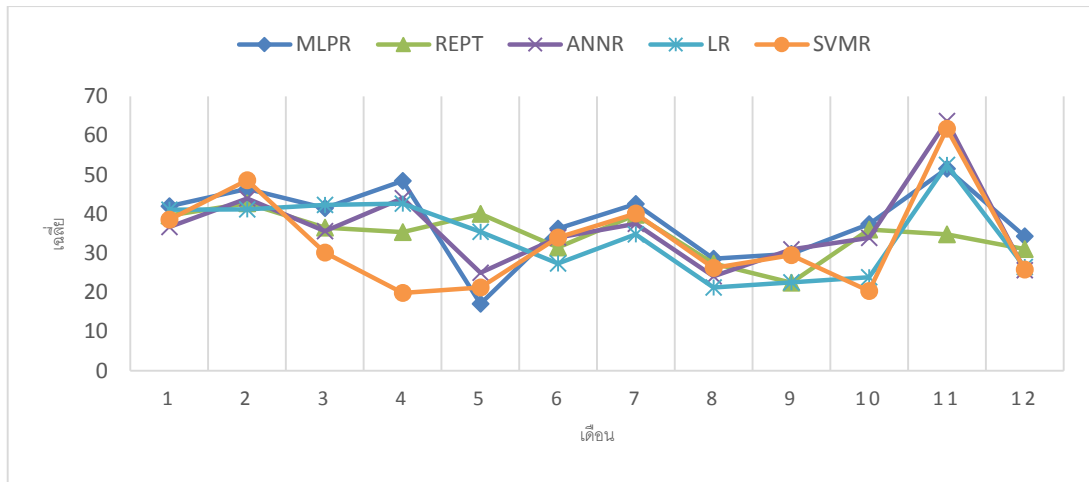
จากภาพประกอบที่ 4.23 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 100 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค LR มีค่าต่ำสุด



ภาพประกอบที่ 4.24 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 200

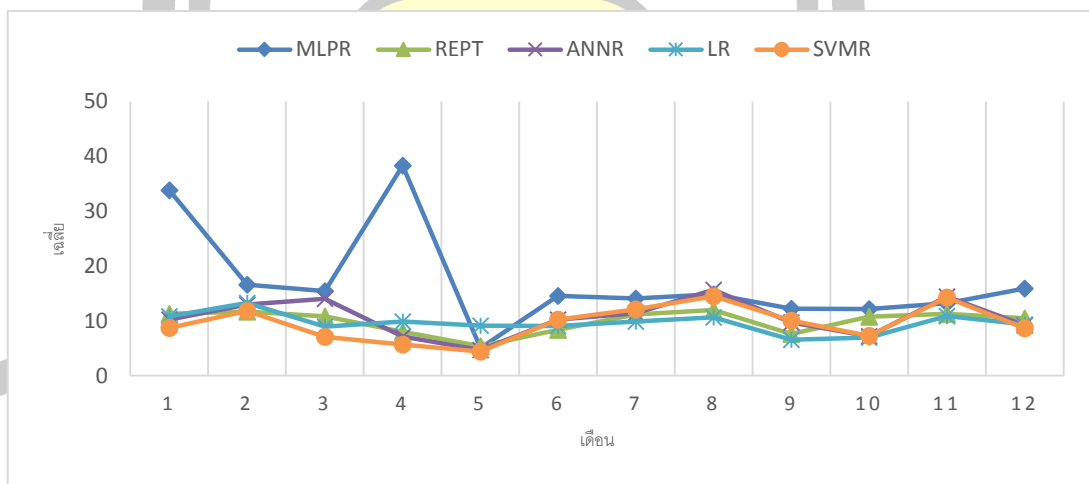
จากภาพประกอบที่ 4.24 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 200 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค ANNR มีค่าต่ำสุด





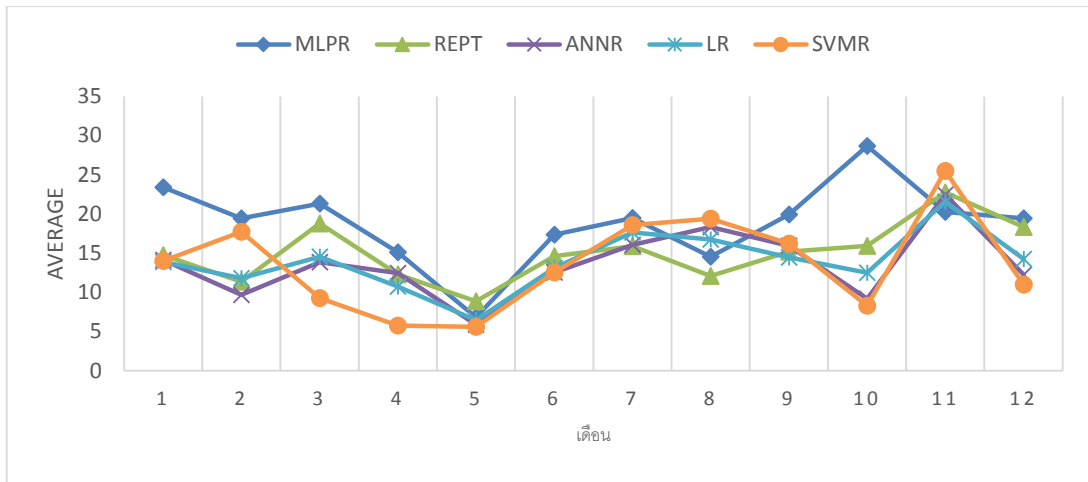
ภาพประกอบที่ 4.25 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 300

จากภาพประกอบที่ 4.25 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 300 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



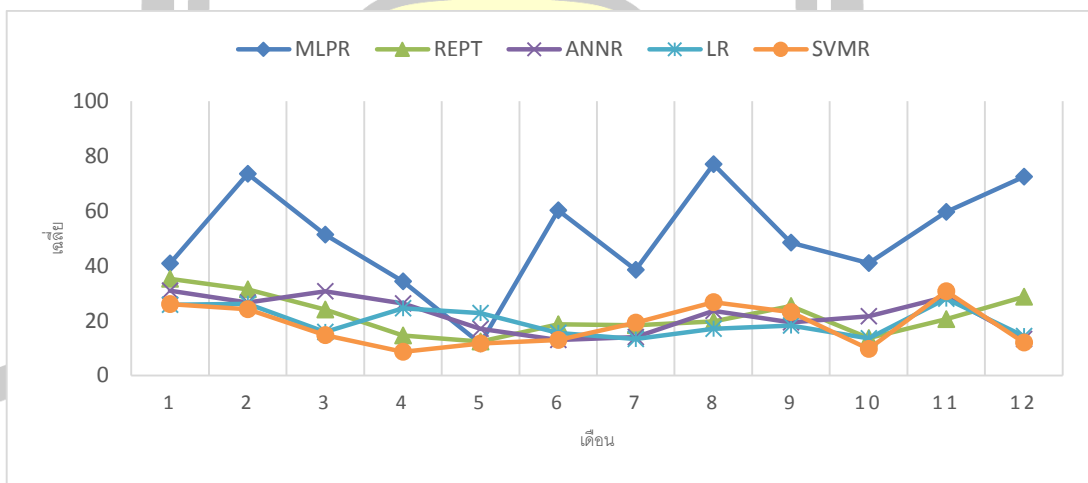
ภาพประกอบที่ 4.26 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 400

จากภาพประกอบที่ 4.26 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 400 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



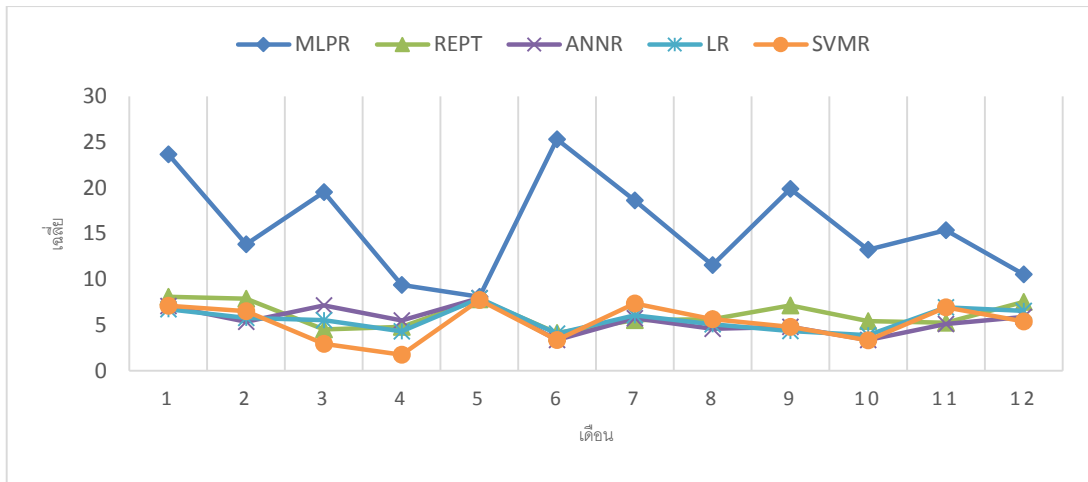
ภาพประกอบที่ 4.27 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 500

จากภาพประกอบที่ 4.27 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 500 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



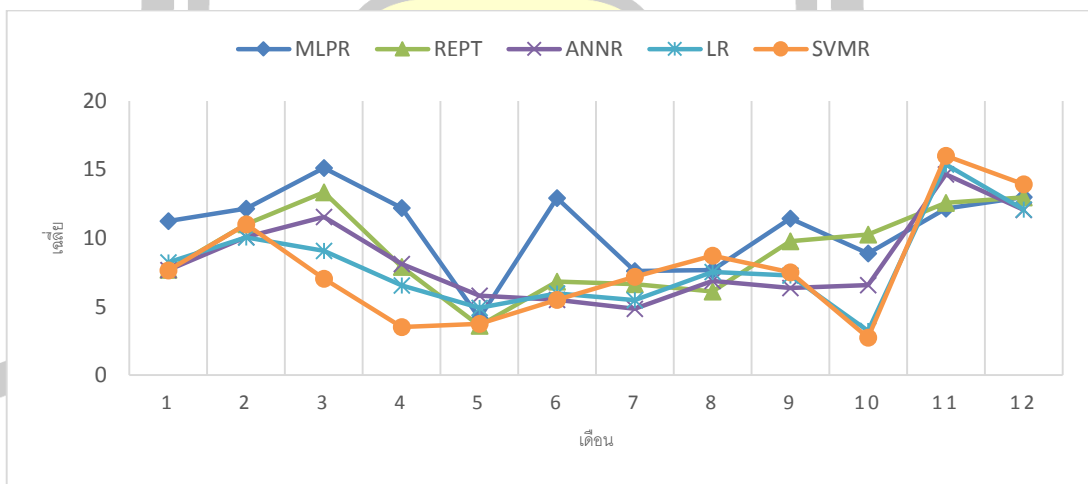
ภาพประกอบที่ 4.28 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 600

จากภาพประกอบที่ 4.28 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 600 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



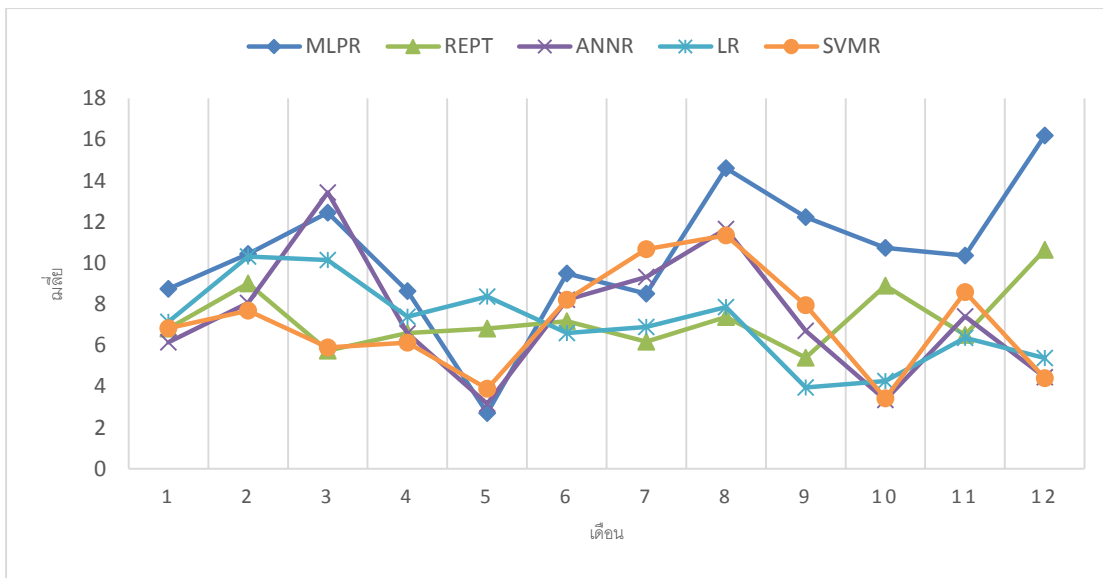
ภาพประกอบที่ 4.29 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 700

จากภาพประกอบที่ 4.29 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 700 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



ภาพประกอบที่ 4.30 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 800

จากภาพประกอบที่ 4.30 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 800 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR มีค่าต่ำสุด



ภาพประกอบที่ 4.31 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 900

จากภาพประกอบที่ 4.31 กราฟแสดงการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ระหว่างการพยากรณ์ด้วยแบบจำลองต่างๆ ของหมวด 900 พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ยแล้ว แบบจำลองที่สร้างด้วยเทคนิค LR มีค่าต่ำสุด

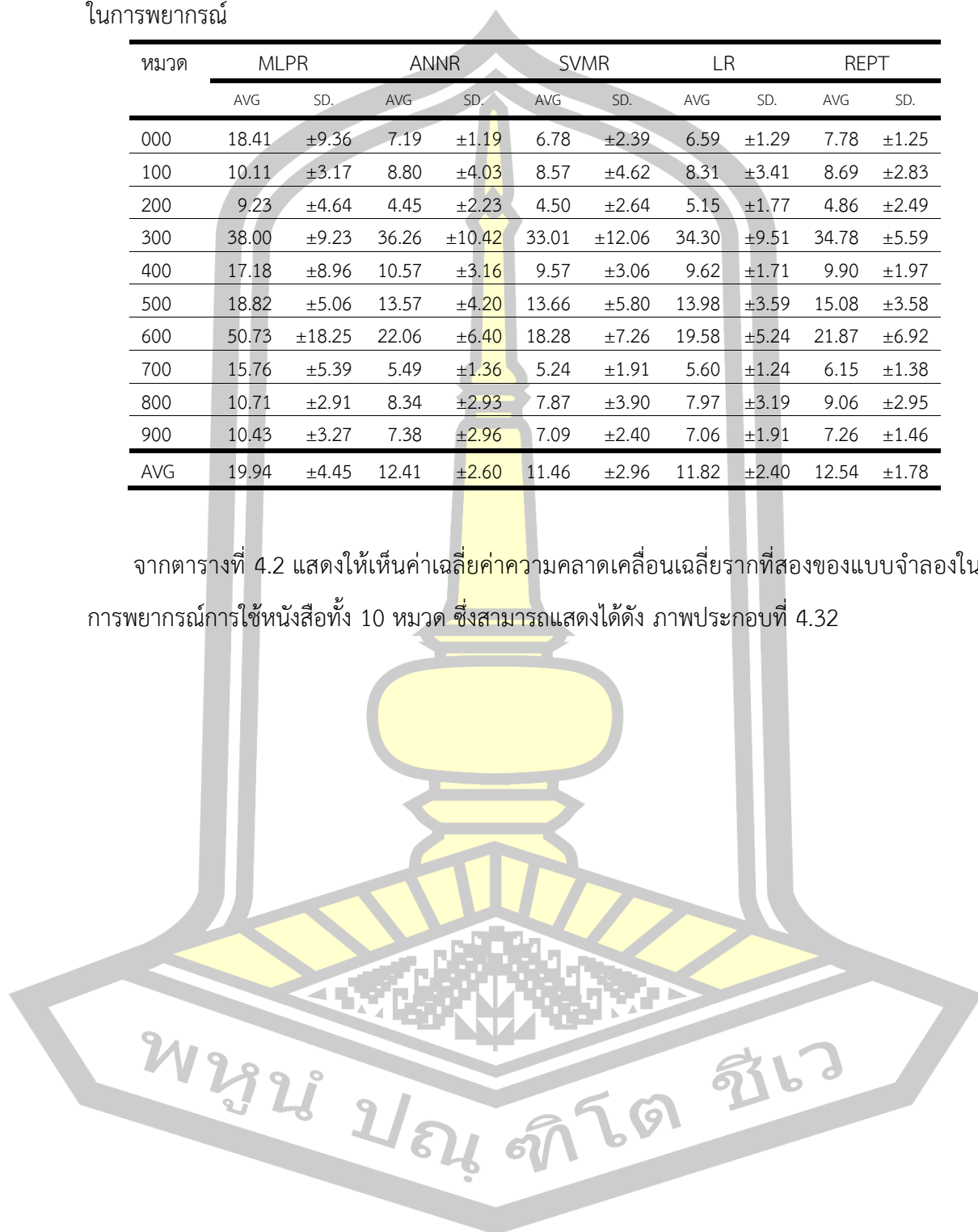
จากภาพประกอบที่ 4.22 – 4.31 แสดงกราฟการเปรียบเทียบระหว่างค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ของการพยากรณ์ด้วยแบบจำลองต่างๆ ในระหว่าง เดือน มกราคม - ธันวาคม 2560 ของหนังสือทั้ง 10 หมวดหมู่ พบว่า ทุกเทคนิคมีค่าความคลาดเคลื่อนที่ผันผวน แต่จากค่าเฉลี่ย ค่าความคลาดเคลื่อนเฉลี่ยรากที่สองแล้ว แบบจำลองที่สร้างด้วยเทคนิค SVMR (6 หมวด) มีค่าต่ำสุด รองลงมา ได้แก่ เทคนิค LR (3 หมวด) และเทคนิค ANNR (1 หมวด) ตามลำดับ

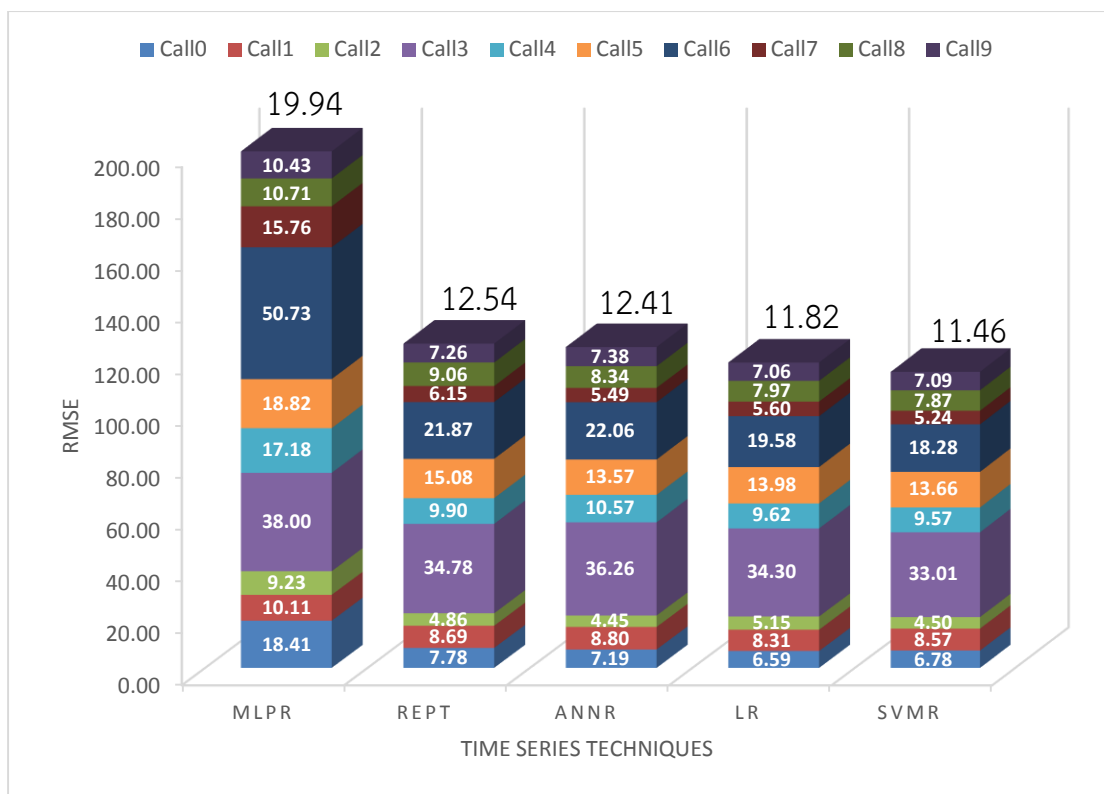
จากค่าเฉลี่ย RMSE ของแบบจำลองในการพยากรณ์การใช้หนังสือในแต่ละหมวด 12 รอบ จึงได้นำมาหาค่าเฉลี่ย (Average : AVG) ของ RMSE ของแต่ละหมวดหนังสือ และค่าความผันผวน (Standard Deviation : SD.) อีกรอบ ดังตารางที่ 4.2

ตารางที่ 4.2 การเปรียบเทียบค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สองของแบบจำลองในการพยากรณ์

| หมวด | MLPR  |        | ANNR  |        | SVMR  |        | LR    |       | REPT  |       |
|------|-------|--------|-------|--------|-------|--------|-------|-------|-------|-------|
|      | AVG   | SD.    | AVG   | SD.    | AVG   | SD.    | AVG   | SD.   | AVG   | SD.   |
| 000  | 18.41 | ±9.36  | 7.19  | ±1.19  | 6.78  | ±2.39  | 6.59  | ±1.29 | 7.78  | ±1.25 |
| 100  | 10.11 | ±3.17  | 8.80  | ±4.03  | 8.57  | ±4.62  | 8.31  | ±3.41 | 8.69  | ±2.83 |
| 200  | 9.23  | ±4.64  | 4.45  | ±2.23  | 4.50  | ±2.64  | 5.15  | ±1.77 | 4.86  | ±2.49 |
| 300  | 38.00 | ±9.23  | 36.26 | ±10.42 | 33.01 | ±12.06 | 34.30 | ±9.51 | 34.78 | ±5.59 |
| 400  | 17.18 | ±8.96  | 10.57 | ±3.16  | 9.57  | ±3.06  | 9.62  | ±1.71 | 9.90  | ±1.97 |
| 500  | 18.82 | ±5.06  | 13.57 | ±4.20  | 13.66 | ±5.80  | 13.98 | ±3.59 | 15.08 | ±3.58 |
| 600  | 50.73 | ±18.25 | 22.06 | ±6.40  | 18.28 | ±7.26  | 19.58 | ±5.24 | 21.87 | ±6.92 |
| 700  | 15.76 | ±5.39  | 5.49  | ±1.36  | 5.24  | ±1.91  | 5.60  | ±1.24 | 6.15  | ±1.38 |
| 800  | 10.71 | ±2.91  | 8.34  | ±2.93  | 7.87  | ±3.90  | 7.97  | ±3.19 | 9.06  | ±2.95 |
| 900  | 10.43 | ±3.27  | 7.38  | ±2.96  | 7.09  | ±2.40  | 7.06  | ±1.91 | 7.26  | ±1.46 |
| AVG  | 19.94 | ±4.45  | 12.41 | ±2.60  | 11.46 | ±2.96  | 11.82 | ±2.40 | 12.54 | ±1.78 |

จากตารางที่ 4.2 แสดงให้เห็นค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สองของแบบจำลองในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด ซึ่งสามารถแสดงได้ดัง ภาพประกอบที่ 4.32



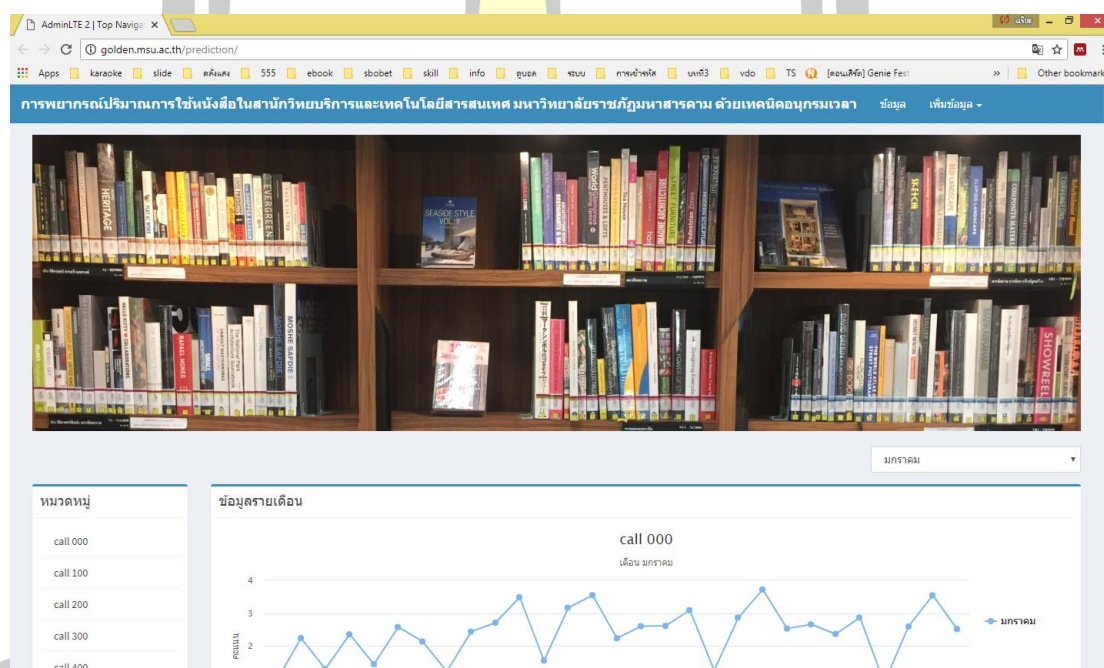


ภาพประกอบที่ 4.32 แสดงค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ของแบบจำลอง ในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด

จากภาพประกอบที่ 4.32 พบว่า ค่าเฉลี่ยค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ของแบบจำลอง ในการพยากรณ์การใช้หนังสือทั้ง 10 หมวด พบว่า ค่าความคลาดเคลื่อนของแบบจำลองในการ พยากรณ์ที่ค่าต่ำสุด อยู่ในระดับดี ได้แก่ เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย (ค่าเฉลี่ย 11.46) แต่มีค่าเบี่ยงเบนมาตรฐาน ( $\pm 2.96$ ) ต่ำกว่าเพียงเทคนิคโครงข่ายประสาทเทียมหลายชั้น แบบถดถอย ( $\pm 4.45$ ) เท่านั้น ค่าความคลาดเคลื่อนเฉลี่ยรากที่สองรองลงมา ได้แก่ เทคนิค การวิเคราะห์การถดถอยโลจิสติก (ค่าเฉลี่ย 11.82) เทคนิคโครงข่ายประสาทเทียมแบบถดถอย (ค่าเฉลี่ย 12.41) เทคนิคการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอย อัลกอริทึมการตัดกิ่งต้นไม้ (ค่าเฉลี่ย 12.54) ตามลำดับ และความคลาดเคลื่อนของแบบจำลองในการพยากรณ์ที่ค่าสูงสุด อยู่ในระดับน้อย คือ เทคนิคโครงข่ายประสาทเทียมแบบหลายชั้นแบบถดถอย (ค่าเฉลี่ย 19.94)

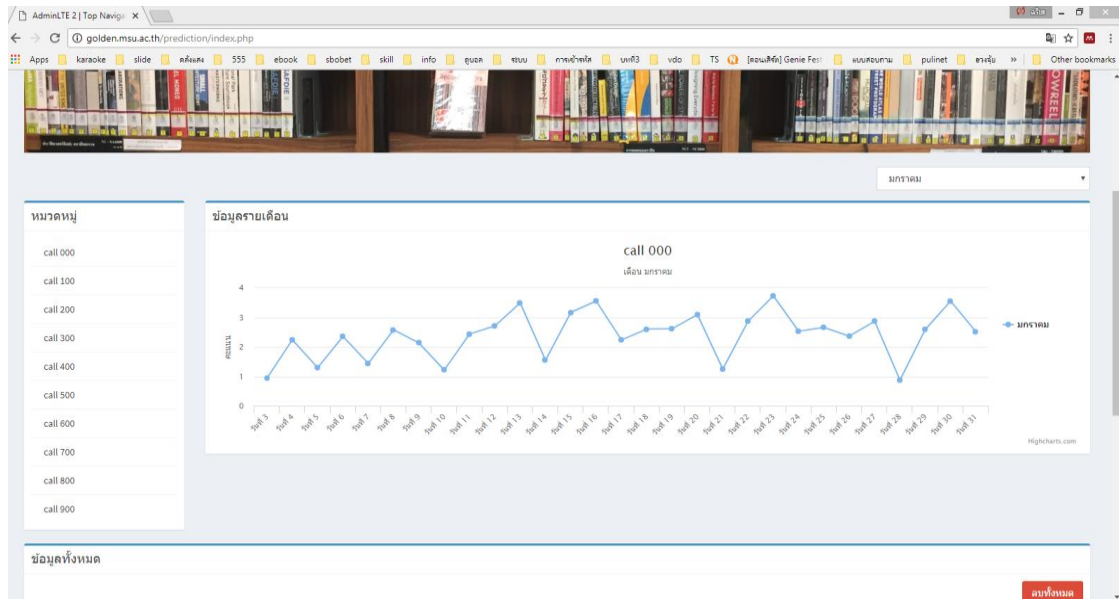
### 4.3 เว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ

จากการพยากรณ์การใช้หนังสือ ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม พบว่า เทคนิคสนับสนุนเส้นสมมุติแบบถดถอยมีค่าการพยากรณ์ปริมาณการใช้หนังสือที่ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด และเมื่อนำมาวัดค่าประสิทธิภาพด้วยค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองแล้ว มีค่าเข้าใกล้ 0 มากที่สุด จึงได้นำแบบจำลองที่สร้างด้วยเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย มาสร้างแบบจำลองเพื่อพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม แล้วนำผลที่ได้มาเป็นข้อมูลในการพัฒนาเว็บไซต์ เพื่อแสดงผลการพยากรณ์การใช้หนังสือ ดังภาพประกอบที่ 4.33



ภาพประกอบที่ 4.33 เว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ ของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

ภาพประกอบที่ 4.33 เป็นหน้าจอเว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ ผ่านทางเว็บไซต์ URL [HTTP://golden.msu.ac.th/prediction/index.php](http://golden.msu.ac.th/prediction/index.php) สามารถแสดงผลการพยากรณ์การใช้หนังสือเป็นกราฟ แบบรายเดือนแยกตามหมวดได้ และแสดงผลแบบรายปีได้ ดังภาพประกอบที่ 4.34 – 4.35 ตามลำดับ



ภาพประกอบที่ 4.34 การแสดงผลการพยากรณ์การใช้หนังสือเป็นกราฟ แบบรายเดือนแยกตามหมวด

call 900

ข้อมูลทั้งหมด

Show 10 entries

| ปี         | call 0 | call 1 | call 2  | call 3  | call 4  | call 5  | call 6  | call 7  | call 8 | call 9  |
|------------|--------|--------|---------|---------|---------|---------|---------|---------|--------|---------|
| 2018-01-03 | 0.9398 | 3.1657 | -0.1893 | 30.6331 | 6.9176  | 6.3462  | -3.8042 | -0.232  | 4.5776 | 2.9608  |
| 2018-01-04 | 2.2349 | 5.2713 | -0.0246 | 32.5883 | 11.9639 | 14.8656 | 10.1548 | 1.5916  | 5.9962 | -0.7487 |
| 2018-01-05 | 1.2901 | 3.4278 | 1.4317  | 39.655  | 12.4581 | 13.4368 | 10.2421 | 1.3923  | 6.4252 | -0.3581 |
| 2018-01-06 | 2.3484 | 6.3908 | 1.9193  | 57.1299 | 11.324  | 12.3868 | 13.1005 | -0.5805 | 2.4463 | -1.0209 |
| 2018-01-07 | 1.4388 | 2.7922 | 0.6795  | 33.0878 | 10.0583 | 10.42   | 2.0582  | 2.7293  | 4.1979 | -0.5116 |
| 2018-01-08 | 2.5715 | 1.4069 | 0.7372  | 37.8649 | 9.0781  | 10.5555 | 11.9723 | 2.5302  | 5.1402 | 0.0226  |
| 2018-01-09 | 2.1302 | 2.078  | 1.0518  | 43.2904 | 9.0834  | 13.7947 | 15.694  | 3.2263  | 4.917  | -0.1668 |
| 2018-01-10 | 1.2122 | 1.5983 | 0.3121  | 35.2692 | 7.5963  | 9.2822  | 5.5978  | 2.1554  | 5.1785 | -0.3595 |
| 2018-01-11 | 2.4323 | 1.4118 | 0.3615  | 34.8156 | 7.3863  | 12.8884 | 10.4609 | 2.3782  | 5.493  | -1.2189 |
| 2018-01-12 | 2.7069 | 2.6943 | 1.1708  | 38.2712 | 7.4028  | 12.3643 | 6.8883  | 1.8728  | 3.786  | -1.7154 |

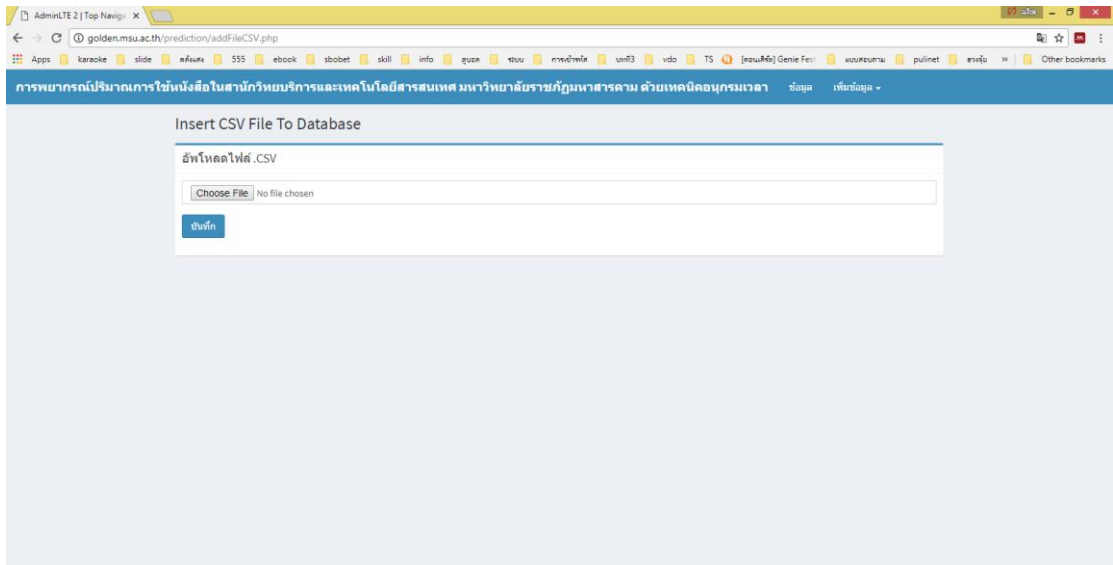
Showing 1 to 10 of 343 entries

Previous 1 2 3 4 5 ... 35 Next

ภาพประกอบที่ 4.35 การแสดงผลแบบรายปี

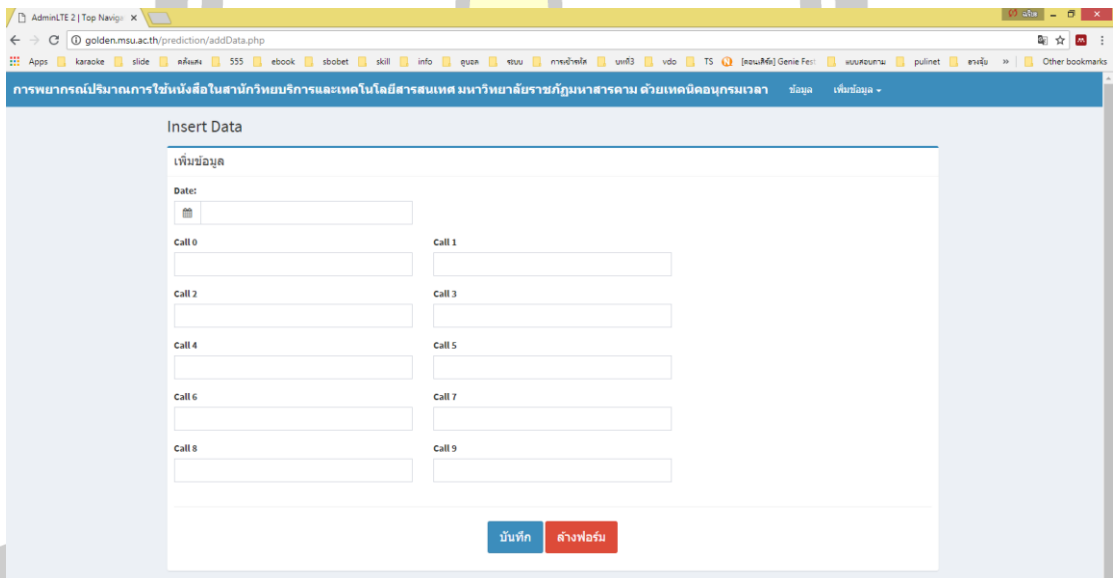
การเพิ่มข้อมูลเพื่อแสดงผลการพยากรณ์ สามารถเพิ่มได้ 2 วิธีโดยการ คือ การเพิ่มข้อมูลด้วยการเพิ่มไฟล์นามสกุล CSV และการพิมพ์ข้อมูล ดังภาพประกอบที่ 4.36 – 4.37



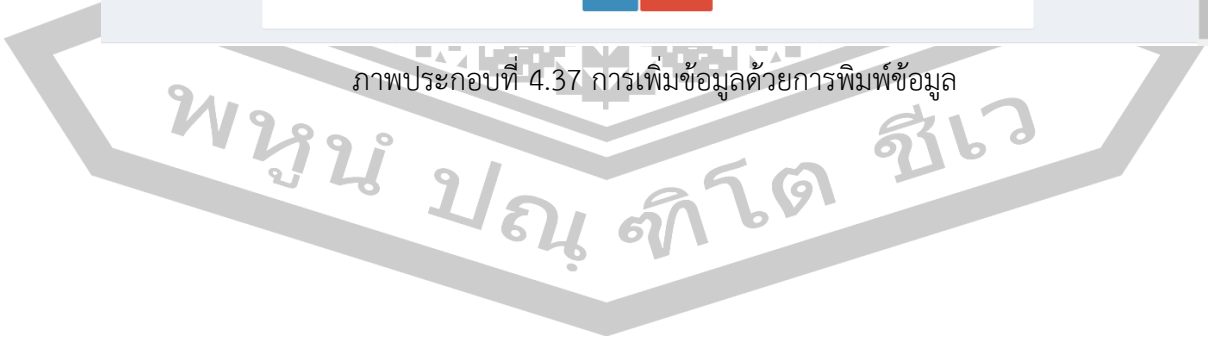


Copyright © 2014-2016 Almsaeed Studio. All rights reserved. Version 2.4.0

ภาพประกอบที่ 4.36 การเพิ่มข้อมูลด้วยการเพิ่มไฟล์นามสกุล CSV



ภาพประกอบที่ 4.37 การเพิ่มข้อมูลด้วยการพิมพ์ข้อมูล



## บทที่ 5

### วิจารณ์ สรุป และข้อเสนอแนะ

การศึกษาในครั้งนี้ มีวัตถุประสงค์ดังนี้

1. เพื่อพัฒนาแบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
  2. เพื่อพัฒนาเว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ เพื่อเป็นแนวทางในการจัดหาหนังสือมาให้บริการของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม
- สามารถสรุปผลการศึกษา อภิปรายผล รวมทั้งข้อเสนอแนะแนวทางในการวิจัย ได้ดังต่อไปนี้

#### 5.1 สรุปผล

5.1.1 การพัฒนาแบบจำลองการพยากรณ์การใช้หนังสือครั้งนี้ ใช้โปรแกรม WEKA เวอร์ชัน 3.9.2 มาใช้สร้างแบบจำลอง ด้วยการนำเทคนิคแบบต่างๆ ที่มีประสิทธิภาพสามารถวิเคราะห์ข้อมูล การถดถอยเชิงเส้นได้มาสร้างแบบจำลอง ประกอบด้วย เทคนิคโครงข่ายประสาทเทียมหลายชั้น แบบถดถอย เทคนิคโครงข่ายประสาทเทียมแบบถดถอย เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย เทคนิคการวิเคราะห์การถดถอยโลจิสติก และการเรียนรู้ต้นไม้ตัดสินใจแบบถดถอยอัลกอริทึมการตัดกิ่งต้นไม้ เนื่องจากข้อมูลการใช้หนังสือที่นำมาวิเคราะห์นั้น เมื่อนำมาแสดงในรูปแบบของกราฟเส้นจะเห็นได้ว่า กราฟเส้นมีลักษณะการเคลื่อนไหวของอนุกรมเวลาแบบการผันแปรเป็น การเปลี่ยนแปลงตามฤดูกาล (Seasonal variation) ภาพประกอบที่ 3.2

จากผลการวิจัย พบว่า แบบจำลองการพยากรณ์การใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ ที่สร้างด้วยการนำเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย มีผลการพยากรณ์ปริมาณการใช้หนังสือที่ใกล้เคียงกับปริมาณการใช้หนังสือจริง เมื่อพิจารณาตามจำนวนหมวด (7 หมวด) มากที่สุด ดังภาพประกอบที่ 4.4 – 4.10

เมื่อนำมาวัดค่าประสิทธิภาพ พบว่า แบบจำลองที่สร้างด้วยการนำเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย มีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ที่เข้าใกล้ 0 มากที่สุด เมื่อพิจารณาตามจำนวนหมวด (7 หมวด) มากที่สุด ดังภาพประกอบที่ 4.14 – 4.20 และมีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์รวมทุกหมวดมีค่า (9.42) น้อยที่สุด ดังภาพประกอบที่ 4.21

ผลการเปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยราก พบว่า แบบจำลองที่สร้างด้วยเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย มีค่าความคลาดเคลื่อนเฉลี่ยรากที่สองที่เข้าใกล้ 0 มากที่สุด

เมื่อพิจารณาตามจำนวนหมวด (6 หมวด) มากที่สุด ดังภาพประกอบที่ 4.25 – 4.30 และมีค่าเฉลี่ยของค่าความคลาดเคลื่อนเฉลี่ยรากที่สองรวมทุกหมวดมีค่า (11.46) น้อยที่สุด ดังภาพประกอบที่ 4.32

จากผลการเปรียบเทียบประสิทธิภาพที่ได้กล่าวข้างต้นนี้ แสดงให้เห็นว่าข้อมูลการใช้หนังสือที่นำมาวิเคราะห์นั้น เป็นข้อมูลอนุกรมเวลาแบบการผันแปรเป็น การเปลี่ยนแปลงตามฤดูกาล (Seasonal variation) โดยแนวโน้มของการพยากรณ์ส่วนใหญ่มีทิศทางไปทางเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย ซึ่งสอดคล้องกับผลงานวิจัยของ Fan Zhang et al. [3] และงานวิจัยของ Ervin Ceperic and Vladimir Ceperic [25] โดยผลที่ได้จากการพยากรณ์ เมื่อเปรียบเทียบกับอัลกอริทึมอื่น เทคนิคสนับสนุนเส้นสมมุติแบบถดถอยมีค่าการพยากรณ์ปริมาณการใช้หนังสือที่ใกล้เคียงกับปริมาณการใช้หนังสือจริงมากที่สุด และเมื่อนำมาวัดค่าประสิทธิภาพ ด้วยค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองแล้ว เทคนิคดังกล่าว มีจำนวนหมวดที่มีค่าเข้าใกล้ 0 จำนวนมากที่สุด รวมถึงมีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ยรากที่สองรวมทุกหมวดมีค่าเข้าใกล้ 0 มากที่สุด

5.1.2 จากข้อสรุปขั้นต้น จึงได้นำข้อมูลการใช้หนังสือของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม มาวิเคราะห์ด้วยแบบจำลองที่สร้างด้วยเทคนิคสนับสนุนเส้นสมมุติแบบถดถอย แล้วนำผลที่ได้มาเป็นข้อมูลในการพัฒนาเว็บไซต์แสดงผลการพยากรณ์การใช้หนังสือ ดังภาพประกอบที่ 4.33 – 4.37 เพื่อเป็นแนวทางในการจัดหาหนังสือมาให้บริการของสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

## 5.2 ข้อเสนอแนะ

งานวิจัยนี้จะก่อให้เกิดประโยชน์สูงสุด เมื่อนำผลการวิจัยไปใช้เป็นส่วนหนึ่งในการกำหนดนโยบายการจัดการทรัพยากรสารสนเทศ โดยการพยากรณ์การใช้หนังสือของห้องสมุด ซึ่งถือเป็นข้อมูลที่สำคัญในการตัดสินใจของผู้บริหารห้องสมุด

การวิจัยในครั้งต่อไป ควรมีการทดลองปรับค่าพารามิเตอร์ในการสร้างแบบจำลอง และควรนำการพยากรณ์การใช้หนังสือไปใช้กับการจัดหมวดหมู่หนังสืออื่น เช่น ระบบหอสมุดรัฐสภาอเมริกัน รวมถึงการสร้างระบบที่สามารถพยากรณ์การใช้หนังสือ จากประวัติการใช้หนังสือของระบบห้องสมุดอัตโนมัติได้เองเพื่อเพิ่มความสะดวกในการใช้งาน

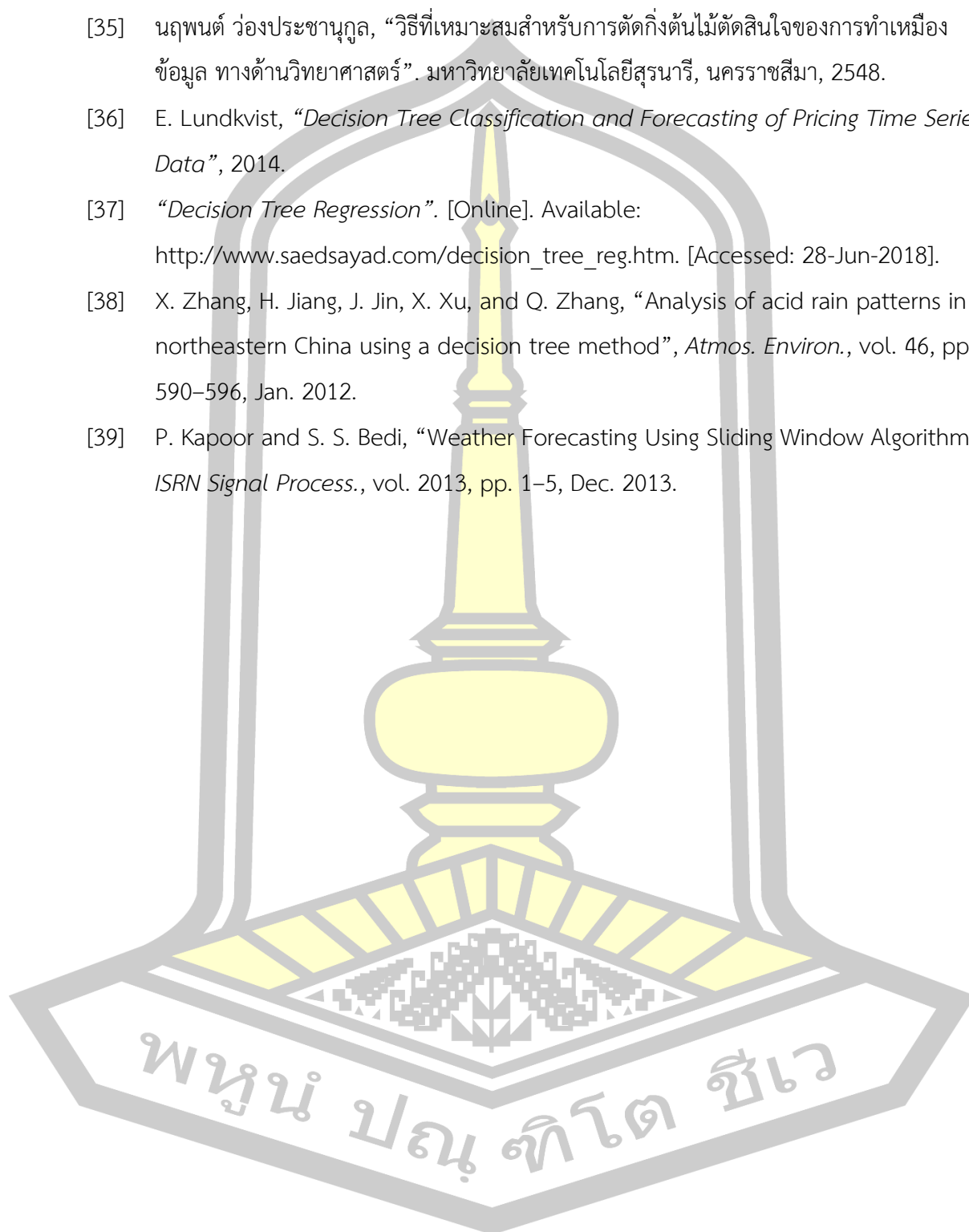
## บรรณานุกรม

- [1] M. A. Peña, A. Brenning, and R. Liao, "Classifying fruit-tree crops by Landsat-8 time series", *IEEE International Symposium of Geoscience and Remote Sensing (GRSS-CHILE)*, 2017, pp. 15–16, 2017.
- [2] J.-M. Igor Aizenberg, Leonid Sheremetov, LuisVilla-Vargas, "Multilayer Neural Network with Multi-Valued Neurons in time series forecasting of oil production", *Neurocomputing*, vol. 175, pp. 980–989, Jan, 2016.
- [3] K. W. S. Fan Zhang, Chirag Deb, Siew Eang Lee, Junjing Yang, "Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique", *Energy Build.*, vol. 126, pp. 94–103, Aug, 2016.
- [4] E. Baidoo and J. Priestley, "An Analysis of Accuracy using Logistic Regression and Time Series", *Grey Lit. From PhD Candidates*, Jan, 2016.
- [5] K. K. W. Y. Geoffrey K.F. Tso, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks", *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep, 2007.
- [6] L. Siguenza-Guzman, V. Saquicela, E. Avila-Ordóñez, J. Vandewalle, and D. Cattrysse, "Review: Literature Review of Data Mining Applications in Academic Libraries", *J. Acad. Librariansh.*, vol. 41, pp. 499–510, Jul, 2015.
- [7] สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม, "ประวัติความเป็นมา สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม", 2558. [Online]. Available: [http://arcm.rmu.ac.th/arit2/?page\\_id=440](http://arcm.rmu.ac.th/arit2/?page_id=440). [Accessed: 02-Jun-2018].
- [8] อัมพร ทีชะระ, "คู่มือการจัดหมู่ระบบหอสมุดรัฐสภาอเมริกัน". ภาควิชาบรรณารักษศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2535.
- [9] รุ่งเรือง สิทธิจันทร์, น้ำลีน เทียมแก้วสมจิต รัตนสวนจิก, "การพัฒนากการจัดหนังสือชั้น สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม". มหาวิทยาลัยมหาสารคาม, มหาสารคาม, 2556
- [10] พวา พันธุ์เมฆา, "ดีดีซี 22 : การแบ่งหมู่หนังสือและแผนการแบ่งหมู่ ระบบทศนิยมของดีวี จากต้นฉบับพิมพ์ครั้งที่ 22". ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2551.

- [11] biografiasyvida, “Biografia de Melvil Dewey”, 2018. [Online]. Available: [www.biografiasyvidas.com/biografia/d/dewey\\_melvil.htm](http://www.biografiasyvidas.com/biografia/d/dewey_melvil.htm). [Accessed: 28-Jun-2018].
- [12] วราพร งามสุข, “การพยากรณ์อนุกรมเวลาโดยการเปรียบเทียบวิธีแบบฉบับและวิธีบ็อกซ์-เจนกินส์กรณีศึกษาจนวนการเกิดอุบัติเหตุในประเทศไทย”. มหาวิทยาลัยบูรพา, ชลบุรี, 2555.
- [13] มังกร เจนชัย, “การพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมา”. มหาวิทยาลัยมหาสารคาม, มหาสารคาม, 2560.
- [14] มังกร เจนชัย, “การพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณน้ำท่าเฉลี่ยของจังหวัดนครราชสีมา”, มหาวิทยาลัยมหาสารคาม, 2560.
- [15] J. Han, M. Kamber, and J. (Computer scientist) Pei, *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann, 2012.
- [16] อนันต์ ปินะเต, “การใช้เทคนิคเหมืองข้อมูลในการเลือกกลุ่มสาขาวิชาที่เหมาะสมสำหรับการศึกษาระดับปริญญาตรี”. มหาวิทยาลัยมหาสารคาม, มหาสารคาม, 2557.
- [17] บรรจบ ตลกกุล, “แบบจำลองเพื่อพยากรณ์การเกิดแผลที่เท้าของผู้ป่วยโรคเบาหวานโดยเทคนิคเหมืองข้อมูล”. มหาวิทยาลัยมหาสารคาม, มหาสารคาม, 2557.
- [18] ชัชชฎา วันดี, “การศึกษาปัจจัยที่มีผลต่อการเลือกอาชีพของนิสิตระดับปริญญาตรีหลังสำเร็จการศึกษาโดยใช้เทคนิคเหมืองข้อมูล”. มหาวิทยาลัยมหาสารคาม, มหาสารคาม, 2557.
- [19] E. Bahadir, “Using Neural Network and Logistic Regression Analysis to Predict Prospective Mathematics Teachers Academic Success upon Entering Graduate Education”, *Educ. Sci. Theory Pract.*, vol. 16, no. 3, pp. 943–964, Jun. 2016.
- [20] เดช ธรรมศิริพยุง มีสีจ, “การรวมกลุ่มจำแนกข้อมูลด้วยเทคนิคต้นไม้ช่วยตัดสินใจ เทคนิคโครงข่ายประสาทเทียมและเทคนิคซัพพอร์ตเวกเตอร์แมชชีน”, *การประชุมวิชาการระดับประเทศทางด้านเทคโนโลยีสารสนเทศ*, หน้า 13–17, 2553.
- [21] N. Kourentzes, D. K. Barrow, and S. F. Crone, “Neural network ensemble operators for time series forecasting”, *Expert Syst. Appl.*, vol. 41, pp. 4235–4244, Jul. 2014.
- [22] G. P. Zhang and M. Qi, “Neural network forecasting for seasonal and trend time series”, *Eur. J. Oper. Res.*, vol. 160, no. 2, pp. 501–514, Jan. 2005.
- [23] B. Lenz and B. Barak, “Data Mining and Support Vector Regression Machine Learning in Semiconductor Manufacturing to Improve Virtual Metrology”, *2013 46th Hawaii International Conference on System Sciences*. pp. 3447–3456, 2013.

- [24] นรินทร์ พนาวาสนิเวส จิระวิชิตชัย, “การจำแนกมะเร็งเม็ดเลือดขาวโดยใช้เทคนิคการลดมิติข้อมูลด้วย Chi-square”, *การประชุมวิชาการทางคอมพิวเตอร์และเทคโนโลยีสารสนเทศ 2553 ครั้งที่ 3*, pp. 7–12, 2553.
- [25] E. Ceperic, V. Ceperic, and A. Baric, “A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines”, *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4356–4364, Nov. 2013.
- [26] E. Bozpolat, “Investigation of the Self-Regulated Learning Strategies of Students from the Faculty of Education Using Ordinal Logistic Regression Analysis.”, *Educ. Sci. Theory Pract.*, vol. 17, no. 1, pp. 301–318, Feb. 2017.
- [27] “Exam pass logistic curve.jpeg - Wikimedia Commons”. [Online]. Available: [https://commons.wikimedia.org/wiki/File:Exam\\_pass\\_logistic\\_curve.jpeg](https://commons.wikimedia.org/wiki/File:Exam_pass_logistic_curve.jpeg). [Accessed: 28-Jun-2018].
- [28] M. Abdel-Aty, N. Uddin, A. Pande, F. Abdalla, and L. Hsia, “Predicting Freeway Crashes from Loop Detector Data by Matched Case-Control Logistic Regression”, *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1897, pp. 88–95, Jan. 2004.
- [29] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar, and V. A. Kanade, “Analysis of customer churn prediction in telecom industry using decision trees and logistic regression”, *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*. pp. 1–4, 2016.
- [30] L. Rokach and O. Z. Maimon, “*Data mining with decision trees : theory and applications*”. Singapore ; Hackensack, 2008.
- [31] J. R. Quinlan and J. R., “*Induction of Decision Trees*”, *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [32] กาญจน์ ณ ศรีระ, “การวิเคราะห์ความทนต่อข้อมูลรบกวนของอัลกอริทึมแบบรวมกลุ่ม”. *มหาวิทยาลัยมหาสารคาม*, มหาสารคาม, 2557.
- [33] อมรภัทร์ หาญโคกกรวด, “แบบจำลองทำนายความเสี่ยงการเกิดอุบัติเหตุบนท้องถนนในจังหวัดสกลนครโดยเทคนิคเหมือง”. *มหาวิทยาลัยมหาสารคาม*, มหาสารคาม, 2556.
- [34] S. Krishnaveni and M. Hemalatha, “A Perspective Analysis of Traffic Accident using Data Mining Techniques”, *Int. J. Comput. Appl.*, vol. 23, no. 7, pp. 40–48, Jun. 2011.

- [35] นฤพนธ์ ว่องประชาณุกุล, “วิธีที่เหมาะสมสำหรับการตัดกิ่งต้นไม้ตัดสินใจของการทำเหมืองข้อมูล ทางด้านวิทยาศาสตร์”. มหาวิทยาลัยเทคโนโลยีสุรนารี, นครราชสีมา, 2548.
- [36] E. Lundkvist, “*Decision Tree Classification and Forecasting of Pricing Time Series Data*”, 2014.
- [37] “*Decision Tree Regression*”. [Online]. Available: [http://www.saedsayad.com/decision\\_tree\\_reg.htm](http://www.saedsayad.com/decision_tree_reg.htm). [Accessed: 28-Jun-2018].
- [38] X. Zhang, H. Jiang, J. Jin, X. Xu, and Q. Zhang, “Analysis of acid rain patterns in northeastern China using a decision tree method”, *Atmos. Environ.*, vol. 46, pp. 590–596, Jan. 2012.
- [39] P. Kapoor and S. S. Bedi, “Weather Forecasting Using Sliding Window Algorithm”, *ISRN Signal Process.*, vol. 2013, pp. 1–5, Dec. 2013.



## ประวัติผู้เขียน

|                      |   |
|----------------------|---|
| ชื่อ                 | นายฉริยะ อัครวรรณ   |
| วันเกิด              | วันที่ 1 พฤศจิกายน พ.ศ. 2520  |
| สถานที่เกิด          | อำเภอเมือง จังหวัดมหาสารคาม   |
| สถานที่อยู่ปัจจุบัน  | สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม<br>ตำบลขามเรียง อำเภอกันทรวิชัย<br>จังหวัดมหาสารคาม รหัสไปรษณีย์ 44150   |
| ตำแหน่งหน้าที่การงาน | นักวิชาการคอมพิวเตอร์ปฏิบัติการ   |
| สถานที่ทำงานปัจจุบัน | สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม ตำบลขามเรียง<br>อำเภอกันทรวิชัย จังหวัดมหาสารคาม รหัสไปรษณีย์ 44150  |
| ประวัติการศึกษา      | พ.ศ. 2539 มัธยมศึกษาตอนปลาย โรงเรียนกาฬสินธุ์พิทยาสรรพ์<br>ตำบลกาฬสินธุ์ อำเภอเมืองกาฬสินธุ์ จังหวัดกาฬสินธุ์<br>พ.ศ. 2541 ประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.)<br>สถาบันเทคโนโลยีราชมงคลวิทยาเขตกาฬสินธุ์ ตำบลกาฬสินธุ์<br>อำเภอเมืองกาฬสินธุ์ จังหวัดกาฬสินธุ์<br>พ.ศ. 2543 ปริญญาบริหารธุรกิจบัณฑิต (บธ.บ.) สาขาคอมพิวเตอร์ธุรกิจ<br>คณะบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม<br>พ.ศ. 2561 ปริญญาวิทยาศาสตรมหาบัณฑิต (วท.ม.)<br>สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ<br>มหาวิทยาลัยมหาสารคาม |

พจนัน ปณุกิตโต ชีวะ