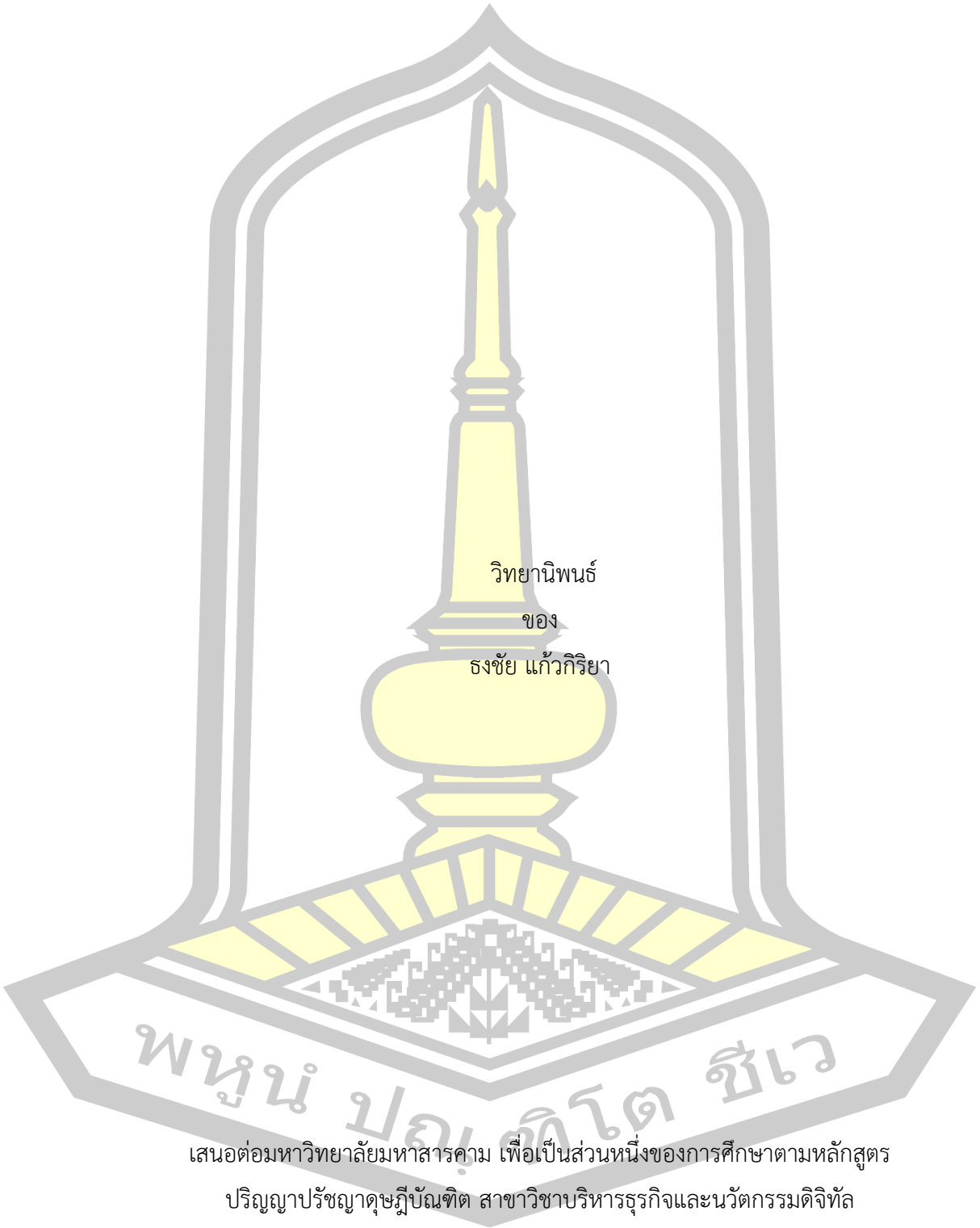
การพัฒนาแบบจำลองพยากรณ์ลูกค้าในการลงทุนด้านกองทุนรวมโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม

วิทยานิพนธ์
ของ
ธงชัย แก้วกิริยา

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล
ธันวาคม 2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การพัฒนาแบบจำลองพยากรณ์ลูกค้าในการลงทุนด้านกองทุนรวมโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม



วิทยานิพนธ์

ของ

ธงชัย แก้วกิริยา

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล

ธันวาคม 2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Development of Customer Predictive Model for Mutual Fund Investment Using Ensemble Learning Technique

Thongchai Kaewkiriya

A Thesis Submitted in Partial Fulfillment of Requirements

for Doctor of Philosophy (Business Administration and Digital Innovation)

December 2023

Copyright of Mahasarakham University

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายธงชัย แก้วกิริยา แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รศ. ดร. ธีระวัฒน์ เจริญราษฎร์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รศ. ดร. กิตติพล วิแสง)

.....กรรมการ
(ดร. การันต์ กิจระการ)

.....กรรมการ
(ดร. กาญจนา หินเธาว์)

.....กรรมการ
(ดร. พงศธร ตันตระบัณฑิตย์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา ปรัชญาดุษฎีบัณฑิต สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล ของมหาวิทยาลัยมหาสารคาม

พหุน ปญฺ ฑิโต ชีเว

....

(ดร. ชลธิชา ธรรมวิญญู)
คณบดีคณะการบัญชีและการจัดการ

(รศ. ดร. กริสน์ ชัยมูล)
คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การพัฒนาแบบจำลองพยากรณ์ลูกค้าในการลงทุนด้านกองทุนรวมโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม
ผู้วิจัย ธงชัย แก้วกิริยา
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. กิตติพล วิแสง
ปริญญา ปรัชญาดุษฎีบัณฑิต **สาขาวิชา** บริหารธุรกิจและนวัตกรรมดิจิทัล
มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2566

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1. เพื่อพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม 2. เพื่อเปรียบเทียบประสิทธิภาพอัลกอริทึมที่ใช้ในแบบจำลองการพยากรณ์ลูกค้า และ 3. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการใหม่ที่นำเสนอกับงานวิจัยที่ผ่านมาเพื่อยืนยันผลลัพธ์ในการนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน วิธีการดำเนินการวิจัยแบ่งออกเป็น 4 ขั้นตอนได้แก่ ขั้นตอนที่ 1 วิเคราะห์ปัญหาและข้อมูลในงานวิจัย ขั้นตอนที่ 2 ออกแบบกรอบแนวคิดการวิจัย ขั้นตอนที่ 3 พัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน ขั้นตอนที่ 4 ทดสอบเปรียบเทียบประสิทธิภาพการวิจัย งานวิจัยนี้ใช้ชุดข้อมูลจำนวน 19,577 ชุดข้อมูล

ผลของการวิจัยสรุปได้ดังนี้ 1. ผลการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม ประกอบด้วย 3 ส่วนคือ 1) ส่วนการจัดการข้อมูล 2) ส่วนของการพัฒนาแบบจำลองพยากรณ์และเปรียบเทียบประสิทธิภาพ 3) ส่วนการพัฒนาแบบจำลองพยากรณ์ด้วยเทคนิคการเรียนรู้แบบกลุ่มและเปรียบเทียบ 2. ผลการเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึม พบว่า อัลกอริทึมเคเนียร์เรสเนเบอร์ จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K 5, 7, และ 9 ตามลำดับ โดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR (LTF/RMF) โดยมีค่าความแม่นยำอยู่ที่ 97.46%. อัลกอริทึมเคเนียร์เรสเนเบอร์ จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K เท่ากับ 7 โดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF โดยมีค่าความแม่นยำอยู่ที่ 83.21% อัลกอริทึมเคเนียร์เรสเนเบอร์ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K เท่ากับ 3 มีค่าความแม่นยำอยู่ที่ 90.94% อัลกอริทึมนาอีฟเบย์จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 97.23% อัลกอริทึมนาอีฟเบย์จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุน

ประเภท MF (Mutual Fund) มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 83.62% อัลกอริทึมนาอีฟเบย์สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำอยู่ที่ 76.09% อัลกอริทึมต้นไม้ตัดสินใจ จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 97.46% อัลกอริทึมต้นไม้ตัดสินใจ จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 83.62% อัลกอริทึมต้นไม้ตัดสินใจ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำอยู่ที่ 93.28% อัลกอริทึมกฎการอุปนัย จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 97.46% อัลกอริทึมกฎการอุปนัย จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 83.62% อัลกอริทึมกฎการอุปนัยสำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำอยู่ที่ 90.52% อัลกอริทึมโครงข่ายประสาทเทียม จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 97.46% อัลกอริทึมโครงข่ายประสาทเทียม จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 83.62% อัลกอริทึมโครงข่ายประสาทเทียมสำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำอยู่ที่ 93.43% อัลกอริทึมการเรียนรู้แบบกลุ่ม จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 97.46% อัลกอริทึมการเรียนรู้แบบกลุ่ม จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF มากที่สุดโดยมีค่าความแม่นยำอยู่ที่ 83.62% อัลกอริทึมการเรียนรู้แบบกลุ่ม สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำอยู่ที่ 92.38% 3. ผลการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ที่นำเสนอกับงานวิจัยที่ผ่านมาโดยพิจารณาจากค่าความแม่นยำของโมเดลพบว่าวิธีการใหม่ที่นำเสนอมีประสิทธิภาพดีกว่า

คำสำคัญ : พยากรณ์ลูกค้า, กองทุมรวม, การเรียนรู้แบบกลุ่ม, การเรียนรู้ของเครื่อง, การลงทุน

TITLE Development of Customer Predictive Model for Mutual Fund Investment Using Ensemble Learning Technique

AUTHOR Thongchai Kaewkiriya

ADVISORS Associate Professor Kittipol Wisaeng , Ph.D.

DEGREE Doctor of Philosophy **MAJOR** Business Administration and Digital Innovation

UNIVERSITY Mahasarakham University **YEAR** 2023

ABSTRACT

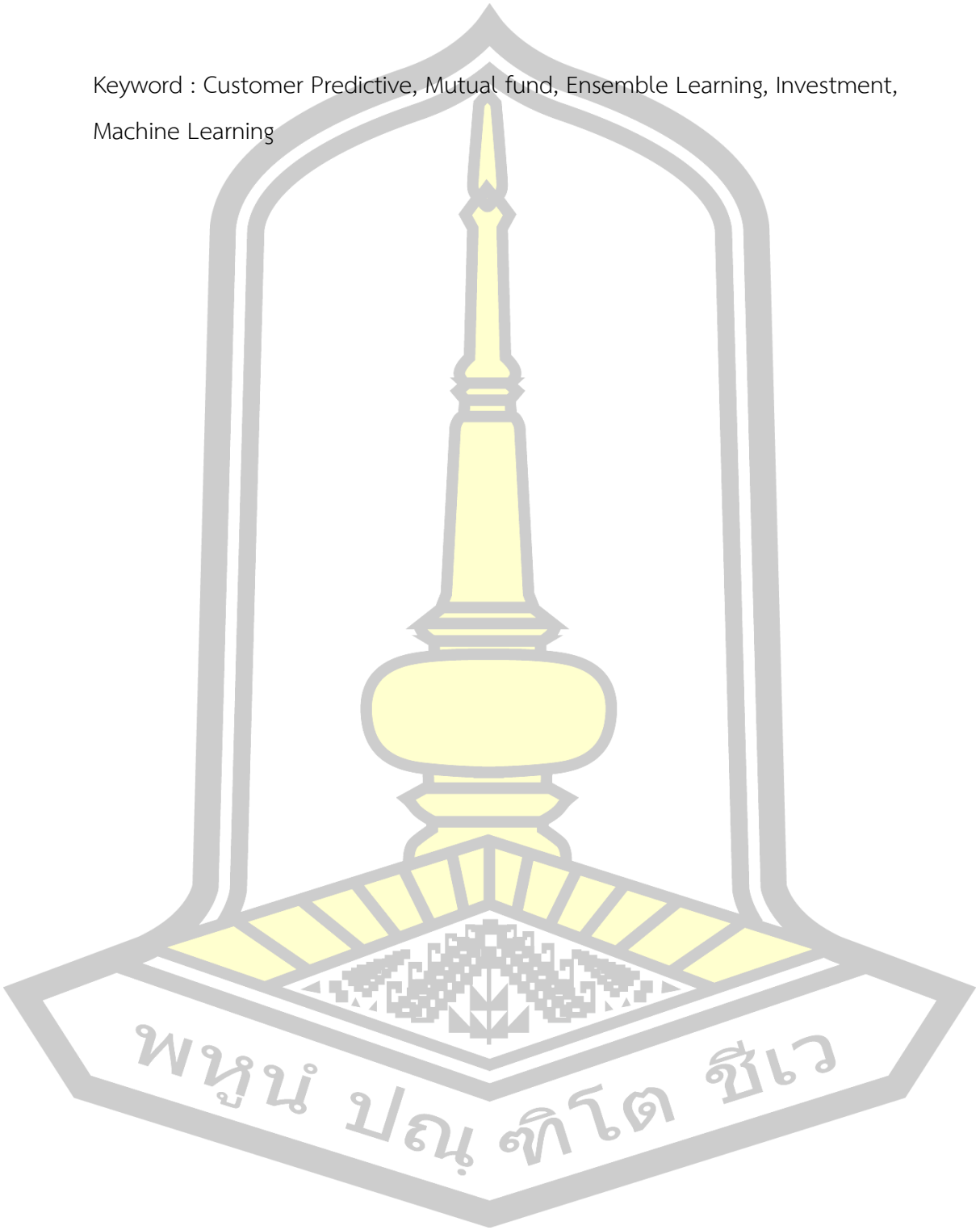
The purpose of this research is to: 1. Develop a customer investment prediction model using group-based learning techniques. 2. Compare the effectiveness of algorithms used in the customer prediction model. 3. Compare the performance of the newly proposed method with previous research to validate the results for the development of the customer investment prediction model. The research methodology consists of 4 steps: Step 1 Analyzing the problem and research data. Step 2 Designing the research framework and concepts. Step 3 Develop the customer investment prediction model. Step 4 Testing and comparing the research performance. This research utilizes a dataset consisting of 19,577 data points.

The research results can be summarized as follows: The outcomes of developing a customer investment prediction model using group-based learning techniques consist of three parts: Part 1: Data Management Part 2: Development of the prediction model and performance comparison Part 3: Development of the prediction model using group-based learning techniques and comparison. The comparison of algorithm performance reveals the following: The K-Nearest Neighbors algorithm, using data management format 1, achieves the highest accuracy when setting K to 5, 7, and 9 successively, with an accuracy rate of 97.46%. This is for predicting products in the LR fund category. The K-Nearest Neighbors algorithm, using

data management format 2, achieves the highest accuracy when setting K to 7, with an accuracy rate of 83.21%. This is for predicting products in the MF fund category. The K-Nearest Neighbors algorithm, for predicting fund types, achieves the highest accuracy when setting K to 3, with an accuracy rate of 90.94%. The Naive Bayes algorithm, using data management format 1, achieves the highest accuracy with a label of LR fund products, reaching an accuracy rate of 97.23%. The Naive Bayes algorithm, using data management format 2, achieves the highest accuracy with a label of MF fund products, reaching an accuracy rate of 83.62%. The Naive Bayes algorithm for predicting fund types achieves an accuracy rate of 76.09%. The Decision Tree algorithm, using data management format 1, achieves the highest accuracy with a label of LR fund products, reaching an accuracy rate of 97.46%. The Decision Tree algorithm, using data management format 2, achieves the highest accuracy with a label of MF fund products, also at an accuracy rate of 83.62%. The Decision Tree algorithm for predicting fund types achieves an accuracy rate of 93.28%. The Induction law algorithm, using data management format 1, achieves the highest accuracy with a label of LR fund products, at an accuracy rate of 97.46%. The Induction law algorithm, using data management format 2, achieves the highest accuracy with a label of MF fund products, at an accuracy rate of 83.62%. For predicting fund types, the Induction law algorithm achieves an accuracy rate of 90.52%. The Neural Network algorithm, using data management format 1, achieves the highest accuracy with a label of LR fund products, at an accuracy rate of 97.46%. The Neural Network algorithm, using data management format 2, achieves the highest accuracy with a label of MF fund products, at an accuracy rate of 83.62%. For predicting fund types, the Neural Network algorithm achieves an accuracy rate of 93.43%. The Ensemble Learning algorithm, using data management format 1, achieves the highest accuracy with a label of LR fund products, at an accuracy rate of 97.46%. The Ensemble Learning algorithm, using data management format 2, achieves the highest accuracy with a label of MF fund products, at an accuracy rate of 83.62%. For predicting fund types, the Ensemble Learning algorithm achieves an accuracy rate of 92.38%. 3. When comparing the performance of the newly introduced method with previous research, considering the accuracy of the models,

it was found that the newly proposed method performs better.

Keyword : Customer Predictive, Mutual fund, Ensemble Learning, Investment, Machine Learning



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จสมบูรณ์ได้ด้วยความกรุณาอนุเคราะห์จากอาจารย์รองศาสตราจารย์ ดร.ธีระวัฒน์ เจริญราษฎร์ ประธานกรรมการวิทยานิพนธ์ รองศาสตราจารย์ ดร.กิตติพล วิแสง อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก อาจารย์ ดร.การันต์ กิจระการ กรรมการ อาจารย์ ดร.ดวงรัตน์ ธารดำรง กรรมการ อาจารย์ ดร.พงศธร ตันตระบัณฑิตย์ กรรมการ อาจารย์ ดร.กาญจนา หินเธาว์ กรรมการ และบุคคลหลากหลายท่านผู้วิจัยขอขอบพระคุณผู้ที่ให้ ความกรุณาอนุเคราะห์ในด้านต่างๆ ดังนี้

ขอขอบพระคุณ รองศาสตราจารย์ ดร.กิตติพล วิแสง อาจารย์ที่ที่ปรึกษาวิทยานิพนธ์หลัก ที่ได้กรุณาอนุเคราะห์เสียสละเวลาอันมีค่าในการให้คำปรึกษาแนะนำด้านต่างๆ ให้ทั้งความรู้ ความเข้าใจ เกิดทักษะและทัศนคติที่ดีเพื่อให้วิทยานิพนธ์ฉบับนี้สำเร็จอย่างสมบูรณ์

ท้ายนี้ผู้จัดทำขอขอบพระคุณผู้มีอุปการะทุกท่าน รวมทั้งผู้ที่มีส่วนเกี่ยวข้องคอยให้กำลังใจตลอดระยะเวลาที่ผ่านมาจนสามารถทำงานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

ธงชัย แก้วกิริยา

พหูน ปณฺฑิโต ชีเว

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... ฉ

กิตติกรรมประกาศ ..... ฌ

สารบัญ ..... ญ

บทที่ 1 บทนำ 1

ภูมิหลัง 1

ความมุ่งหมายของการวิจัย 5

ความสำคัญของการวิจัย 6

กรอบแนวคิดที่ใช้ในการวิจัย 6

ขอบเขตของการวิจัย 6

ประโยชน์ที่คาดว่าจะได้รับ 7

นิยามศัพท์เฉพาะ 7

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 9

แนวคิดด้านการลงทุน 9

ความรู้เกี่ยวกับผลิตภัณฑ์กองทุน 11

การตัดสินใจลงทุน พฤติกรรม และการวิเคราะห์ทางด้านประชากรศาสตร์ 14

หลักการวิเคราะห์ข้อมูลและกระบวนการเรียนรู้ของเครื่อง (Machine learning) 18

เทคนิคการแบ่งกลุ่มข้อมูลเคมีน (K-Means) 21

เทคนิคการจำแนกข้อมูล (Classification) 23

เทคนิคการเรียนรู้แบบกลุ่ม 33

เอกสารและงานวิจัยที่เกี่ยวข้อง 36

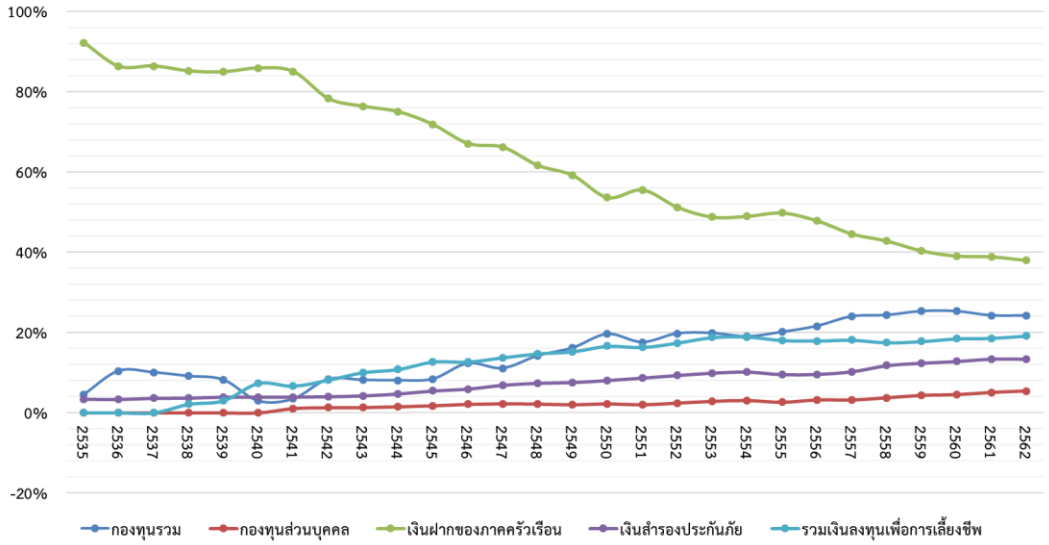
บทที่ 3 วิธีการดำเนินการวิจัย 44
วิเคราะห์ปัญหาและข้อมูลในงานวิจัย 44
ออกแบบกรอบแนวคิด 49
พัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน 55
ทดสอบเปรียบเทียบประสิทธิภาพ 57
บทที่ 4 ผลการวิเคราะห์ข้อมูล 58
ผลการจัดการข้อมูลตัวแปรที่ใช้ในการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน 58
ผลการเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึม 67
ผลการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน 92
ผลการเปรียบเทียบการจัดการข้อมูลที่ส่งผลต่อประสิทธิภาพของรูปแบบพยากรณ์ 96
ผลการเปรียบเทียบประสิทธิภาพของวิธีการกับงานวิจัยที่ผ่านมา 99
บทที่ 5 สรุปผล อภิปราย และข้อเสนอแนะ 1
ความมุ่งหมายของการวิจัย 1
สรุปผลการวิจัย 1
อภิปรายผลการวิจัย 3
ข้อเสนอแนะ 9
บรรณานุกรม 11
ประวัติผู้เขียน 18

พหุน ปณ ฑิโต ชีเว

บทที่ 1
บทนำ

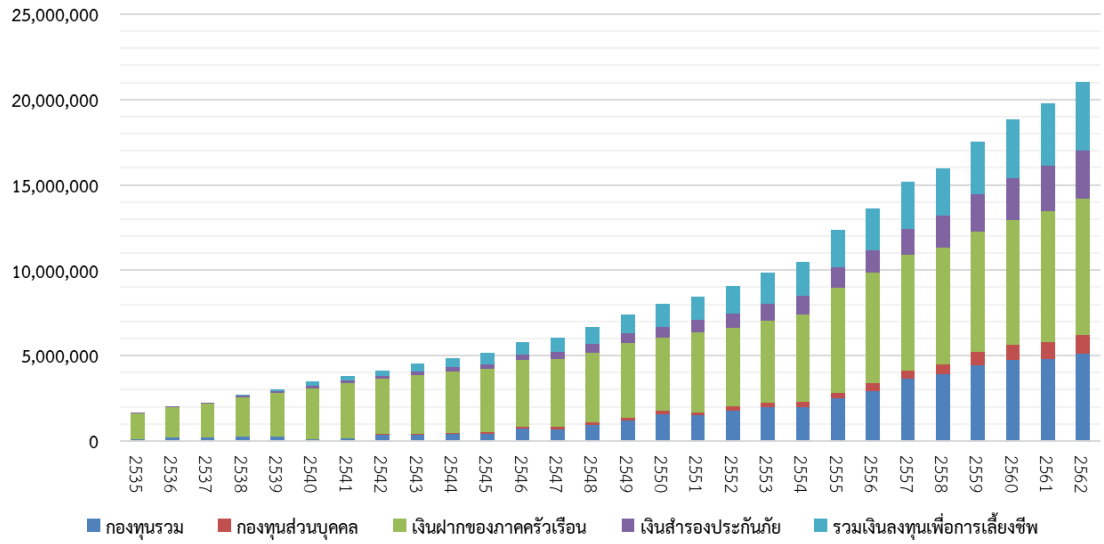
ภูมิหลัง

จากสถานการณ์สภาวะดอกเบี้ยเงินฝากของสถาบันการเงิน มีแนวโน้มลดลงอย่างต่อเนื่อง โดยส่งผลกระทบกับผู้ที่มีเงินฝากที่มีความต้องการผลตอบแทนให้เพียงพอต่อความต้องการ แต่หากกล่าวถึงการลงทุนในกองทุน กลับให้ผลตอบแทนได้มากกว่าการลงทุนในรูปแบบเงินฝาก ส่งผลให้ผู้คนหันมาให้ความสนใจการลงทุนเพิ่มมากขึ้น เนื่องจากผลตอบแทนที่สูงขึ้นหรือความต้องการจะได้รับสิทธิประโยชน์ทางภาษี เช่น การลงทุนผ่านกองทุนรวมหุ้นระยะยาว (Long Term Equity Fund: LTF) มีจุดมุ่งหมายในการส่งเสริมการลงทุนระยะยาวในตลาดหุ้น โดยผู้ลงทุนจะได้รับสิทธิประโยชน์ทางภาษีเป็นสิ่งจูงใจในการลงทุน กองทุนรวมเพื่อการเลี้ยงชีพ (Retirement Mutual Fund: RMF) กองทุนรวมที่มีจุดมุ่งหมายเพื่อส่งเสริมให้เกิดการออมเงินระยะยาวไว้สำหรับใช้จ่ายยามเกษียณอายุ กองทุนรวมตลาดเงิน (Money Market Fund: MMF) กองทุนรวมที่เน้นลงทุนในเงินฝากของธนาคาร หรือตราสารหนี้ระยะสั้น สำหรับลงทุนที่เหมาะสมสำหรับคนที่ต้องการบริหารสภาพคล่อง กองทุนที่มีกำหนดอายุโครงการ กองทุนรวมตราสารหนี้ที่มีระยะเวลา (Term Fund: TF) กองทุนที่มีกำหนดอายุโครงการที่ชัดเจนคล้ายกับการฝากเงินแบบประจำ แต่เน้นบริหารกองทุนเพื่อให้สร้างผลตอบแทนที่สูงกว่าเงินฝากประจำ และกองทุนรวม (Mutual Fund: MF) กองทุนที่นำเงินที่ระดมทุนได้ไปลงทุนในหลักทรัพย์ หรือทรัพย์สินประเภทต่างๆ ตามนโยบายการลงทุนที่ได้ระบุไว้ในหนังสือชี้ชวนเสนอขายแก่นักลงทุน กองทุนรวมมีจุดเด่นตรงที่มีเงินน้อยก็ลงทุนได้ แถมยังมีนโยบายการลงทุนที่หลากหลาย โดยรูปแบบการลงทุน ผู้ลงทุนในแต่ละรายจะได้รับเป็น “หน่วยลงทุน” เป็นหลักฐานการมีส่วนร่วมในเงินกองทุนนั้นตามสัดส่วนเงินของตนเองที่ได้ลงทุน โดยผู้ที่มีหน้าที่บริหารเงินภายในกองทุนเพื่อให้เกิดเป็นผลกำไร ได้แก่ “บริษัทหลักทรัพย์จัดการกองทุน” หรือที่เรียก โดยย่อว่า บลจ.



ภาพประกอบ 1 อัตราส่วนการออมและการลงทุนภาคครัวเรือน

จากงานศึกษาสัดส่วนการออมและการลงทุนภาคครัวเรือน ตามภาพประกอบ 1 พบว่าประชาชนมีรูปแบบการลงทุนผ่านบัญชีเงินฝากธนาคารลดลงอย่างมีนัยสำคัญ และพบว่าอัตราดอกเบี้ยส่งผลต่อการออมเงินในบัญชีออมทรัพย์ของลูกค้า และจากกราฟแสดงสัดส่วนระหว่างเงินฝากภาคครัวเรือนและการลงทุนผ่านบริษัทหลักทรัพย์จัดการกองทุน พบว่ามีสัดส่วนเพิ่งสูงขึ้น ตามภาพประกอบ 2 ข้อมูลจาก สมาคมบริษัทจัดการลงทุน กองทุนบำเหน็จบำนาญข้าราชการ สำนักงานประกันสังคม ธนาคารแห่งประเทศไทย และสมาคมประกันชีวิตไทย (2562)



ภาพประกอบ 2 สัดส่วนการออมและการลงทุนภาคครัวเรือน

จากการรวบรวมข้อมูลบริษัทที่ได้รับใบอนุญาต การจัดการกองทุนรวมประเภทบริษัทหลักทรัพย์ (สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์, 2564) จะพบว่ามีบริษัทหลักทรัพย์จัดการกองทุนจำนวนมากถึง 27 บริษัท ที่ผู้ลงทุนสามารถเลือกหรือสนใจในการสมัครเลือกซื้อกองทุน ตามตาราง 1 โดยสามารถติดต่อผ่านบริษัทหลักทรัพย์จัดการกองทุน ทั้งจากช่องทางธนาคารที่เป็นตัวแทน หรือกับบริษัทหลักทรัพย์โดยตรง แน่นอนว่าการลงทุนในกองทุนมีรูปแบบหลากหลายตามผลตอบแทนที่ต้องการ

ปัจจุบันเทคโนโลยีสารสนเทศ เข้ามามีบทบาทในหลายหน่วยธุรกิจ ตลอดจนถึงการดำเนินงานของบริษัทหลักทรัพย์จัดการกองทุน ซึ่งมีการเก็บค่าการซื้อกองทุน ตลอดจนข้อมูลลูกค้า เช่น สถิตการเลือกซื้อกองทุนผ่านใบขอซื้อหลักทรัพย์หรือที่นิยมอย่างระบบออนไลน์ ซึ่งในปัจจุบันนักลงทุนบริษัทหลักทรัพย์จัดการกองทุน มีพฤติกรรมการเลือกลงทุนผลิตภัณฑ์ที่แตกตามวัตถุประสงค์ของนักลงทุน ซึ่งจากการค้นคว้าข้อมูล และสอบถามผู้เชี่ยวชาญด้านการลงทุน พบว่ามีเป็นจำนวนมากที่เลือกประเภทการลงทุนยังไม่เหมาะสมหรือยังขาดการวางแผนที่ดี ทำให้ได้ผลตอบแทนที่ไม่คุ้มค่าเท่าที่ควร

ตาราง 1 รายชื่อบริษัทที่ได้รับใบอนุญาต การจัดการกองทุนรวมประเภทบริษัทหลักทรัพย์

	ชื่อบริษัท	
บมจ. บลจ. กรุงไทย	บลจ. ธนชาต จำกัด	บลจ. แลนด์ แอนด์ เฮ้าส์ จำกัด
บลจ. กรุงศรี จำกัด	บลจ. บัวหลวง จำกัด	บลจ. วรรณ จำกัด
บลจ. กสิกรไทย จำกัด	บลจ. บางกอกแคปปิตอล จำกัด	บลจ. วี จำกัด
บลจ. เกียรตินาคินภัทร จำกัด	บลจ. พรินซิเพิล จำกัด	บลจ. อเบอร์ดีน สแตนดาร์ด (ประเทศไทย) จำกัด
บลจ. คิง ไว (เอเชีย) จำกัด	บลจ. เพื่อผู้ลงทุนต่างด้าว จำกัด	บลจ. เอ็กซ์สปริง จำกัด
บลจ. ทหารไทย จำกัด	บลจ. ฟิลลิป จำกัด	บลจ. เอเซีย เวลท์ จำกัด
บลจ. ทาลิส จำกัด	บลจ. เมอร์ชั่น พาร์ทเนอร์ จำกัด	บมจ. บลจ. เอ็มเอฟซี
บลจ. ทิสโก้ จำกัด	บลจ. ยูโอบี (ประเทศไทย) จำกัด	บลจ. เอไอเอ (ประเทศไทย)
บลจ. ไทยพาณิชย์ จำกัด	บลจ. เรนเนสซานซ์ จำกัด	บลจ. แอสเซท พลัส จำกัด

ทั้งนี้ ข้อมูลสถิติอันเกี่ยวเนื่องกับ รูปแบบประชากร อาชีพ รายได้ ถิ่นฐาน และฐานะทางการเงินผ่านบัญชีเงินฝาก ตลอดจนความถี่ในการซื้อกองทุน ที่แตกต่างกัน ส่งผลต่อการเลือกผลิตภัณฑ์กองทุน และมีรูปแบบการลงทุนที่แตกต่างกันด้วย แม้กระทั่งพฤติกรรมของนักลงทุนที่มีความต้องการในการออมเงินเพื่อหวังเงินปันผล หรือทำกำไรในระยะสั้น ทำให้เป็นที่สนใจของนักวิจัยและนักวิเคราะห์ในการแยกประเภท และอธิบายถึงรูปแบบของการลงทุนที่ต่างกัน

จากที่กล่าวมานั้น ผู้วิจัยจึงสนใจในการนำเทคนิคการทำเหมืองข้อมูลมาใช้วิเคราะห์หาความต้องการของลูกค้า โดยการทำเหมืองข้อมูล (Data Mining) เป็นเครื่องมือกระบวนการทางด้านการวิเคราะห์ข้อมูลเพื่อค้นหาองค์ความรู้ที่ได้รับความนิยม และได้รับการยอมรับว่าเป็นเครื่องมือที่จะช่วยให้หลายประเภทธุรกิจสามารถเข้าใจความต้องการของลูกค้า เพื่อนำไปสู่การปรับเปลี่ยนกลยุทธ์ให้ทันสถานการณ์ในภาวะที่มีการแข่งขันสูง และมีเทคโนโลยีที่ทันสมัย การทำเหมืองข้อมูลจึงเป็นเครื่องมือด้านสารสนเทศที่มีประโยชน์อย่างแท้จริง อีกทั้งผู้วิจัยยังสนใจในการนำกระบวนการเรียนรู้ของเครื่อง (Machine learning) เข้ามาช่วยในการประมวลผลข้อมูลเพื่อรองรับกับการวิเคราะห์และเรียนรู้จากข้อมูลจำนวนมาก ให้เกิดประโยชน์และเป็นแนวทางในการแนะนำผลิตภัณฑ์กองทุนได้ เช่น การใช้โครงข่ายประสาทเทียมสำหรับพยากรณ์ราคาสูงสุดและราคาต่ำสุดบนตลาดอัตราแลกเปลี่ยน (ฆฤณ ชินประสาทศักดิ์, 2555) การพยากรณ์อัตราแลกเปลี่ยนของคู่สกุลเงิน ยูโร/ดอลลาร์สหรัฐ เป็นต้น โดยการใช้ข้อมูลรับเข้าเป็นตัวบ่งชี้ทางเทคนิค และวัดประสิทธิภาพของแบบจำลองด้วยระบบจำลองการซื้อขาย ซึ่งประกอบด้วย กลยุทธการลงทุน วิธีจัดการความเสี่ยง และหลักการบริหารเงิน โดยผลการทดลองแสดงให้เป็นว่าโครงข่ายประสาทเทียมที่นำเสนอให้ผลตอบแทนที่ดีกว่าแบบจำลองจากงานวิจัยต้นแบบ หรือยกตัวอย่างองค์กรขนาดใหญ่ (Marr, 2020) อย่างธนาคารซิตี้แบงก์ (Citibank) ที่ให้ประโยชน์ในแวดวงธนาคารในการรักษาลูกค้า และการจัดหาลูกค้า โดยการวิเคราะห์ข้อมูลและกำหนดเป้าหมายการใช้จ่ายเพื่อการส่งเสริมการขายโดยใช้อัลกอริทึมในกระบวนการเรียนรู้ของเครื่อง อีกประการหนึ่งคือ การเก็บรวบรวมบันทึกธุรกรรมเพื่อระบุความผิดปกติ ซึ่งในกรณีของลูกค้าของธนาคารซิตี้แบงก์ อาจหมายถึงการเรียกเก็บเงินที่ไม่ถูกต้อง ผิดปกติ หรือเป็นการฉ้อโกง ซึ่งค่าใช้จ่ายที่เกิดจากความผิดปกติเหล่านี้จัดการได้ง่ายกว่ามาก หากตรวจพบปัญหาอย่างรวดเร็ว หรือแม้กระทั่งก่อนที่จะเกิดขึ้น โดยผ่านการสร้างแบบจำลองเชิงคาดการณ์ หรืออเมริกัน เอ็กซ์เพรสประเทศไทย (American Express, Thailand) กับการใช้กระบวนการเรียนรู้ของเครื่อง ด้วยข้อมูลการใช้บัตรมากกว่า 110 ล้านใบ และธุรกรรมที่ประมวลผลมากกว่าหนึ่งล้านล้านดอลลาร์ ทำให้อเมริกัน เอ็กซ์เพรสมีการไหลของข้อมูลที่หลากหลายเพื่อใช้ประโยชน์ โดยในการวิเคราะห์ข้อมูล

โดยเฉพาะกระบวนการเรียนรู้ของเครื่อง เป็นหัวใจสำคัญของการตัดสินใจของอเมริกัน เอ็กซ์เพรส คือ การตรวจจับการฉ้อโกง และการนำผู้ค้ากับลูกค้าเข้าใกล้กันมากขึ้น โดยการตรวจจับ และป้องกันการฉ้อโกงบัตรเครดิต อาศัยอัลกอริทึมกระบวนการเรียนรู้ของเครื่องอย่างมาก อันมีเป้าหมายคือการตรวจจับธุรกรรมฉ้อโกงโดยเร็วที่สุดเพื่อลดการสูญเสีย โดยใช้ข้อมูลจากแหล่งข้อมูลที่หลากหลาย รวมถึงข้อมูลสมาชิกบัตร รายละเอียดการใช้จ่าย และข้อมูลร้านค้าเพื่อตรวจจับเหตุการณ์ที่น่าสงสัย และตัดสินใจ

จากที่กล่าวมา อาจคาดการณ์ได้ว่าในอนาคตมีแนวโน้มว่าจะมีผู้ลงทุนรายใหม่เข้าสู่ตลาดหลักทรัพย์อย่างต่อเนื่อง เนื่องจากบริษัทหลักทรัพย์จัดการกองทุนมีการทำกิจกรรมประชาสัมพันธ์โดยตลอด อีกทั้งการพัฒนาทางเทคโนโลยีสารสนเทศ รวมถึงจำนวนบริษัทที่ได้รับใบอนุญาต การจัดการกองทุนรวมประเภทบริษัทหลักทรัพย์ที่เพิ่มขึ้น ผู้วิจัยมีความประสงค์ในการนำเสนอการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Ensemble Learning) เครื่องมือการพยากรณ์ที่มีประสิทธิภาพในการทำเหมืองข้อมูลและกระบวนการเรียนรู้ของเครื่องเพื่อพัฒนาการทำงานในการแนะผลิตภัณฑ์กองทุน จากข้อมูลลูกค้าต่างๆ ที่มีอยู่แล้วของธนาคารมาใช้ในการสนับสนุนวิเคราะห์พัฒนาการบริการที่มีความแม่นยำและส่งผลประโยชน์ต่อองค์กร ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนารูปแบบการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม สำหรับใช้เป็นรูปแบบแนะนำผลิตภัณฑ์กองทุนให้กับลูกค้ารายใหม่ รวมถึงเปรียบเทียบประสิทธิภาพอัลกอริทึมและรูปแบบในการจัดการข้อมูลก่อนนำไปพัฒนารูปแบบพยากรณ์

ความมุ่งหมายของการวิจัย

1. เพื่อพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม
2. เพื่อเปรียบเทียบประสิทธิภาพอัลกอริทึมที่ใช้ในแบบจำลองการพยากรณ์ลูกค้า
3. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการใหม่ที่นำเสนอกับงานวิจัยที่ผ่านมาเพื่อยืนยันผลลัพธ์ในการนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

พหุน ปณ ฑิโต ชีเว

ความสำคัญของการวิจัย

1. เพื่อเป็นข้อมูลในการพัฒนาระบบ หรือกระบวนการในการแนะนำผลิตภัณฑ์กองทุนให้กับลูกค้าด้วยแบบจำลองการพยากรณ์ ที่มีประสิทธิภาพและเหมาะสมมากที่สุด

2. เพื่อเป็นสารสนเทศในการสนับสนุนการตัดสินใจในการบริหารและจัดการทักษะของบุคคลากรในการทำงาน การแนะนำผลิตภัณฑ์กองทุนและพัฒนากลยุทธ์ทางการตลาดที่มีความแม่นยำ ตรงตามความต้องการของลูกค้าได้อย่างมีประสิทธิภาพ

3. เพื่อเป็นข้อข้อมูลในการนำไปประยุกต์ใช้ในการวิเคราะห์ ออกแบบหรือพัฒนา เพื่อแก้ไขปัญหาการแนะนำผลิตภัณฑ์กองทุนได้อย่างเหมาะสม

กรอบแนวคิดที่ใช้ในการวิจัย

1. การวิจัยเรื่อง การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม เสนอเป็นข้อมูลพื้นฐาน ปัจจัย และข้อเสนอแนะในลักษณะของรูปแบบพยากรณ์ด้วยกระบวนการวิเคราะห์ข้อมูลและกระบวนการเรียนรู้ของเครื่อง เพื่อใช้ประกอบการพิจารณาในการกำหนดแผนกลยุทธ์ในการแนะนำกองทุนให้กับลูกค้ารายใหม่ และศึกษาพฤติกรรมลูกค้าจากข้อมูลเดิมที่มีอยู่ในระบบธนาคาร

2. เพื่อศึกษาและเก็บข้อมูลประสิทธิภาพของรูปแบบการเรียนรู้ (Performances) และความแม่นยำ ความน่าเชื่อถือ และการต่อยอดเพื่อศึกษาพฤติกรรมลูกค้ามาช่วยในการแนะนำและพัฒนาผลิตภัณฑ์กองทุน

ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาศึกษาเป็นข้อมูลเกี่ยวกับนักลงทุนในกองทุนต่าง ๆ โดยจัดเก็บจากบริษัทหลักทรัพย์จัดการกองทุนแห่งหนึ่ง

2. กลุ่มตัวอย่างที่ใช้ในการวิจัย ผู้ที่ทำการซื้อกองทุนผ่านบริษัทหลักทรัพย์จัดการกองทุนแห่งหนึ่ง ภายในเดือน มกราคม - กรกฎาคม 2562 ผ่านช่องทางธนาคาร ที่มีมูลค่าการเลือกซื้อผลิตภัณฑ์กองทุน 500 บาทขึ้นไป จำนวน 19,577 ราย

3. ข้อมูลลูกค้าบริษัทหลักทรัพย์จัดการกองทุน ได้แก่ เพศ ภูมิลำเนา การศึกษา อาชีพ รายได้ ความถี่ในการซื้อขายกองทุน สถาณะการเงินในบัญชีฝากประจำ บัญชีออมทรัพย์ และบัญชีพนักงาน ที่บริษัทมีนิติบุคคลกับธนาคาร

4. ตัวแปรที่เป็นเป้าหมายในการพยากรณ์เป็นผลิตภัณฑ์กองทุนที่ผู้ลงทุนเลือกลงทุนผ่านบริษัทหลักทรัพย์จัดการกองทุน

5. ด้านเทคนิคและวิธีการวิเคราะห์ข้อมูลจะใช้ การเรียนรู้แบบกลุ่ม (Ensemble Learning) และเทคนิคการค้นหารายการส่วนใหญ่ในรายการในลักษณะที่มีประสิทธิภาพมากที่สุด (Voting Algorithm) ซึ่งประกอบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Neural Network) เคเนียร์เรสเนเบอร์ (K-NN) นาอีฟเบย์ (Naïve Bayes) และกฎการอุปนัย (Rule Induction)

ประโยชน์ที่คาดว่าจะได้รับ

1. มีความรู้ ความเข้าใจในทฤษฎีในการพัฒนารูปแบบและกระบวนการวิเคราะห์ข้อมูลด้วยกระบวนการเรียนรู้ของเครื่อง

2. สามารถนำรูปแบบพยากรณ์ที่ปรับใช้ในการคำแนะนำผลิตภัณฑ์กองทุนประเภทต่าง ๆ ให้แก่ผู้ลงทุนที่สนใจได้อย่างเหมาะสม

3. เรียนรู้ประสิทธิภาพของแต่ละอัลกอริทึม ที่มีผลต่อการพัฒนารูปแบบพยากรณ์จากข้อมูลที่ได้ใช้ในการวิจัย

นิยามศัพท์เฉพาะ

1. ผู้ลงทุน หมายถึง ผู้ที่มีความต้องการลงทุน มีอำนาจตัดสินใจในการลงทุน

2. กองทุนรวม หมายถึง โครงการลงทุนที่จัดตั้งขึ้น และบริหารโดยบริษัทหลักทรัพย์จัดการกองทุนรวมที่ได้รับอนุญาตให้ประกอบกิจการจากธนาคารแห่งประเทศไทย มีวัตถุประสงค์ที่จะระดมทุนจากนักลงทุนทั่วไป หรือผู้มีเงินออมโดยการเสนอขายหน่วยลงทุนแล้ว นำเงินที่ได้ไปลงทุนในหลักทรัพย์ประเภทต่างๆ ที่กฎหมายอนุญาตให้ลงทุนได้

3. หน่วยลงทุน หมายถึง ตราสารแสดงสิทธิในทรัพย์สินของโครงการจัดการลงทุนของกองทุนที่ออกให้กับผู้ลงทุน

4. บริษัทหลักทรัพย์จัดการกองทุน (บลจ.) หมายถึง บริษัทที่มีความเชี่ยวชาญในการจัดการลงทุน และได้รับใบอนุญาตจาก สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ (ก.ล.ต.) เป็นผู้บริหารจัดการเงินลงทุนก้อนนั้น

5. ลักษณะบุคคล หมายถึง ข้อมูล หรือสิ่งใดๆ ก็ตามที่เป็นเครื่องบงชี้ให้เห็น และเขาใจถึงเรื่องราว หรือลักษณะเฉพาะตัวของบุคคลหนึ่งบุคคลใด เช่น เพศ อายุ ระดับการศึกษา รายได้เฉลี่ยต่อเดือน อาชีพที่มีผลตอแนวโนมการตัดสินใจลงทุนในกองทุนรวม

6. เหมืองข้อมูล (Data Mining) หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ ตลอดถึงความสัมพันธ์ที่แฝงอยู่ในชุดข้อมูลนั้น ๆ จากฐานข้อมูลขนาดใหญ่เพื่อให้ได้สารสนเทศ (Useful Information) จากข้อมูลที่ยังไม่ทราบ (Unknown Data) โดยจะเป็นข้อมูลที่มีเหตุผล และสามารถนำไปใช้ได้ ซึ่งเป็นสิ่งสำคัญที่จะช่วยสนับสนุนการตัดสินใจ

7. กระบวนการเรียนรู้ของเครื่อง (Machine learning) หมายถึง หมายถึงการที่กำหนดให้เครื่อง (คอมพิวเตอร์) สามารถปฏิบัติงาน โดยเรียนรู้จากการกระทำหรือสิ่งที่ทำไปก่อนหน้านั้น โดยใช้หลักการของปัญญาประดิษฐ์ (Artificial intelligence) หรืออาจเป็นการเรียนรู้จากการถูกสั่งให้ทำจากตัวอย่าง จากการเปรียบเทียบ (analogy) โดยปราศจากการป้อนคำสั่งของโปรแกรมเมอร์

8. การแบ่งกลุ่มข้อมูล (Clustering) หมายถึง การแบ่งชุดข้อมูลออกเป็นกลุ่ม (Cluster) โดยนำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน จะอาศัยความเหมือน (Similarity) หรือ ความใกล้ชิด (Proximity) โดยคำนวณจากการวัดระยะระหว่างเวกเตอร์ของข้อมูลเข้า โดยใช้การวัดระยะแบบต่าง ๆ

9. การจำแนกข้อมูล (Classification) กลุ่มของ algorithm ที่เน้นสอน computer โดยการศึกษาจากข้อมูลตัวอย่าง อันมีรูปแบบการประเมินชัดเจนว่าถ้าทำงานแล้วผลลัพธ์ได้แบบนี้คือถูกหรือผิด

10. การเรียนรู้แบบกลุ่ม (Ensemble learning) เทคนิคของการเรียนรู้ของเครื่องขั้นสูง (Advanced Machine Learning Model) ที่ช่วยเพิ่มประสิทธิภาพการพยากรณ์ผล (Prediction Performance) โดยการใช้เทคนิคที่ใช้แบบจำลองการจำแนกข้อมูล (Classification) หลายๆ แบบจำลอง มาช่วยในการหาคำตอบ

พหุน ปณ ฑิโต ชีเว

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่อง การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Development of Customer Predictive Model for Investment using Ensemble Learning Techniques) ได้ประกอบไปด้วยทฤษฎีและงานวิจัยที่เกี่ยวข้อง ดังนี้

1. แนวคิดด้านการลงทุน
2. ความรู้เกี่ยวกับผลิตภัณฑ์กองทุน
3. การตัดสินใจลงทุน พฤติกรรม และการวิเคราะห์ทางด้านประชากรศาสตร์
4. หลักการวิเคราะห์ข้อมูลและกระบวนการเรียนรู้ของเครื่อง (Machine learning)
5. เทคนิคการแบ่งกลุ่มข้อมูลข้อมูลเคมีน (K-Means)
6. เทคนิคการจำแนกข้อมูล
7. เทคนิคการเรียนรู้แบบกลุ่ม
8. เอกสารและงานวิจัยที่เกี่ยวข้อง

แนวคิดด้านการลงทุน

การลงทุน (Investment) หรือการนำเงินที่เก็บสะสมไปสร้างผลตอบแทนที่สูงกว่าการออม โดยอีกความหมาย คือ การเก็บเงินส่วนที่เหลือจากการใช้จ่ายเพื่อบริโภคไปลงทุนเพื่อให้ได้ผลตอบแทนที่มากขึ้นกว่าเดิม เช่น การฝากเงินในสถาบันการเงินการธนาคาร แต่อัตราผลตอบแทนที่ได้รับเป็นดอกเบี้ยไม่สูงมากตามที่ต้องการ ดังนั้น นักลงทุนจึงหาแหล่งลงทุนที่สามารถได้รับอัตราผลตอบแทนสูงกว่า (SangSoi, 2015) นอกจากนี้ ทฤษฎีการลงทุน (ศูนย์ส่งเสริมการพัฒนาความรู้ตลาดทุน, 2553) ได้ถูกกล่าวถึงในรูปแบบการอธิบายความหมายของการลงทุนไว้ว่า เป็นการกันเงินจำนวนหนึ่งในช่วงระยะเวลาหนึ่ง เพื่อก่อให้เกิดกระแสเงินสดรับในอนาคตซึ่งจะชดเชยให้แก่ผู้กันเงิน โดยกระแสเงินสดรับนี้ควรคุ้มกับอัตราเงินเฟ้อ และคุ้มกับความไม่แน่นอนที่จะเกิดขึ้นกับกระแสเงินสดรับในอนาคต ดังนั้น จึงอาจกล่าวสรุปได้ว่า การลงทุน หมายถึง การนำเงินเก็บสะสมจำนวนหนึ่งมาซื้ออสังหาริมทรัพย์ หรือหลักทรัพย์ของบุคคลหรือสถาบัน โดยมีผลตอบแทนจากการลงทุนที่มาควบคู่กับความเสี่ยง

การลงทุนแบ่งเป็น 3 ประเภท คือ การลงทุนเพื่อการบริโภค การลงทุนในธุรกิจ และการลงทุนในหลักทรัพย์ Sungkaew (2015) สามารถอธิบายได้ ดังนี้

1. การลงทุนเพื่อการบริโภค (Consumer Investment) เป็นการลงทุนเพื่อขายสินทรัพย์ถาวร (Durable Goods) เช่น รถยนต์, ตู้เย็น และสิ่งของเครื่องใช้ต่าง ๆ เป็นต้น โดยการลงทุนประเภทนี้เป็นการลงทุนที่ไม่หวังผลกำไร แต่ต้องการความพอใจ

2. การลงทุนในธุรกิจ (Business or Economic Investment) เป็นการซื้อสินทรัพย์เพื่อประกอบธุรกิจน้อยที่สุด เป้าหมายในการลงทุนของธุรกิจคือ กำไร โดยที่กำไรจะเป็นตัวดึงดูดทำให้ผู้ลงทุนนำเงินมาลงทุนมากขึ้น

3. การลงทุนในหลักทรัพย์ (Financial or Securities Investment) การลงทุนในหลักทรัพย์คือการซื้อสินทรัพย์ (Asset) ในรูปของหลักทรัพย์ (Securities) อย่างเช่น การลงทุนพันธบัตร การลงทุนในหุ้นกู้หรือหุ้นทุน กาลงทุนแบบนี้เป็นการลงทุนทางอ้อม ต่างจากการลงทุนในธุรกิจ เพราะผู้ลงทุนไม่ต้องประกอบธุรกิจเอง เป็นการลงทุนที่ผู้ลงทุนจะนำเงินที่ออมได้ ไปซื้อหลักทรัพย์ที่จะลงทุน ผลตอบแทนที่ได้จะเป็น ดอกเบี้ยหรือเงินปันผลแล้วแต่ประเภทของหลักทรัพย์ที่จะลงทุน

การลงทุนที่เกี่ยวข้องในงานวิจัยครั้งนี้ จัดอยู่ในประเภทการลงทุนในหลักทรัพย์ (Financial หรือ Securities Investment) ซึ่งเป็นการซื้อสินทรัพย์ในรูปของหลักทรัพย์ เช่น พันธบัตร หุ้นกู้ หรือหุ้นทุนของหน่วยงานภาครัฐหรือภาคเอกชน โดยการลงทุนในลักษณะนี้เป็นการลงทุนทางอ้อม เนื่องจากมีความเสี่ยง หรือผู้ลงทุนเองยังไม่มีทุนมากพอ โดยผลตอบแทนจากการลงทุนจะอยู่ในรูปของดอกเบี้ยหรือเงินปันผล แล้วแต่ประเภทหลักทรัพย์ที่ลงทุน ซึ่งดอกเบี้ยหรือเงินปันผลที่ได้รับนั้นเป็นผลมาจากการดำเนินงานของหน่วยงานที่ผู้ลงทุนตัดสินใจลงทุน นอกจากนี้ ผู้ลงทุนอาจได้ผลตอบแทนอีกลักษณะหนึ่งคือกำไรจากการขายหลักทรัพย์ เมื่อราคาเสนอขายสูงกว่าราคาเสนอซื้อ ทั้งนี้ ผลตอบแทนดังกล่าวจะอยู่ภายใต้ความเสี่ยงทั้งที่ควบคุมได้และควบคุมไม่ได้ ข้อมูลพื้นฐานสำคัญที่นักลงทุนมักนำมาพิจารณาในการตัดสินใจลงทุน คือ ราคาตลาดของหลักทรัพย์ในขณะที่ราคาของหลักทรัพย์มักมีการเปลี่ยนแปลงตลอดเวลา โดยการเปลี่ยนแปลงที่เพิ่มขึ้นหรือลดลงจะสามารถนำมาวิเคราะห์ประกอบการตัดสินใจในการลงทุนว่าหลักทรัพย์นั้นมีความเหมาะสมที่จะลงทุนหรือไม่

ความรู้เกี่ยวกับผลิตภัณฑ์กองทุน

กองทุนรวม คือ เครื่องมือในการลงทุน (Investment Vehicle) สำหรับผู้ลงทุนรายย่อยที่ประสงค์จะนำเงินมาลงทุนในตลาดเงินตลาดทุน แต่ติดขัดด้วยอุปสรรคหลายประการ ที่ทำให้การลงทุนด้วยตนเองไม่สามารถได้ผลลัพธ์ตามเป้าหมายที่ต้องการ เช่น มีทุนทรัพย์ที่จำกัด ไม่สามารถกระจายการลงทุนในหลักในแต่ละประเภทได้มากพอสำหรับการลดความเสี่ยงจากการลงทุน หรือไม่มีประสบการณ์ ความรู้ ความชำนาญในการลงทุน หรือไม่มีเวลาศึกษา ค้นหา และติดตามข้อมูลเพื่อใช้ในการตัดสินใจการลงทุน กองทุนรวม จึงเป็นเครื่องมือในการลงทุนที่มีประสิทธิภาพ มีการจัดการลงทุนอย่างเป็นระบบ โดยมีจุดมุ่งหมายให้การลงทุนได้รับผลตอบแทนที่ดีสุด ภายใต้กรอบความเสี่ยงที่ผู้ลงทุนยอมรับ

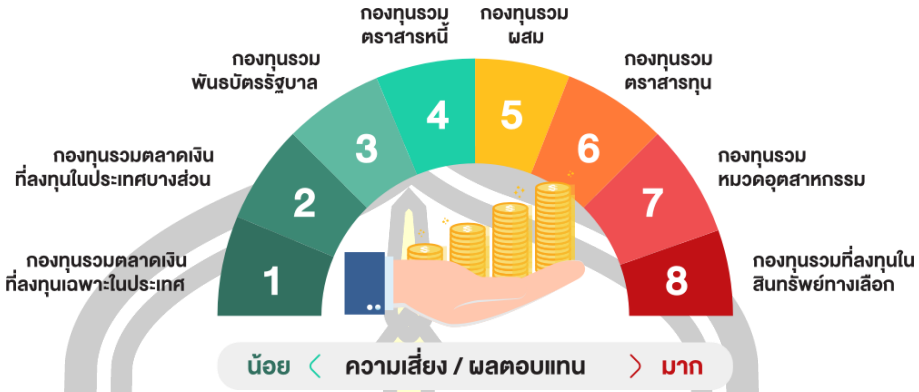
ผู้วิจัยดำเนินศึกษาข้อมูลตามลักษณะของตัวแปรตามจากชุดข้อมูลวิจัย ได้แก่ กองทุนรวม กองทุนตลาดเงิน กองทุนที่มีกำหนดอายุโครงการ และกองทุน RMF (Retirement Mutual Fund) และ LTF (Long Term Equity Fund) โดยสามารถอธิบายได้ ดังนี้

กองทุนรวม (Mutual Fund)

โครงการลงทุนที่ระดมเงินทุนจากนักลงทุนหลาย ๆ คน นำมาลงทุนตามนโยบายที่กองทุนรวมนั้น ๆ แล้วนำไปจดทะเบียนให้มีฐานะเป็นนิติบุคคล จากนั้นนำเงินที่ระดมทุนได้ไปลงทุนในหลักทรัพย์ หรือทรัพย์สินประเภทต่างๆ ตามนโยบายการลงทุนที่ได้ระบุไว้ในหนังสือชี้ชวนเสนอขายแก่นักลงทุน กองทุนรวมมีจุดเด่นตรงที่มีเงินน้อยก็ลงทุนได้ แถมยังมีนโยบายการลงทุนที่หลากหลาย ทั้งกองที่ลงทุนในตราสารหนี้ หุ้นในประเทศ หุ้นต่างประเทศ รวมไปถึงอสังหาริมทรัพย์ ทองคำ และสินทรัพย์อื่น ๆ

ดังนั้น แต่ละกองจึงมีผลตอบแทนและความเสี่ยงที่แตกต่างกันตามประเภทสินทรัพย์ที่กองทุนไปลงทุน นักลงทุนสามารถเลือกกองทุนให้เหมาะกับวัตถุประสงค์การลงทุนและระดับความเสี่ยงที่ยอมรับได้ ซึ่งจะมีการทำแบบทดสอบวัดระดับความเสี่ยงเมื่อเปิดบัญชีเริ่มซื้อขายกองทุน โดยกองทุนรวมที่ขายในประเทศไทย สามารถแบ่งออกเป็น 8 ประเภทหลัก ตามระดับความเสี่ยง (Risk spectrum) ตามภาพประกอบ 3 และมีรายละเอียด (ตลาดหลักทรัพย์แห่งประเทศไทย, 2563) ดังนี้

พหุน ปณฺฑิโต ชีเว



ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย, 2563

ภาพประกอบ 3 ประเภทกองทุนรวมตามระดับความเสี่ยง

1. ความเสี่ยงระดับที่ 1 กองทุนรวมตลาดเงินในประเทศ เป็นกองทุนที่เน้นลงทุนในเงินฝาก ตั๋วเงิน รวมถึงตราสารหนี้ที่มีอายุเฉลี่ยไม่เกิน 1 ปี มีความเสี่ยงต่ำที่สุด เหมาะสำหรับผู้ลงทุนที่ไม่ต้องการความเสี่ยง หรือผู้ที่ต้องการพักเงินในช่วงที่ผลตอบแทนไม่ดี

2. ความเสี่ยงระดับที่ 2 กองทุนรวมตลาดเงินต่างประเทศ เป็นกองทุนที่เน้นลงทุนในเงินฝาก ตั๋วเงิน รวมถึงตราสารหนี้ที่มีอายุเฉลี่ยไม่เกิน 1 ปี เหมือนกลุ่มแรก แต่อาจมีการลงทุนในต่างประเทศบางส่วน จึงทำให้มีความเสี่ยงเรื่องอัตราแลกเปลี่ยนเพิ่มขึ้นมา แต่โดยปกติกองทุนประเภทนี้มักจะมีการป้องกันความเสี่ยงอัตราแลกเปลี่ยนเอาไว้

3. ความเสี่ยงระดับที่ 3 กองทุนรวมพันธบัตรรัฐบาล เป็นกองทุนที่เน้นลงทุนในตราสารหนี้ภาครัฐ เช่น พันธบัตรรัฐบาล หรือพันธบัตรรัฐวิสาหกิจ ซึ่งส่วนใหญ่จะมีอายุเฉลี่ยมากกว่า 1 ปีขึ้นไป จึงมีความผันผวนมากกว่ากองทุนรวมตลาดเงิน

4. ความเสี่ยงระดับที่ 4 กองทุนรวมตราสารหนี้ เป็นกองทุนที่เน้นลงทุนในตราสารหนี้ทั้งภาครัฐและภาคเอกชน เช่น พันธบัตรรัฐบาล พันธบัตรรัฐวิสาหกิจ ตั๋วเงินคลัง และหุ้นกู้เอกชน ซึ่งมีทั้งกองที่ลงทุนในตราสารหนี้ระยะสั้น (อายุเฉลี่ยไม่เกิน 1 ปี) และตราสารหนี้ระยะยาว (อายุเฉลี่ยมากกว่า 1 ปี) จึงเหมาะกับผู้ที่ลงทุนได้ทั้งระยะสั้น ระยะยาว และผู้ที่รับความเสี่ยงได้ไม่มาก คาดหวังผลตอบแทนที่แน่นอน สม่ำเสมอ รวมถึงผู้ที่ต้องการกระจายความเสี่ยงของพอร์ตการลงทุน

5. ความเสี่ยงระดับที่ 5 กองทุนรวมผสม เป็นกองทุนที่สามารถลงทุนในสินทรัพย์อะไรก็ได้ ไม่ว่าจะเป็นเงินฝาก ตราสารหนี้ หุ้น หรืออื่น ๆ ซึ่งสัดส่วนการลงทุนจะลงทุนในอะไรมากกว่ากัน ก็ต้องไปดูในนโยบายการลงทุนของกองทุน เหมาะสำหรับผู้ที่รับความเสี่ยงได้ปานกลางถึงสูง และผู้ที่ไม่มีเวลาในการปรับสัดส่วนกองทุนหรือหุ้น หากตลาดมีความผันผวนมาก

6. ความเสี่ยงระดับที่ 6 กองทุนรวมตราสารทุน เป็นกองทุนที่เน้นลงทุนในหุ้นทั้งในประเทศและต่างประเทศ ซึ่งกองทุนประเภทนี้หมายความรวมไปถึงกองทุน SSF และ RMF เหมาะกับผู้ที่ต้องการผลตอบแทนสูงและรับความเสี่ยงได้สูง หรือผู้ที่ชอบการลงทุนในหุ้น แต่ไม่มีเวลาบริหารการลงทุน

7. ความเสี่ยงระดับที่ 7 กองทุนรวมตามหมวดอุตสาหกรรม เป็นกองทุนที่เน้นลงทุนในหุ้น แต่เจาะจงอุตสาหกรรมมากขึ้น เช่น หุ้นธนาคาร หุ้นสื่อสาร หุ้นโรงพยาบาล ฯลฯ แต่กองทุนประเภทนี้จะมีความเสี่ยงสูงกว่ากองทุนรวมตราสารทุนทั่วไป เนื่องจากมีการลงทุนแบบกระจุกตัว จึงเหมาะกับผู้ที่รับความเสี่ยงได้สูง และมีความรู้ความเข้าใจในอุตสาหกรรมนั้น ๆ อย่างดี

8. ความเสี่ยงระดับที่ 8 กองทุนรวมทางเลือก เป็นกองทุนที่เน้นลงทุนในทางเลือกอื่นๆ ที่นอกเหนือจากสินทรัพย์พื้นฐาน เช่น ทองคำ น้ำมัน เป็นต้น เหมาะกับผู้ที่รับความเสี่ยงได้สูง และผู้ที่ต้องการกระจายการลงทุนไปยังสินทรัพย์อื่น ๆ เพื่อลดความเสี่ยงของพอร์ตการลงทุนโดยรวม แต่อาจต้องศึกษาและทำความเข้าใจรายละเอียดที่เกี่ยวข้องกับทางเลือกนั้นๆ มากยิ่งขึ้น

กองทุนตลาดเงิน (Money Market)

กองทุนรวมที่เน้นลงทุนในเงินฝากของธนาคาร หรือตราสารหนี้ระยะสั้นที่มีอายุคงเหลือไม่เกิน 1 ปี เช่น ตั๋วเงินคลัง ตั๋วแลกเงิน ตั๋วสัญญาใช้เงิน หรือพันธบัตรและหุ้นกู้เอกชนที่มีอายุคงเหลือไม่เกิน 1 ปี ทางเลือกลงทุนที่เหมาะสมสำหรับคนที่ต้องการบริหารสภาพคล่อง เพราะสามารถซื้อขายได้ทุกวันทำการคล้ายกับบัญชีออมทรัพย์ที่สามารถเบิกถอนได้ตลอด และยังให้ผลตอบแทนที่ดีกว่า โดยกองทุนรวมตลาดเงินมีความเสี่ยงต่ำ ทำให้ราคาหน่วยลงทุนไม่ค่อยผันผวนมาก วิธีการลงทุนที่เหมาะสมที่สุดจึงเป็นการซื้อด้วยเงินก้อนใหญ่ เนื่องจากไม่จำเป็นต้องถัวเฉลี่ยราคาต้นทุน กองทุนรวมตลาดเงิน แบ่งได้เป็น 2 ประเภท (ตลาดหลักทรัพย์แห่งประเทศไทย, 2564) คือ

1. กองทุนรวมตลาดเงิน (ตราสารหนี้ภาครัฐ) เน้นลงทุนในตราสารหนี้ภาครัฐเป็นหลัก เป็นกองทุนที่มีความเสี่ยงต่ำ

2. กองทุนรวมตลาดเงิน (ทั่วไป) ที่นอกจากลงทุนในตราสารหนี้ภาครัฐแล้ว ยังกระจายการลงทุนไปในตราสารหนี้ของบริษัทเอกชนที่มีความมั่นคง และเงินฝากของสถาบันการเงินด้วย เพื่อเพิ่มผลตอบแทนให้สูงมากขึ้น

กองทุนที่มีกำหนดอายุโครงการ (Term Fund)

กองทุนรวมตราสารหนี้ที่มีระยะเวลา หรือมีกำหนดอายุโครงการที่ชัดเจน เช่น 3 เดือน 6 เดือน หรือ 1 ปี ดังนั้นการซื้อขายจะต้องเป็นไปตามเงื่อนไขที่บริษัทหลักทรัพย์จัดการกองทุน (บลจ.) กำหนด คล้ายๆ กับการฝากเงินแบบประจำ โดยนโยบายการลงทุนของกองทุนรวม

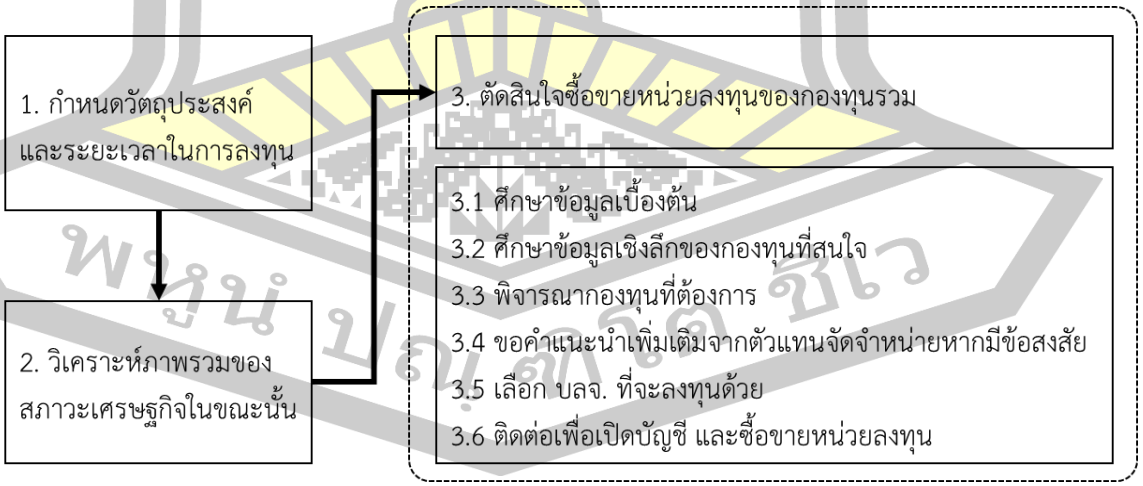
Term Fund จะลงทุนในตราสารหนี้ทั้งในและต่างประเทศ โดยเน้นบริหารกองทุนเพื่อให้สร้างผลตอบแทนที่สูงกว่าเงินฝากประจำ ดังนั้น ตราสารหนี้ที่ลงทุนจึงมีความเสี่ยงต่ำ เช่น พันธบัตรรัฐบาล หุ้นกู้ภาคเอกชนที่มีความน่าเชื่อถือระดับที่ลงทุนได้ (Investment Grade) รวมถึงเงินฝาก

กองทุน RMF (Retirement Mutual Fund) และ LTF (Long Term Equity Fund)

เป็นเครื่องมือการลงทุนอีกชนิดที่มีลักษณะพิเศษแตกต่างจากกองทุนรวมทั่วไป ก็คือ เงินลงทุนในกองทุนรวมทั้งสองประเภทนี้ผู้ลงทุนสามารถนำไปลดหย่อนภาษีได้ถือเป็นสิทธิประโยชน์พิเศษที่ผู้ลงทุนจะได้รับนอกเหนือจากผลตอบแทนจากการลงทุน กองทุน RMF คือกองทุนรวมที่มีจุดมุ่งหมายเพื่อส่งเสริมให้เกิดการออมเงินระยะยาวไว้สำหรับใช้จ่ายยามเกษียณอายุ ซึ่งจะคล้ายกับกองทุนสำรองเลี้ยงชีพของภาคเอกชน และกองทุนบำเหน็จบำนาญข้าราชการ และกองทุน LTF เป็นกองทุนรวมที่มีนโยบายการลงทุนในหุ้นสามัญจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย จัดตั้งขึ้นโดยมีจุดมุ่งหมายในการส่งเสริมการลงทุนระยะยาวในตลาดหุ้น โดยผู้ลงทุนจะได้รับสิทธิประโยชน์ทางภาษีเป็นสิ่งจูงใจในการลงทุน เพื่อแลกกับการลงทุนตามเงื่อนไขที่สรรพากรกำหนด (ตลาดหลักทรัพย์แห่งประเทศไทย, 2558)

การตัดสินใจลงทุน พฤติกรรม และการวิเคราะห์ทางด้านประชากรศาสตร์

ในการซื้อขายหน่วยลงทุนในกองทุนรวม เริ่มตั้งแต่การเตรียมความพร้อม โดยสิ่งที่ผู้ลงทุนต้องให้ความสำคัญคือ การสำรวจสถานะทางการเงินของตนเอง มีเงินที่สามารถนำมาใช้สำหรับการลงทุนเท่าไร และเงินที่จะนำมาลงทุนต้องเป็นเงินที่ปลอกจากภาระผูกพันใด ๆ และไม่ส่งผลกระทบต่อการดำรงชีวิตประจำวัน โดยสามารถอธิบายขั้นตอนได้ ดังนี้



ภาพประกอบ 4 ขั้นตอนการตัดสินก่อนการซื้อขายหน่วยลงทุนในกองทุนรวม

กำหนดวัตถุประสงค์ และระยะเวลาในการลงทุน

ตั้งคำถามต่อผู้ต้องลงทุนเองให้ได้มาซึ่งเป้าหมาย หรือวัตถุประสงค์ในการลงทุนเป็นอย่างไร โดยเป้าหมายควรตั้งอยู่บนพื้นฐานความเป็นจริงมีความชัดเจน และสมเหตุสมผลทั้งในเรื่องของผลตอบแทน หรือจำนวนเงินที่ต้องการได้รับจากการลงทุนและความเสี่ยงที่สามารถยอมรับได้ตลอดจนระยะเวลาในการลงทุนเพื่อความต้องในการใช้เงินในอนาคต เนื่องจากกองทุนรวมแต่ละประเภทความแตกต่างกัน เช่น อัตราผลตอบแทน ความเสี่ยงของตราสารที่กองทุนรวมถืออยู่ และอายุโครงการหรือระยะเวลาในการลงทุนที่แตกต่างกันออกไปด้วย

วิเคราะห์ภาพรวมของสภาวะเศรษฐกิจในขณะนั้น

เป็นการวิเคราะห์ว่าบรรยากาศ และทิศทางของเศรษฐกิจเหมาะสมกับการลงทุนหรือไม่ ผลกระทบของสภาวะเศรษฐกิจมีต่อการลงทุนเป็นอย่างไร มากน้อยแค่ไหน เพราะสภาวะเศรษฐกิจถือเป็นปัจจัยสำคัญที่มีผลต่อการตัดสินใจลงทุนของผู้ลงทุน นอกจากนี้ สภาวะเศรษฐกิจยังส่งผลต่อตราสารการลงทุนประเภทต่าง ๆ อีกด้วย

ตัดสินใจซื้อขายหน่วยลงทุนของกองทุนรวม

1. ศึกษาหาข้อมูลเบื้อต้นเกี่ยวกับกองทุนรวมประเภทต่าง ๆ จากเอกสารเผยแพร่หรือจากสื่อสิ่งพิมพ์ต่าง ๆ หากเป็นโครงการลงทุนออกใหม่ก็ต้องพิจารณาถึงประเภทของกองทุนว่าเป็นกองทุนเปิด หรือกองทุนปิด บริษัทหลักทรัพย์จัดการกองทุน (บลจ.) ใดเป็นผู้จัดตั้ง มีอายุโครงการหรือไม่ สามารถซื้อหน่วยลงทุนได้จากที่ไหน และในราคาเท่าไหร่ หรือหากเป็นกองทุนรวมที่มีอยู่ก่อน ให้พิจารณาดูว่าราคาซื้อขายหน่วยลงทุนนั้นเป็นเท่าใด กองทุนรวมมีขนาดใหญ่น้อยแค่ไหน สามารถซื้อขายหน่วยลงทุนผ่านช่องทางใดได้บ้าง และสภาพคล่องในการซื้อขายของหน่วยลงทุนนั้นเป็นอย่างไร

2. ศึกษาหาข้อมูลเชิงลึกของกองทุนรวมที่สนใจลงทุนอย่างละเอียดจากหนังสือชี้ชวน (Prospectus) โดยพิจารณาว่านโยบายการลงทุนเป็นอย่างไร ใครเป็นผู้จัดการกองทุน มูลค่าทรัพย์สินภายใต้การบริหารจัดการลงทุนมีมากน้อยเพียงใด ผลการดำเนินงานเป็นอย่างไร มีค่าธรรมเนียมและค่าใช้จ่ายอะไรบ้าง นอกจากนี้ มีนโยบายการจ่ายเงินปันผล เงื่อนไข และสิทธิประโยชน์อื่น ๆ ที่พึงได้รับมีอะไรบ้าง

3. พิจารณาให้แน่ใจว่ากองทุนรวมที่ต้องการซื้อ มีนโยบายการลงทุนที่ตรงกับเป้าหมายในการลงทุน ตลอดจนมีระยะเวลาในการลงทุนที่สอดคล้องกับความต้องการใช้เงินจริงในอนาคต

4. ขอคำแนะนำเพิ่มเติมจากผู้ทำหน้าที่ติดต่อกับผู้ลงทุน หรือตัวแทนจัดจำหน่ายหากมีข้อสงสัยที่เกี่ยวกับกองทุนรวม เพื่อให้เกิดความมั่นใจก่อนการตัดสินใจซื้อขายหน่วยลงทุน

5. เลือกบริษัทหลักทรัพย์จัดการกองทุน (บลจ.) ที่จะลงทุนด้วย หลักจากตัดสินใจเลือกประเภทกองทุนรวมได้แล้ว พบว่ามีหลาย บลจ. ที่เสนอขายกองทุน ผู้ลงทุนควรพิจารณาก่อนว่าควรเลือกซื้อกองทุนกับ บลจ. ใด เนื่องจากแต่ละแห่งมีจุดเด่นแตกต่างกัน ซึ่งสามารถพิจารณาได้จาก

5.1 บริการที่เสนอให้ เช่น มีกองทุนรวมรูปแบบที่ตรงกับความต้องการ มีการซื้อขายออนไลน์ มีบริการติดต่อฟรีให้สามารถสอบถามข้อมูล มีการจัดสัมมนาให้ความรู้ เป็นต้น

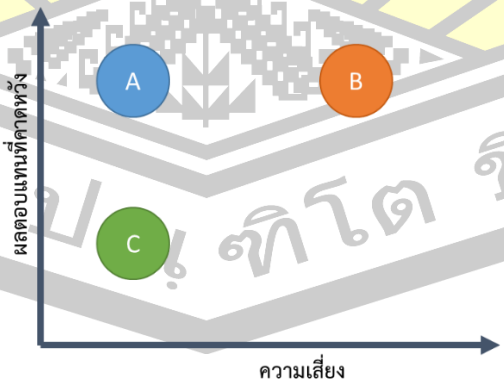
5.2 ความสะดวกสบาย ทั้งในเรื่องการติดต่อซื้อขาย ซึ่งควรดูว่าสถานการณ์ที่ติดต่อซื้อขายมีมากน้อยเพียงใด และในเรื่องการเข้าถึงข้อมูล เช่น นอกจาก บลจ. จะมีการประกาศราคาซื้อขายทางหนังสือพิมพ์แล้วยังมีการเผยแพร่ผ่านเว็บไซต์ด้วยหรือไม่ เป็นต้น

5.3 ความรู้ความสามารถของผู้ติดต่อผู้ลงทุน หรือเจ้าหน้าที่ บลจ. โดยผู้ติดต่อผู้ลงทุนหรือเจ้าหน้าที่ บลจ. ที่มีคุณภาพ ควรให้คำแนะนำเกี่ยวกับการลงทุนได้อย่างมีหลักการ มีความรู้ความเข้าใจในตลาดทุน ติดตามภาวการณ์ลงทุนอย่างต่อเนื่อง เป็นต้น

5.4 ความสามารถในการจัดการกองทุน ดูได้จากผลการดำเนินงานของกองทุนรวมภายใต้การจัดการของ บลจ. โดยควรเปรียบเทียบผลการดำเนินงานของกองทุนรวมที่มีนโยบายเดียวกัน ซึ่งจัดตั้งกองทุนในเวลาใกล้เคียงกันของหลาย ๆ บลจ.

พฤติกรรมของนักลงทุน

การพิจารณานักลงทุน ว่าจะได้รับผลตอบแทนจากการลงทุนได้มากน้อยเพียงใด ขึ้นอยู่กับลักษณะของนักลงทุนแต่ละคนที่สามารถยอมรับผลตอบแทนจากการลงทุน และความเสี่ยงที่แตกต่างกัน โดยทั่วไปแล้ว สามารถแบ่งพฤติกรรมการลงทุน (Investment behavior) หรือแนวคิดที่อธิบายถึงการตัดสินใจของผู้ลงทุนภายใต้ความเสี่ยงที่ไม่เท่ากัน โดยแบ่งนักลงทุนเป็น 3 ประเภท โดยพิจารณาตามภาพประกอบ 5 ดังนี้ (ณรงค์ จารขจรกุล, 2541) (Markowitz, 2015)



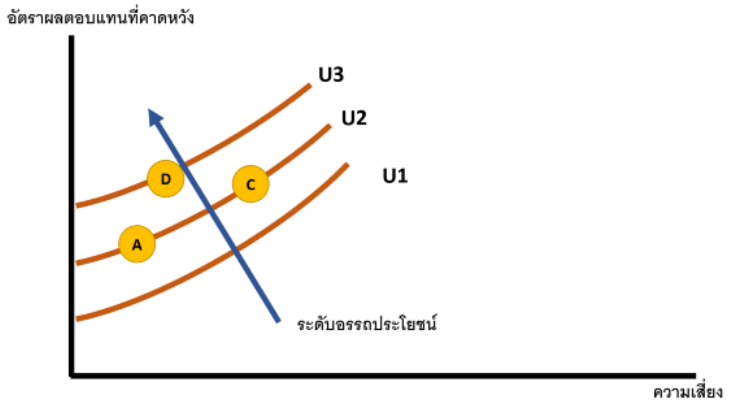
ภาพประกอบ 5 อัตราผลตอบแทนและความเสี่ยงของกลุ่มหลักทรัพย์

1. นักลงทุนที่กลัวความเสี่ยง (Risk Averter) ในระดับผลตอบที่คาดหวังเท่ากัน ผู้ที่กลัวความเสี่ยง จะเลือกลงทุนในกลุ่มสินทรัพย์ที่มีความเสี่ยงต่ำที่สุด จากภาพประกอบ 5 อัตราผลตอบแทนและความเสี่ยงของกลุ่มหลักทรัพย์ ผู้ลงทุนที่กลัวความเสี่ยง (Risk Averter) จะเลือกกลุ่มการลงทุน A เนื่องจากให้ผลตอบแทนที่สูงกว่า C ในระดับความเสี่ยงเดียวกัน และที่ผลตอบแทนที่คาดหวังเท่ากับ B แต่มีความเสี่ยงที่ต่ำกว่า

2. นักลงทุนที่ชอบความเสี่ยง (Risk Lover) ในระดับผลตอบที่ที่คาดหวังเท่ากัน ผู้ที่ชอบความเสี่ยง จะเลือกลงทุนในกลุ่มสินทรัพย์ที่มีความเสี่ยงมากที่สุด จากภาพประกอบ 5 อัตราผลตอบแทน และความเสี่ยงของกลุ่มหลักทรัพย์ หากเป็นนักลงทุนที่ชอบความเสี่ยงจะเลือกลงทุนในกลุ่มหลักทรัพย์ B มากกว่ากลุ่มหลักทรัพย์ A

3. นักลงทุนที่เป็นกลาง (Risk neutral) ผู้ลงทุนที่เป็นกลางจะพิจารณาทางเลือกการลงทุนจากผลตอบแทนเพียงอย่างเดียว โดยไม่สนใจความเสี่ยงเลย จากภาพประกอบ 5 อัตราผลตอบแทน และความเสี่ยงของกลุ่มหลักทรัพย์ ผู้ลงทุนที่เป็นกลางจะเลือก กลุ่มการลงทุน A หรือ B เพราะให้ผลตอบแทนที่เท่ากัน

อย่างไรก็ตาม พบว่า โดยทั่วไปแล้วนักลงทุนส่วนใหญ่เป็น นักลงทุนที่กลัวความเสี่ยง (Risk Averter) โดยหากความเสี่ยงเท่ากันนักลงทุนจะเลือกกลุ่มการลงทุนที่ให้ผลตอบแทนคาดหวังที่มากกว่า แต่หากระดับผลตอบแทนที่คาดหวังเท่ากันนักลงทุนจะเลือกกลุ่มการลงทุนที่มีความเสี่ยงที่น้อยกว่า ดังนั้น สิ่งที่น่าสนใจคือทำไมนักลงทุนส่วนใหญ่ถึงเป็นนักลงทุนที่กลัวความเสี่ยง ซึ่งสามารถอธิบายได้ด้วยทฤษฎีอรรถประโยชน์ (Utility theory)



ที่มา: https://miro.medium.com/max/700/1*fbtCQ7FqiWRJNydiUTeW7A.png

ภาพประกอบ 6 เส้นความพึงพอใจเท่ากัน

จากภาพประกอบ 6 เส้นความพอใจเท่ากัน U1, U2, U3 คือ เส้นความพอใจเท่ากัน ซึ่งทุกจุดบนเส้นมีระดับความพอใจในการลงทุนของผู้ลงทุนเท่ากัน ดังนั้น ผู้ลงทุนจะเลือกจุดใดก็ได้บนเส้นความพอใจเดียวกันเช่น A กับ C และเส้นที่อยู่สูงว่าจะมีระดับความพอใจที่สูงกว่าด้านล่าง เช่นที่จุด D ให้ความพอใจแก่ผู้ลงทุนสูงกว่าที่จุด A และ C โดยหากพิจารณาที่จุด A และจุด C จะเห็นได้ว่า เมื่อระดับความเสี่ยงสูงขึ้นผู้ลงทุนต้องการผลตอบแทนที่เพิ่มขึ้นเพื่อชดเชยระดับความเสี่ยงที่ผู้ลงทุนได้รับ

ทฤษฎีอรรถประโยชน์ข้างต้นเป็นพื้นฐานการอธิบายพฤติกรรมของนักลงทุนในทฤษฎีกลุ่มหลักทรัพย์โดยที่มีข้อสนับสนุน คือการที่ผู้ลงทุนซื้อกรมธรรม์ประกันชีวิตและประกันภัย

จากความหมายในการลงทุน จะเห็นว่านักลงทุนมักเลือกลงทุนในหลักทรัพย์ที่ให้ผลตอบแทนตามที่นักลงทุนคาดหวัง การที่ธุรกิจให้ผลตอบแทนจากการลงทุนที่ดีไม่ว่าจะอยู่ในรูปของเงินปันผล ดอกเบี้ย หรือกำไรจากการซื้อขายหลักทรัพย์ ย่อมทำให้นักลงทุนอยากลงทุนในหลักทรัพย์นั้นมากยิ่งขึ้น โดยพบว่าผลการดำเนินงาน หรือผลประกอบการที่ดีของธุรกิจมีความสัมพันธ์กับผลตอบแทนจากการลงทุนที่เพิ่มขึ้นด้วยเช่นกัน (Haowcharoen, 2016) ระบุได้ว่า ยิ่งธุรกิจสามารถให้ผลตอบแทนที่เป็นไปตามความคาดหวังของนักลงทุนได้มากเพียงใด ก็จะส่งผลให้นักลงทุนมีความต้องการลงทุนในหลักทรัพย์นั้นเพิ่มมากยิ่งขึ้น ทำให้เกิดปริมาณการซื้อขายหลักทรัพย์ที่เพิ่มขึ้น และส่งผลต่อราคาหลักทรัพย์ในที่สุด ดังนั้น อัตราส่วนวัดความสามารถในการทำกำไรจึงมีความสัมพันธ์กับราคาของหลักทรัพย์ที่สะท้อนให้เห็นว่า ธุรกิจมีความสามารถในการทำกำไรเท่าไร และมีผลการดำเนินงานเป็นไปตามที่นักลงทุนคาดหวังหรือไม่

ข้อมูลจากงบการเงินเป็นเพียงรายการและตัวเลข ซึ่งจะมีประโยชน์และมีคุณค่าก็ต่อเมื่อผู้ใช้ข้อมูลในงบการเงินสามารถนำความสำคัญของตัวเลขทางการเงิน โดยการวิเคราะห์งบการเงินเพื่อให้บรรลุวัตถุประสงค์ของตนในการได้มาซึ่งข้อมูล เพื่อนำไปใช้ในการตัดสินใจต้องเป็นไปอย่างมีหลักเกณฑ์และถูกต้องแม่นยำ และเป็นการลดความเสี่ยงจากการตัดสินใจที่ผิดพลาด (Hirunratsamee, 2015)

หลักการวิเคราะห์ข้อมูลและกระบวนการเรียนรู้ของเครื่อง (Machine learning)

การทำเหมืองข้อมูล (Data Mining)

เป็นกระบวนการในการสกัด (Extract) เพื่อค้นหารูปแบบ (Patterns) หรือความรู้ (Knowledge) จากฐานข้อมูลขนาดใหญ่ เพื่อให้ได้มาซึ่งสารสนเทศที่ต้องการ น่าสนใจหรือไม่คาดคิดมาก่อนและเป็นประโยชน์มาใช้สนับสนุนการตัดสินใจต่าง ๆ เปรียบเสมือนวิวัฒนาการหนึ่งในการ

จัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล โดยสามารถสรุปสาระสำคัญได้ (Agarwal, 2013) ดังนี้

1. กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

2. การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย

3. การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

4. การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

ตัวอย่างการนำเหมืองข้อมูลมาประยุกต์ใช้งาน เช่น การใช้สารสนเทศจากการทำเหมืองข้อมูลจากฐานข้อมูลประวัติคนไข้ เพื่อใช้วินิจฉัยผู้ป่วยหรือพยากรณ์โรคในทางการแพทย์ หรือการใช้สารสนเทศจากการทำเหมืองข้อมูลจากฐานข้อมูลสิเชื่อมาวิเคราะห์ในการปล่อยสินเชื่อให้ลูกค้ารายใหม่ในทางการธนาคาร หรือใช้สารสนเทศที่ได้จากการทำเหมืองข้อมูลจากฐานข้อมูลลายนิ้วมือหรือใบหน้ามาวิเคราะห์เพื่อหาเจ้าของลายนิ้วมือหรือเจ้าของใบหน้า

เห็นได้ว่า การทำเหมืองข้อมูล (Data Mining) อาจไม่ได้เน้นไปที่การสร้างระบบอัจฉริยะหรืออุปกรณ์เพื่ออำนวยความสะดวกแก่ผู้ใช้งาน แต่จะมุ่งเน้นไปที่การวิเคราะห์ข้อมูลให้ได้มาซึ่งสิ่งที่น่าสนใจ โดยใช้เทคนิควิธีการเรียนรู้ของเครื่อง (Machine Learning) เฉกเช่นเดียวกับปัญญาประดิษฐ์

การเรียนรู้ของเครื่อง (Machine Learning)

การทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยใช้ ข้อมูล หรือการใช้งานเสมือนเป็นสมองของปัญญาประดิษฐ์ (Artificial Intelligence (AI)) ใช้ การเรียนรู้ของเครื่อง ในการสร้างความฉลาด มักจะใช้เรียกแบบจำลองที่เกิดจากการเรียนรู้ของปัญญาประดิษฐ์โดยไม่ได้เกิดจากการเขียนโดยใช้มนุษย์ มนุษย์มีหน้าที่เขียนโปรแกรมให้ AI เรียนรู้จากข้อมูลเท่านั้น ที่เหลือเครื่องจะจัดการเองโดยการเรียนรู้ของเครื่อง เรียนรู้จากสิ่งที่เราส่งเข้าไปกระตุ้น แล้วจดจำเอาไว้เป็นมันสมอง ส่งผลลัพธ์ออกมาเป็นตัวเลขหรือ Code ที่ส่งต่อไปแสดงผลหรือก็คือการเรียนรู้ของเครื่อง เป็นปัญญาประดิษฐ์ชนิดหนึ่งที่ช่วยให้ Application ซอฟต์แวร์มีความแม่นยำมากขึ้นในการพยากรณ์ผลลัพธ์โดยไม่ต้องตั้งโปรแกรมอย่างชัดเจนให้ทำ ซึ่งอัลกอริทึมการเรียนรู้ของเครื่องใช้ข้อมูลประวัติเป็นข้อมูลรับเข้า (Input) เพื่อพยากรณ์ค่าข้อมูลส่งออกใหม่ (Output)

การเรียนรู้ของเครื่อง สามารถเอาไปใช้งานได้หลายรูปแบบ ต้องอาศัยกลไกที่เป็นโปรแกรม หรือเรียกว่าอัลกอริทึมที่มีหลากหลายแบบ โดยมีนักวิทยาศาสตร์ข้อมูลเป็นผู้ออกแบบ โดยหนึ่งในอัลกอริทึมที่ได้รับความนิยมสูง คือการเรียนรู้เชิงลึก (Deep Learning) ซึ่งถูกออกแบบมาให้ใช้งานได้ง่าย และประยุกต์ใช้ได้หลายลักษณะงาน

อย่างไรก็ตาม ในการทำงานจริงของนักวิทยาศาสตร์ข้อมูลจำเป็นต้องออกแบบตัวแปรต่าง ๆ และต้องหาอัลกอริทึมอื่น ๆ มาเป็นคู่เปรียบเทียบ เพื่อมองหาอัลกอริทึมที่เหมาะสมที่สุดในการใช้งานจริง ซึ่ง การเรียนรู้ของเครื่อง จะประกอบไปด้วยข้อมูล และเครื่องมือทางสถิติเพื่อพยากรณ์ผลลัพธ์ออกมาและถูกใช้เพื่อทำประโยชน์กับภายในองค์กรเชิงลึกต่อไป โดย การเรียนรู้ของเครื่องเกี่ยวข้องอย่างมากกับการทำเหมืองข้อมูลโดยทั่วไปจะเป็นการคอยสนับสนุนด้านการแนะนำ สำหรับวิธีการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เพื่อปรับปรุงตามประสบการณ์ของผู้ใช้

รูปแบบการเรียน สามารถแบ่งออกเป็น 3 ประเภทหลัก ๆ ได้แก่

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นกลุ่มของอัลกอริทึมที่เน้นสอนคอมพิวเตอร์ โดยการศึกษาจากข้อมูลตัวอย่าง อาทิ เมื่อต้องการให้คอมพิวเตอร์แยกภาพสุนัขออกจากภาพสัตว์ชนิดอื่นๆ จำเป็นต้องมีข้อมูลภาพตัวอย่างของสุนัข นำเข้าให้คอมพิวเตอร์รู้ว่า ถ้ารูปภาพมีลักษณะแบบที่นำเข้าไป ภาพนั้นคือภาพสุนัข หรือ ถ้าเราต้องการให้คอมพิวเตอร์ประมวลผลว่าคนที่มาขอสินเชื่อควรได้รับการอนุมัติหรือไม่ เราก็ต้องทำการหากลุ่มตัวอย่างเพื่อมาสอนคอมพิวเตอร์ว่าคนที่มีลักษณะแบบนี้ มีประวัติการเงินแบบนี้ควรอนุมัติหรือถ้ามีประวัติการเงินแบบนี้ไม่ควรอนุมัติ สรุปคือมีรูปแบบการประเมินชัดเจนว่าถ้าทำงานแล้วผลลัพธ์ได้แบบนี้คือถูกหรือผิด อ้างอิงจากตัวอย่างที่ได้ทำการสอนไป ซึ่งเราจะเรียกผลลัพธ์ของตัวอย่างที่ใช้ในการสอนว่า Label

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้ที่ไม่จำเป็นต้องมีค่าเป้าหมายของแต่ละข้อมูลตัวอย่าง ในระหว่างการเรียนรู้หรือก็คือการเรียนรู้แบบไม่มีการระบุผล (Target Variable) ที่ต้องการ ให้คอมพิวเตอร์หาความสัมพันธ์จากข้อมูลเอง แม้ว่าการเรียนรู้แบบไม่มีผู้สอนจะไม่ต้องการผู้สอน แต่ก็ต้องการแนวทางในการจัดกลุ่ม เช่น การจัดกลุ่มอาจจะจัดตามรูปทรง สี หรือวิธีการใช้งานของวัตถุต่าง ๆ ที่จะนำมาจัด เป็นต้น ดังนั้น หากไม่มีการให้แนวทางที่ชัดเจนว่า การจัดกลุ่มควรเป็นไปตามคุณลักษณะใด ทำให้การจัดกลุ่มอาจไม่ประสบความสำเร็จในแง่การนำมาใช้งานจริงก็ได้ ตัวอย่างเช่น การให้เด็กจัดของไว้บนชั้นวางของนั้น เด็กอาจจะจัดตามใจชอบและไม่เป็นหมวดหมู่ ทำให้ไม่สะดวกต่อการนำสิ่งของใหม่ ๆ เนื่องจากขั้นตอนการเรียนรู้แบบไม่มีผู้สอนจะมีการระบุกลุ่มของข้อมูลตัวอย่างก่อน เมื่อตัดสินใจได้แล้วว่า ข้อมูลใหม่มี

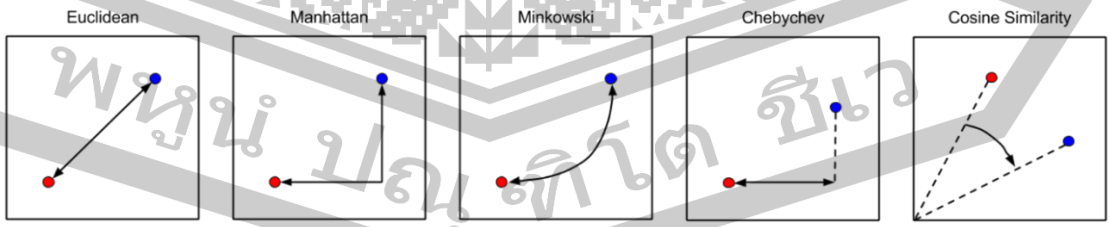
ลักษณะที่ควรจะจัดรวมเข้ากลุ่มใดหรืออาจเป็นกลุ่มใหม่ก็ได้ หลังจากนั้น จึงมีการปรับคุณลักษณะของกลุ่ม โดยการนำลักษณะของข้อมูลใหม่นี้มาช่วยกำหนดแนวทางการจัดด้วย

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เทคนิคการทำ AI ที่ใช้วิธีการให้รางวัล (Reward) หรือลงโทษ (Punishment) เพื่อที่จะขับเคลื่อนให้ Agent นั้น ๆ ไปในทิศทางเป้าหมายที่ระบุไว้ได้ ซึ่งการเรียนรู้แบบเสริมกำลังได้นำมาใช้เพื่อเทรนระบบที่สามารถเอาชนะ หากแต่ปัญหาคือ framework ของการเรียนรู้แบบเสริมกำลังนั้นจะต้องใช้เวลาอย่างมากในการไปให้ถึงจุดหมายที่ต้องการ ซึ่งมีโอกาสที่จะไม่มีความยืดหยุ่น และอาจจะไม่เสถียรเสมอไป

เทคนิคการแบ่งกลุ่มข้อมูลเคมีน (K-Means)

การแบ่งกลุ่ม K-means บางครั้งถูกเรียกว่าการวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) มีหลักการในการใช้งานคือ ใช้หลักการตัดแบ่ง (Partition) และแบ่งวัตถุออกเป็น k กลุ่ม โดยการแทนแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ซึ่งใช้จุดศูนย์กลางของกลุ่มในการวัดระยะห่างของตัวอย่างในกลุ่มเดียวกัน โดยชนิดของตัวแปรที่ใช้ในเทคนิค K-Means Clustering จะต้องเป็นตัวแปรเชิงปริมาณ

วิธีการวิเคราะห์ข้อมูลที่ใช้ใน Machine Learning หรือ Data Mining โดยจะแบ่งชุดข้อมูลออกเป็นกลุ่ม (Cluster) โดยนำข้อมูลที่มีคุณลักษณะเหมือนกัน หรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน ขั้นตอนวิธีที่ใช้ในการแบ่งกลุ่มจะอาศัยความเหมือน (Similarity) หรือ ความใกล้ชิด (Proximity) โดยคำนวณจากการวัดระยะระหว่างเวกเตอร์ของข้อมูลเข้า โดยใช้การวัดระยะแบบต่าง ๆ เช่น การวัดระยะแบบยูคลิด (Euclidean distance) หรือการวัดระยะแบบแมนฮัตตัน (Manhattan distance) เป็นต้น ตามภาพประกอบ 7

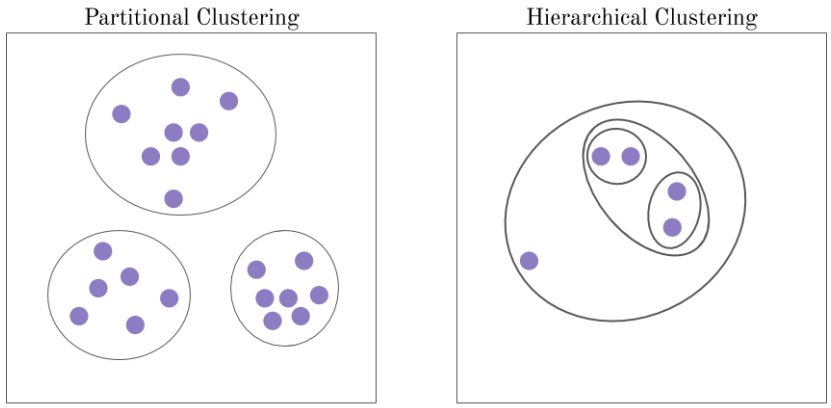


ที่มา: https://subscription.packtpub.com/book/big_data_and_business_intelligence/

ภาพประกอบ 7 ตัวอย่างวิธีการการวัดระยะ

การแบ่งกลุ่มข้อมูลจะแตกต่างจากการแบ่งประเภทข้อมูล (Classification) โดยจะแบ่งกลุ่มข้อมูลจากความคล้าย โดยไม่มีการกำหนดประเภทของข้อมูลไว้ก่อน ยกตัวอย่างอัลกอริทึมที่ใช้ในการแบ่งกลุ่ม เช่น การจัดกลุ่มเคมีน (K-means Clustering) การจัดกลุ่มไฮร์เออราชิคอล (Hierarchical Clustering) เป็นต้น การแบ่งกลุ่มข้อมูลอาจใช้เป็นขั้นตอนเบื้องต้นของการวิเคราะห์ข้อมูล เพื่อช่วยในการลดขนาดข้อมูลหรือคัดเฉพาะบางกลุ่มเพื่อทำการวิเคราะห์ต่อไปหรือแยกการวิเคราะห์ออกเป็นสำหรับแต่ละกลุ่ม โดยการแบ่งกลุ่มข้อมูล จัดอยู่ในประเภทการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) การแบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม การจัดระเบียบข้อมูลลงในการจัดกลุ่มที่สมเหตุสมผลถือเป็นหนึ่งในรูปแบบพื้นฐานของความเข้าใจและการเรียนรู้ (Ahmad, 2007) โดยการจัดกลุ่มถูกใช้อย่างกว้างขวางในการระบุกลุ่มลูกค้าในทางการตลาด (Customer Segmentation) การจัดกลุ่มลักษณะประชากรในทางสังคมศาสตร์ การค้นหาข้อมูลผิดปกติ (Outlier) การจัดกลุ่มเป็นเทคนิคแรก ๆ ที่วิเคราะห์ข้อมูลที่ยังไม่มีโครงสร้าง เพื่อค้นหากลุ่มทางธรรมชาติ (Natural Group)

ขั้นตอนวิธีในการแบ่งกลุ่มข้อมูล โดยทั่วไปแบ่งได้เป็น 2 ประเภทใหญ่ๆ คือ การแบ่งแบบเป็นลำดับขั้น (Hierarchical) และการแบ่งแบบตัดเป็นส่วน (Partition) การแบ่งแบบเป็นลำดับขั้นนั้นจะมีทำการแบ่งกลุ่มจากกลุ่มย่อยที่ถูกแบ่งไว้ก่อนหน้านั้นซ้ำหลายครั้ง ส่วนการแบ่งแบบตัดเป็นส่วนนั้น การแบ่งจะทำเพียงครั้งเดียว แสดงตามภาพประกอบ 8



ที่มา: https://miro.medium.com/

ภาพประกอบ 8 การแบ่งกลุ่มแบบ partitional และ hierarchical

โดยในงานวิจัยนี้ จะใช้การจัดระเบียบข้อมูลในการจัดกลุ่ม ถือเป็นหนึ่งในรูปแบบพื้นฐานของความเข้าใจและการเรียนรู้ โดยที่การวิเคราะห์กลุ่ม คือการศึกษาอย่างเป็นทางการของวิธีการและ

อัลกอริทึมสำหรับการจัดกลุ่ม หรือการจัดกลุ่มวัตถุตามลักษณะเฉพาะที่วัดได้ หรือรับรู้หรือความคล้ายคลึงกัน และหนึ่งในอัลกอริทึมการจัดกลุ่มที่ได้รับความนิยมและง่ายที่สุด คือ K-means แม้ว่าจะมีการนำเสนอ K-means มาเป็นเวลานาน แต่ยังคงใช้กันอย่างแพร่หลายจนถึงปัจจุบัน

เทคนิคการจำแนกข้อมูล (Classification)

เทคนิคการจำแนกข้อมูล จัดอยู่ในประเภทการเรียนรู้แบบ Supervised learning เป็นกลุ่มของ Algorithm ที่เน้นสอน Computer โดยการศึกษาจากข้อมูลตัวอย่าง อันมีรูปแบบการประเมินชัดเจนว่าถ้าทำงานแล้วผลลัพธ์ได้แบบนี้คือถูกหรือผิด อ้างอิงจากตัวอย่างที่ได้ทำการสอนไป ซึ่งเราจะเรียกผลลัพธ์ของตัวอย่างที่ใช้ในการสอนว่า Label โดยเทคนิคการจำแนกข้อมูล เป็นการพยายามที่จะพยากรณ์คำตอบที่เป็น Discrete Output หรือคำตอบที่ไม่ต่อเนื่องกันเช่น ตอบคำถามว่าเป็นสุนัขหรือแมว

โดยในงานวิจัยนี้ ทำการศึกษาอัลกอริทึมเกี่ยวกับการจำแนกข้อมูล ที่ใช้ในงานวิจัยทั้งหมด 5 อัลกอริทึม ซึ่งสามารถอธิบายได้ ดังนี้

เคเนียร์เรสเนเบอร์ (K-nearest Neighbors Algorithm: K-NN)

วิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุดนั้น โดย K-NN จัดอยู่ในลักษณะพิเศษคือ Lazy learning มาจากรูปแบบการจำแนกประเภทข้อมูลโดยไม่มีการสร้างแบบจำลองเตรียมไว้ล่วงหน้าเมื่อมีข้อมูลให้ที่ต้องการจำแนกประเภท เป็นเพียงการนำมาเทียบกับข้อมูลเดิมและพิจารณาความคล้ายคลึงกันของข้อมูลใหม่และข้อมูลเดิม (Hulett et al., 2012)

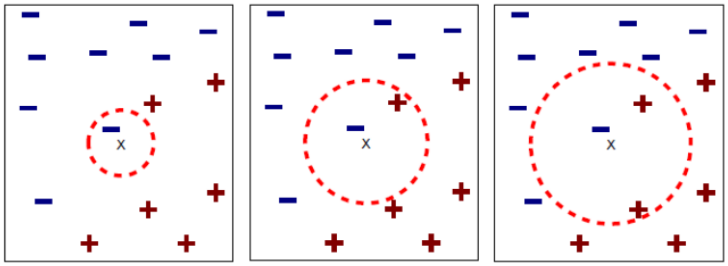
โดยมีขั้นตอนการจำแนกข้อมูลที่ใช้วิธีการหาระยะห่างระหว่างคุณลักษณะของแต่ละข้อมูล ซึ่งวิธีการจะเหมาสำหรับข้อมูลที่เป็นตัวเลข (วิภาวรรณ บัวทอง, 2557) ตามภาพประกอบ 9 ดังนี้

1. กำหนดค่า K โดยนิยมกำหนดให้เป็นเลขคี่ เพื่อให้ลดข้อผิดพลาดในการพยากรณ์
2. คำนวณระยะห่าง (Distance) ของข้อมูลที่ต้องการพิจารณากับชุดข้อมูลสอน โดยสามารถคำนวณได้จากสมการระยะทาง โดยยกตัวอย่างสมการของยูคลิด (Euclidean distance) ดังสมการที่ 1

$$\text{Dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (1)$$

3. จัดลำดับของระยะห่างจากน้อยไปมาก และเลือกชุดข้อมูลที่น้อยที่สุดตามจำนวน K

4. กำหนดให้คำตอบของข้อมูลที่ต้องการพยากรณ์ คือกลุ่มที่มีจำนวนมากที่สุดในกลุ่มของชุดข้อมูล K



ที่มา: https://miro.medium.com/max/2400/1*iGkSp2sfZpaLwiAWgPbTDQ.png

ภาพประกอบ 9 ตัวอย่างเทคนิค K-NN โดยการกำหนดค่า K ต่างๆ

โดย K-NN มีข้อดี ซึ่งหากเงื่อนไขในการตัดสินใจมีความซับซ้อน วิธีนี้สามารถนำไปสร้างแบบจำลองที่มีประสิทธิภาพได้ แต่มีข้อเสีย เช่น ใช้เวลาคำนวณนาน ถ้า attribute มีจำนวนมากจะเกิดข้อผิดพลาดในการคำนวณค่า และคำนวณค่าได้เฉพาะข้อมูลประเภท Nominal เช่น ข้อมูลเพศ ชายหญิง อาชีพ เป็นต้น

นาอีฟเบย์อัลกอริทึม (Naïve Bayes algorithm)

ความน่าจะเป็นแบบเบย์ เป็นทฤษฎีที่พูดถึงความน่าจะเป็นในการเกิดสิ่งหนึ่ง ก็ต่อเมื่ออีกสิ่งที่ได้เกิดขึ้น หรือที่เรียกกันว่า “Given” ซึ่งถ้าใครได้เรียนมา ลักษณะของการเขียนสัญลักษณ์ คือ P(A|B) อ่านว่า ความน่าจะเป็นของเหตุการณ์ A เมื่อเกิดเหตุการณ์ B แล้ว โดยการเรียนรู้แบบเบย์ (Bayesian Learning) เป็นการจำแนกประเภทรูปแบบหนึ่งที่อาศัยหลักการของความน่าจะเป็นเข้ามาช่วยในการหาคำตอบของประเภทตัวอย่างใหม่ (Friedman et al., 2001)

พื้นฐานคุณสมบัติและขั้นตอนวิธีการ Naïve Bayes โดยการเรียนรู้แบบเบย์ เป็นเทคนิคที่ใช้ทฤษฎีความน่าจะเป็น ตามกฎของเบย์ (Bayes’ Theorem) เพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior Knowledge) (นพมาศ ปักเข็ม, 2015) ได้แก่ ความน่าจะเป็นก่อนหน้าสำหรับสมมติฐานหนึ่ง ๆ ร่วมกับข้อมูล เช่น ความน่าจะเป็นที่สังเกตได้สำหรับสมมติฐานหนึ่ง ๆ เพื่อหาสมมติฐานที่ดีที่สุด เมื่อกล่าวถึงพื้นฐานนิยามแล้ว สามารถยกตัวอย่างวิธีการได้คือ จำนวนเหตุการณ์ที่เราสนใจ / จำนวนเหตุการณ์ทั้งหมดที่เป็นไปได้

ความน่าจะเป็น (Probability) ในทางคณิตศาสตร์จะเป็นค่าที่อยู่ในช่วงระหว่าง (Probability axioms) 0 (ไม่มีทางเป็นไปได้) และ 1 (เป็นไปได้อย่างแน่นอน) โดยค่าความน่าจะเป็นที่มีค่าระหว่าง 0 ถึง 1 หมายถึงมีโอกาสเป็นไปได้ทั้ง 2 กรณี คือ P (success) = จำนวนครั้งที่สำเร็จ / จำนวนค่าเป็นไปได้ทั้งหมด และ P (failure) = จำนวนครั้งที่ล้มเหลว / จำนวนค่าเป็นไปได้ทั้งหมด

โดยถ้าความน่าจะเป็นของเหตุการณ์ A ขึ้นอยู่กับเหตุการณ์ B เรียกความน่าจะเป็นนี้ว่า Conditional Probability ซึ่งการเรียนรู้แบบเบย์ อาศัยหลักการของการคำนวณความน่าจะเป็นของแต่ละสมมติฐาน โดยการเรียนรู้แบบเบย์เป็นการเรียนรู้เพิ่มได้ เนื่องจากตัวอย่างใหม่ที่ได้มา ถูกนำมาปรับเปลี่ยน แจกแจงซึ่งมีผลต่อการเพิ่มหรือลดความน่าจะเป็น ทำให้มีการเรียนรู้ที่เปลี่ยนไป วิธีการนี้ตัวแบบจะถูกปรับเปลี่ยนไปตามตัวอย่างใหม่ที่ได้ โดยผนวกรวมกับความรู้เดิมที่มี ซึ่งการพยากรณ์ค่าคลาสเป้าหมายของตัวอย่างใช้ความน่าจะเป็นมากที่สุดของทุกสมมติฐานจากทฤษฎีของเบย์ โดยสามารถคำนวณความน่าจะเป็นของสมมติฐานต่างๆ โดยสามารถเขียนอยู่ในรูปอย่างง่ายซึ่งกำหนดด้วยสมการที่ 2

$$P(A|B) = P(B|A) * P(A) / P(B) \quad (2)$$

โดย B แทนข้อมูลที่นำมาใช้ในการคำนวณการแจกแจงความน่าจะเป็น posteriori probability ของสมมติฐาน A คือ P(A|B) ตามทฤษฎี

P (A) คือ ความน่าจะเป็นก่อนหน้าของสมมติฐาน A

P (B) คือ ความน่าจะเป็นก่อนหน้าของชุดข้อมูลตัวอย่าง B

P (A|B) คือ ความน่าจะเป็นของ A เมื่อรู้ B

P (B|A) คือ ความน่าจะเป็นของ B เมื่อรู้ A

ซึ่งหากนำทฤษฎี Bayes มาใช้สำหรับเหตุการณ์ที่มีสมมุติฐาน (H) หลายๆ สมมุติฐานได้จากสมการที่ 3

$$P(H_i|E) = P(E|H_i) * P(H_i) / \sum_{n=1,k} P(E|H_n) * P(H_n) \quad (3)$$

$P(H_i|E)$ = ความน่าจะเป็นที่สมมุติฐาน Hi จะเป็นจริงภายใต้เหตุการณ์ E

$P(E|H_i)$ = ความน่าจะเป็นที่เหตุการณ์ E จะเป็นจริงภายใต้สมมุติฐาน Hi

$P(H_i)$ = ความน่าจะเป็นที่สมมุติฐาน Hi จะเป็นจริง

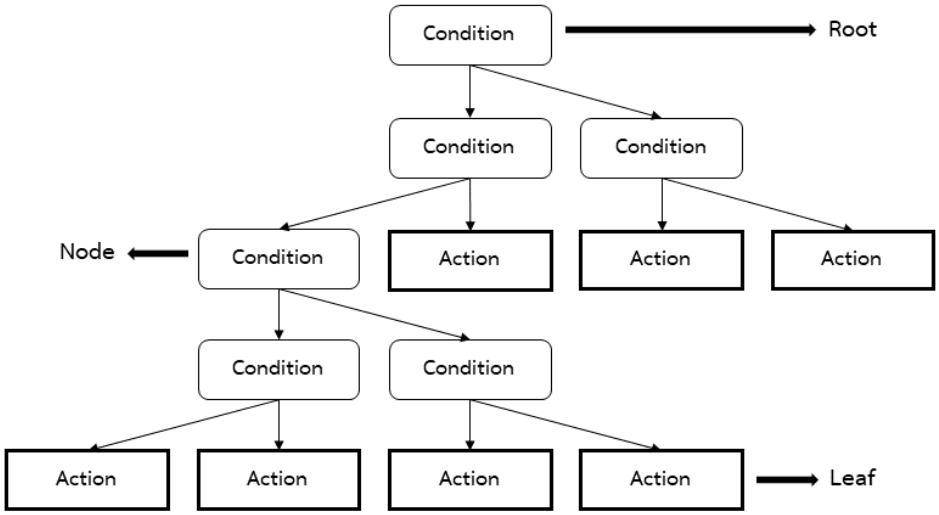
ต้นไม้ตัดสินใจ (Decision Tree Algorithm)

วิธีการต้นไม้ตัดสินใจ เป็นเทคนิคการสร้างแบบจำลองแบบพยากรณ์ล่วงหน้าโดยทั่วไป ต้นไม้ตัดสินใจ ถูกสร้างขึ้นด้วยวิธีอัลกอริทึมที่ระบุวิธีการแยกชุดข้อมูลตามเงื่อนไขต่าง ๆ หรืออธิบายได้ว่า วิธีการต้นไม้ตัดสินใจ เป็นรูปแบบของการวิเคราะห์หลายตัวแปร โดยการวิเคราะห์ตัวแปรหลายรูปแบบทำให้สามารถพยากรณ์ อธิบาย หรือจัดหมวดหมู่ผลลัพธ์ได้ ซึ่งความสามารถในการวิเคราะห์ตัวแปรหลายตัวของ ต้นไม้ตัดสินใจ ช่วยให้สามารถทำได้มากกว่าความสัมพันธ์แบบสาเหตุเดียว เพื่อค้นหา และอธิบายสิ่งต่าง ๆ ในบริบทของอิทธิพลหลายอย่าง ในการแก้ปัญหา เพราะผลลัพธ์ที่สำคัญเกือบทั้งหมดที่กำหนดความสำเร็จขึ้นอยู่กับปัจจัยหลายอย่าง

แบบจำลองหนึ่งสำหรับวิธีการ ต้นไม้ตัดสินใจ ถูกสร้างขึ้นโดย J. Ross Quinlan จาก University of Sydney อัลกอริทึมสำหรับการสร้าง Decision Tree ในปี ค.ศ.1975 ถูกเรียกว่า Dichotomiser 3 (ID3) อัลกอริทึมนี้ถูกสร้างขึ้นตามหลักการของ Occam ด้วยความคิดในการสร้าง Decision Tree ที่เล็กที่สุดและมีประสิทธิภาพมากที่สุดต่อมา Quinlan พัฒนาแบบจำลองต่อด้วยการสร้างอัลกอริทึม C4.5 (J48) ในปี ค.ศ. 1992 โดยเทคนิค Decision Tree หรือแบบจำลองทางคณิตศาสตร์เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองเพื่อพยากรณ์ในรูปแบบโครงสร้างต้นไม้ ซึ่งจัดอยู่ในประเภทการเรียนรู้แบบมีผู้สอน (Supervise Learning) (Quinlan, 2014)

การจำแนกข้อมูลด้วย Decision Tree จะเป็นกระบวนการสร้างต้นไม้เพื่อใช้ในการตัดสินใจจากข้อมูลที่เป็นหมวดหมู่ โดยมีส่วนประกอบตามภาพประกอบ 9 ดังนี้

1. โหนด (Node) คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าผลลัพธ์เป็นไปในทิศทางใดหรือค่าที่เป็นไปได้จากการทดสอบตัวแปรนั้น ๆ โดยโหนดที่อยู่สูงสุดจะถูกเรียกว่า Root Node
2. กิ่ง (Branch) คือ คุณสมบัติหรือทิศทางของโหนดต่าง ๆ ที่แตกออกมาจากค่าความเป็นได้นั้น ๆ โดยกิ่งจะเท่ากับคุณสมบัติของโหนด
3. ใบ (Leaf) คือ กลุ่มของผลลัพธ์จากการแยกแยะหรือจำแนกข้อมูลออกมา



ภาพประกอบ 10 ส่วนประกอบของต้นไม้ตัดสินใจ (Decision Tree)

การสร้างต้นไม้ตัดสินใจในรูปแบบ ID3 และ C4.5 ได้ทำการประยุกต์ใช้วิธีการ Greedy Approach ในการสร้างต้นไม้ภายใต้วิธีการแบบ Top-Down Recursive Divide-and-Conquer โดยทำการพิจารณาชุดข้อมูลสำหรับเรียนรู้ด้วยการแบ่งข้อมูลออกเป็นส่วนย่อย ๆ ในระหว่างกระบวนการ โดยการคัดเลือกคุณสมบัติ เป็นวิธีการที่ช่วยลดจำนวนตัวแปรที่ใช้ในการพยากรณ์ในแบบจำลอง เพื่อเลือกคุณสมบัติที่ดีที่สุดเพียงคุณสมบัติเดียวหรือเลือกกลุ่มคุณสมบัติที่มีความสำคัญต่อการพยากรณ์ กระบวนการคัดเลือกคุณสมบัติเป็นกระบวนการที่สำคัญในการเตรียมข้อมูลสำหรับการทำเหมืองข้อมูลเพื่อให้การพยากรณ์ของแบบจำลองมีประสิทธิภาพมากขึ้นและในวิธีการต้นไม้ตัดสินใจเนื่องจากมีการลดจำนวนคุณสมบัติที่ไม่จำเป็นต่อการพยากรณ์ หรือทำให้ค่าของการพยากรณ์ผิดพลาดออกไป โดยการคัดเลือกคุณสมบัติสามารถแบ่งประเภทได้ 2 แบบคือ

1. การคัดเลือกสับเซตจากคุณสมบัติทั้งหมด (Feature Subset Selection) คือการคัดเลือกเอาเพียงบางสับเซตของคุณสมบัติ จากจำนวนคุณสมบัติทั้งหมดโดยสับเซตที่เลือกมีแนวโน้มที่จะทำให้ประสิทธิภาพของการพยากรณ์จากแบบจำลองดีขึ้น

2. การจัดลำดับคุณสมบัติ (Feature Ranking) คือการคำนวณคะแนนของแต่ละคุณสมบัติ (Feature Score) แล้วทำการเรียงลำดับของแต่ละคุณสมบัติตามคะแนนที่ได้โดยเรียงจากมากไปน้อย ยกตัวอย่างเทคนิคการจัดลำดับคุณสมบัติที่ใช้ในการสร้าง ต้นไม้ตัดสินใจ ได้แก่ การจัดลำดับคุณสมบัติแบบ Information Gain (IG) มีหลักการพื้นฐานจากการสุ่มตัวอย่าง (Entropy) โดยค่าที่ได้จาก Information Gain คือค่าของความต่างระหว่างตัวแปร X ที่เป็นตัวแปรเป้าหมาย

(Target Variable) กับตัวแปรอิสระ (Independent Variable) ซึ่งลักษณะของ Information Gain จะทำการลด Entropy ของตัวแปรเป้าหมาย X โดยการเรียนรู้จากสถานะของ ตัวแปรอิสระ การคำนวณของเทคนิค Information Gain นั้นจะพิจารณาระหว่างคุณสมบัติ X หมายถึงตัวแปรเป้าหมายกับคุณสมบัติที่เป็นคลาสของข้อมูล Y หรือตัวแปรอิสระ จากนั้นทำการวิเคราะห์ได้จากจากความน่าจะเป็นที่เกิดขึ้นระหว่างค่าของคุณสมบัติ X กับค่าของคุณสมบัติY ว่ามีมากน้อยเพียงใด หากมีความน่าจะเป็นเกิดขึ้นน้อยก็จะทำการให้คะแนนของคุณสมบัติ X น้อยตามไปด้วย โดยสามารถเขียนเป็นสมการที่ 4

$$IG(X; Y) = H(Y) - H(Y|X) \quad (4)$$

เมื่อ H (Y) คือ ค่าความน่าจะเป็นจากการสุ่มตัวอย่างของ Y

H (Y|X) คือ ค่าความน่าจะเป็นจากการสุ่มตัวอย่างของ Y เทียบกับ X

G (Y; X) คือ ค่าของคะแนนจากการสุ่มตัวอย่างที่คำนวณได้โดยจะมีตั้งแต่ 0 ถึง 1

ค่าของ Y คือ ค่าคุณสมบัติที่เป็นคลาสของข้อมูลตั้งแต่ $\{ Y_1, Y_2 , \dots , Y_n\}$

ค่าของ X คือ ค่าคุณสมบัติอื่น ๆ ที่ไม่ใช่คลาส $\{ X_1, X_2 , \dots , X_n\}$

โดยที่ค่าของ H (Y) และ H (Y|X) คำนวณมาจาก

$$H(Y) = \sum (i=1, i=k)\ P(Y = y_i) * \log 2\ P(Y = y_i)$$

$$H(Y|X) = \sum (i=1, i=k)\ P(X = x_i) * H(Y|X = x_i)$$

เมื่อ $P(Y = y_i)$ คือ ค่าความน่าจะเป็นตั้งแต่ Y_1 จนถึง Y_k

$P(X = x_i)$ คือ ค่าความน่าจะเป็นตั้งแต่ X_1 จนถึง X_k

กฎการอุปนัย (Rule Induction Algorithm)

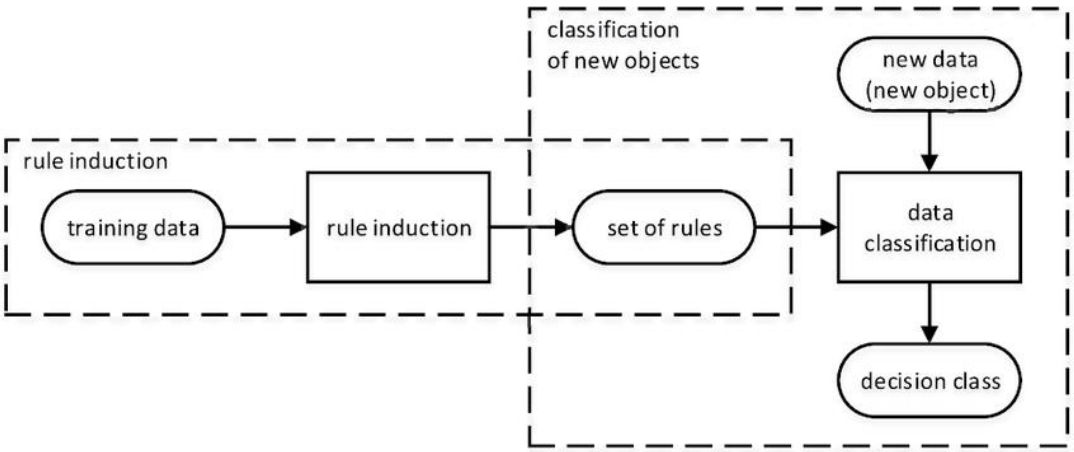
กฎการอุปนัย คือการสร้างอัลกอริทึมโดยไม่ต้องเขียนโปรแกรมโดยมนุษย์ แต่เป็นการวิเคราะห์โครงสร้างข้อมูลที่มีอยู่ หรือวิธีการดึงเอาชุดของกฎเกณฑ์ต่าง ๆ มาเพื่อจัดแบ่งเงื่อนไขหรือกรณีโครงสร้างต้นไม้สามารถสร้างชุดของกฎต่าง ๆ บางครั้งเรียกวิธีการแบบนี้ว่าการสร้างกฎใหม่

จากตัวอย่าง (ฝนทิพย์ คูณแก้ว, 2555) แต่ยังมีความหมายที่แตกต่างกันออกไป เนื่องจากวิธีการอุปนัยจะสร้างชุดของกฎที่เป็นอิสระ ซึ่งไม่จำเป็นต้องอยู่ในรูปโครงสร้างของต้นไม้ เพราะตัวสร้างการอนุมานกฎ (rule induction) ไม่ได้บังคับการแตกข้อมูลเป็นแต่ละระดับ แต่อาจจะสามารถค้นหารูป (Pattern) ที่แตกต่างกันได้ และบางครั้งอาจดีกว่าสำหรับการจัดแบ่ง Class ของผลลัพธ์

ในกรณีที่ง่ายที่สุด กฎจะแสดงด้วยคำสั่ง If-Then ที่สร้างขึ้นด้วยอัลกอริทึม ID3 สำหรับการเรียนรู้แผนผังการตัดสินใจ (Freitas, 2002) ซึ่งจะเป็นการฝึกข้อมูลรับเข้า และสร้างกฎโดยแบ่งพาร์ติชั่นตารางด้วยการวิเคราะห์กลุ่ม (Cluster Analysis) ทางเลือกอื่นที่เป็นไปได้ที่ดีกว่าอัลกอริทึม ID3 คือการเขียนโปรแกรมทางพันธุกรรมซึ่งจะพัฒนาโปรแกรมจนกว่าจะเหมาะสมกับข้อมูล

โดยในเอกสารแรพพิทไมเนอร์ สตูดิโอ (Rapidminer Studio) ได้อธิบายไว้ว่า กฎการอุปนัย (Rule Induction) จะทำงานคล้ายกับกฎประพจน์ (Propositional rule) ที่เป็นการตัดแต่งกิ่งที่เพิ่มขึ้นซ้ำ ๆ เพื่อลดข้อผิดพลาดหรือ Repeated Incremental Pruning to Produce Error Reduction (RIPPER) (Cohen, 1995) เริ่มต้นด้วยคลาสที่มีการแตกกิ่งน้อยกว่า และตัดกฎจนไม่มีตัวอย่างที่มีอัตราความผิดพลาดมากกว่า 50% ในระยะถัดมา กฎแต่ละข้อจะเพิ่มเงื่อนไขลงในกฎจนกว่าจะสมบูรณ์ เช่น มีค่าความถูกต้อง 100% โดยในขั้นตอนนี้ จะพยายามทุกค่าที่เป็นไปได้ของแต่ละ Attribute และเลือกเงื่อนไขที่มีค่า Information Gain (IG) มากที่สุด ในระยะการแบ่งกิ่ง สำหรับแต่ละกฎจะถูกตัดแต่งด้วยเมตริกการตัดแต่งกิ่ง p/(p+n)

กฎการอุปนัย มักจะถูกเปรียบเทียบกับ ต้นไม้ติดสินใจ แต่ กฎการอุปนัย มีข้อได้เปรียบในเรื่องการทำความเข้าใจผลลัพธ์ แสดงได้ในรูปแบบตรรกะ ข้อเสียที่สำคัญของ กฎการอุปนัย คือการปรับขนาดได้ไม่ดีกับขนาดชุดการฝึก และมีปัญหากับข้อมูลที่รบกวน โดยอัลกอริทึม RIPPER สามารถลดข้อเสียเหล่านี้ได้ค่อนข้างมาก ซึ่งปัญหาหลักของ ต้นไม้ติดสินใจ คือการ Overfitting หรือการที่แบบจำลองตอบสนองต่อการรบกวน (Noise) จำนวนมาก จนเริ่มเรียนจากการรบกวนและรายละเอียดของข้อมูลที่ไม่ถูกต้อง แล้วแบบจำลองของจะไม่เหมาะสมสำหรับการพยากรณ์ข้อมูล เช่น แบบจำลองทำงานได้ดีในชุดการฝึก แต่ทำงานได้ไม่ดีในชุดการตรวจสอบ โดยแสดงกระบวนการตามภาพประกอบ 11



ที่มา: https://www.researchgate.net/figure/Rule-induction-for-data-classification_fig1_268686483

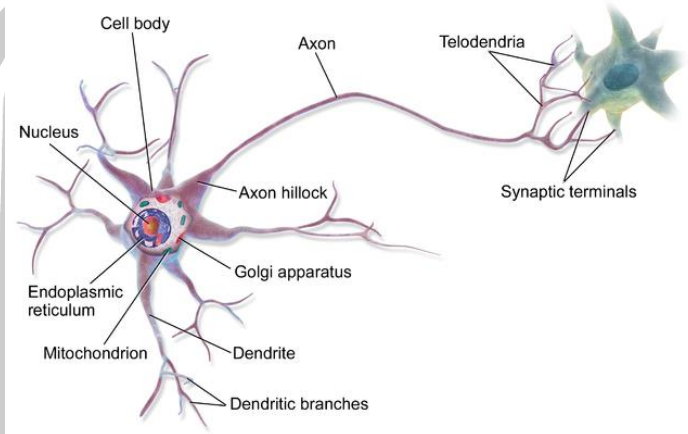
ภาพประกอบ 11 ขั้นตอนของกฎการอุปนัย

โครงข่ายประสาทเทียม (Neural Network Algorithm)

โครงข่ายประสาทเทียม เป็นศาสตร์สาขาหนึ่งของเทคโนโลยี ปัญญาประดิษฐ์มีแนวคิดตั้งแต่ปี 1943 คือ การพัฒนาเครือข่ายของ Algorithm ให้ทำงานในรูปแบบโครงสร้างและการทำงานของการประมวลผลเหมือนกับสมองของสิ่งมีชีวิตได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ซึ่งมีปรับเปลี่ยนตัวเองต่อการตอบสนองของอินพุตตามกฎของการเรียนรู้ (Learning Rule) หลังจากที่โครงข่ายได้เรียนรู้สิ่งที่ต้องการแล้ว โครงข่ายนั้นจะสามารถทำงานที่กำหนดไว้ได้โครงข่ายประสาทเทียมได้ถูกพัฒนาคิดค้นจากการทำงานของสมองมนุษย์โดยสมองมนุษย์ประกอบไปด้วยหน่วยประมวลผลเรียกว่า เซลล์ประสาท หรือ Neuron ซึ่งจำนวนนิวรอลในสมองมนุษย์มีอยู่ประมาณและมีการเชื่อมต่อกันอย่างมากมายหรือข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน ทำให้สมองมนุษย์จึงสามารถกล่าวได้ว่าเป็นคอมพิวเตอร์ที่มีการปรับตัวเอง (Adaptive) ไม่เป็นเชิงเส้น (Nonlinear) และทำงานแบบขนาน (Parallel) ในการดูแลจัดการการทำงานร่วมกันของนิวรอนในสมอง การคำนวณเชิงนิวรอลเป็นการคำนวณที่เลียนแบบมาจากการทำงานของสมองมนุษย์นั่นเอง (Anthony & Bartlett, 2009)

โดยโครงข่ายประสาทเทียมได้ถูกพัฒนาขึ้นโดยอาศัยหลักการทำงานของสมองมนุษย์ซึ่งสมองประกอบด้วยหน่วยประมวลผลพื้นฐานที่เรียกว่า Neuron ภายในสมองประกอบด้วยนิวรอลจำนวนมหาศาลและมีจุดต่อจำนวนโครงข่ายประสาทประกอบขึ้นด้วยส่วนสำคัญ 3 ส่วน (ชิตพงษ์ กิตตินราดร, 2563) คือ ใยประสาท (Dendrite) ตัวเซลล์ (Soma) และแกนประสาท (Axon) ในแต่

ละโครงข่ายประสาทจะเชื่อมต่อกันโดยจุดประสานประสาท (Synapse) ซึ่งสามารถเปลี่ยนค่าความต้านทานได้ตามสัญญาณที่ส่งระหว่างกันของเซลล์ประสาท การส่งสัญญาณระหว่างเซลล์ประสาททำได้โดยการถ่ายเทสารประกอบโซเดียมและโพแทสเซียม ตามภาพประกอบ 12



ที่มา: https://commons.wikimedia.org/w/index.php?curid=28761830

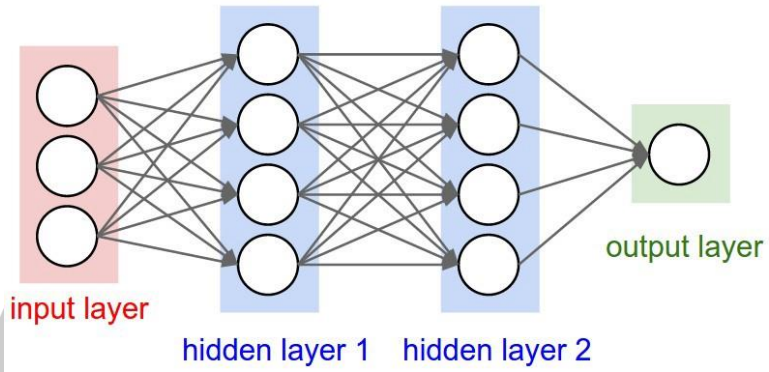
ภาพประกอบ 12 ส่วนประกอบการทำงานของสมองมนุษย์

ส่วนประกอบของโครงข่ายประสาทเทียม ตามภาพประกอบ 12 ได้แก่

1. Input Layer มีหน้าที่ในการรับข้อมูลเข้ามาในโครงข่ายประสาทโดย Input Layer จะเพียงชั้นเดียวเท่านั้นและมีหน้าส่งข้อมูลไปยังชั้นถัดไป (Hidden Layer) โดยจำนวนของโหนดขึ้นอยู่กับจำนวนของ input ว่ามีข้อมูลอะไรบ้างที่นำเข้ามาคิดในแบบจำลอง

2. Hidden Layer มีหน้าที่รับข้อมูลจาก Layer ก่อนหน้า โดย Hidden Layer สามารถมีจำนวนมากกว่า 1 ได้ โดยพื้นฐาน หากยิ่งต้องการความแม่นยำที่มากขึ้นเราก็จะเพิ่มจำนวนชั้นของ Hidden Layer และจำนวน Neurons ให้มากขึ้นก็จะช่วยได้ซึ่งไม่เสมอไป โดย Hidden Layer เป็นชั้นที่อยู่ระหว่างกลาง ซึ่งจะมีผลอย่างมากต่อประสิทธิภาพในการเรียนรู้ของแบบจำลอง โดยจะมีชั้นก็ได้แล้วแต่ผู้สร้างและแต่ละชั้นจะมีจำนวนของ Neuron เท่าไหร่ก็ได้เช่นกัน

3. Output Layer มีหน้าที่รอรับค่าจาก Hidden Layer โดยในชั้น output นั้น แต่ละ neurons จะมีค่าน้ำหนัก (Weight) ของคลาสอยู่ เป็นชั้นที่นำเอาข้อมูลจากการคำนวณไปใช้ จำนวนของโหนดในชั้นนี้ ขึ้นอยู่กับรูปแบบของ output ที่ต้องการ



ที่มา: https://miro.medium.com/max/791/0*CXpXUW-LyfJEA7Fg

ภาพประกอบ 13 ส่วนประกอบของ Neural Network

โดยการประมวลผลต่างๆ เกิดขึ้นในหน่วยประมวลผลย่อย เรียกว่า โหนด (Node) ซึ่งโหนดเป็นการจำลองลักษณะการทำงานมาจากเซลล์การส่งสัญญาณ ระหว่างโหนดที่เชื่อมต่อกันจำลองมาจากการเชื่อมต่อของใยประสาท และแกนประสาทในระบบประสาทของสมองมนุษย์ ภายในโหนด จะมีฟังก์ชันกำหนดสัญญาณส่งออกที่เรียกว่า ฟังก์ชันการแปลง (Transfer Function) ซึ่งทำหน้าที่เปรียบเสมือนกระบวนการทำงานในเซลล์ (อำภา สาระศิริ, 2559) ประกอบด้วย 6 องค์ประกอบ ตามภาพประกอบ 13 ดังนี้

1. ข้อมูลอินพุต (Input) เป็นข้อมูลที่เป็นตัวเลข แต่หากเป็นข้อมูลเชิงคุณภาพต้องทำการแปลงข้อมูลให้อยู่ในรูปเชิงปริมาณที่โครงข่ายประสาทเทียมยอมรับได้

2. ข้อมูลเอาต์พุต (Output) คือ ผลลัพธ์ที่เกิดขึ้นจริงจากกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม

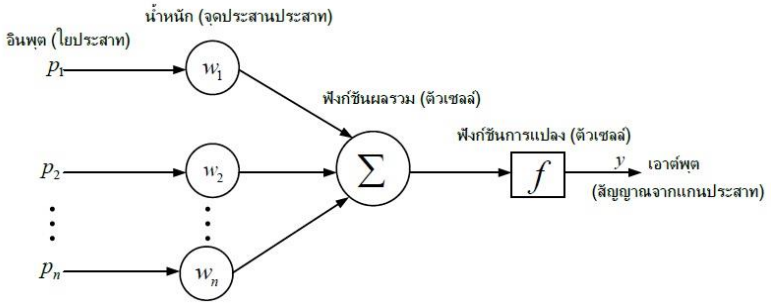
3. Neurous โดยข้างใน Neuron จะต่างกันตาม Layer ที่อยู่ ได้แก่ Input ข้างในตัวมันจะมีข้อมูลที่รับมา Hidden Layer จะมีสมการที่ช่วยในการคำนวณเพื่อพยากรณ์ว่าเป็นคลาสอะไร หรือคำนวณแบบถดถอย (Regression) และ Output ก็จะเป็นตัวที่บ่งบอกว่าเป็นคลาสอะไร

4. ค่าน้ำหนัก (Weights) คือ สิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม หรือ ค่าความรู้ (Knowledge) ค่านี้จะถูกเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่น ๆ ที่อยู่ในรูปแบบเดียวกัน โดยทุก ๆ neurons ใน Hidden Layer จะต้องมี Bias เข้าไปคำนวณเพื่อให้ขอบเขตการตัดสินใจ (Decision Boundary) โดย Bias คือตัวเลขที่บวกเข้าไปเพื่อปรับให้ค่าที่คำนวณออกมาถูกต้องมากขึ้น โดยค่านี้จะไม่ขึ้นอยู่กับ X ใด ๆ เป็นตัวแทนของ ธรรมชาติของข้อมูล ส่วน Weight จะเป็นน้ำหนักซึ่งส่งผลทุกๆ Neurons มีค่า Output ที่ไม่เท่ากัน ทำให้แต่ละคลาสมีน้ำหนักไม่เท่ากัน

เวลาคำนวณว่าเป็นคลาสไหน ทำให้สามารถแยกว่าข้อมูลนี้เป็นคลาสอะไรได้ด้วยการดูตัวเลขที่ Output หรือจะเรียกว่า ค่าความรู้ (knowledge) ค่าที่ถูกเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่น ๆ ที่อยู่ในรูปแบบเดียวกัน

5. ฟังก์ชันผลรวม (Summation function: S) เป็นผลรวมของข้อมูลป้อนเข้าและค่าน้ำหนัก

6. ฟังก์ชันการแปลง (Transfer Function) เป็นส่วนที่ทำหน้าที่รวมค่าเชิงตัวเลขจากเอาต์พุตของนิวรอล แล้วทำการตัดสินใจว่าจะส่งสัญญาณเอาต์พุตออกไปในรูปใด ฟังก์ชันการแปลงสามารถเป็นได้ทั้งแบบเชิงเส้นหรือไม่เป็นเชิงเส้น การเลือกใช้ฟังก์ชันการแปลงจะขึ้นอยู่กับลักษณะของระบบ ที่นำเอาโครงข่ายประสาทเทียมไปประยุกต์ใช้



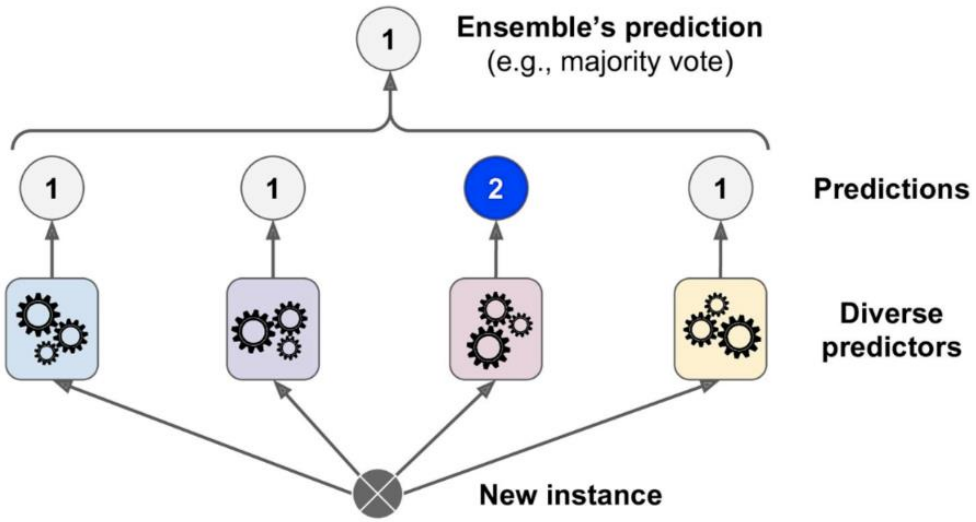
ที่มา: https://www.mut.ac.th/upload/from_eng/a96_4.jpg

ภาพประกอบ 14 กระบวนการประมวลผลของโครงข่ายประสาทเทียมและองค์ประกอบ

เทคนิคการเรียนรู้แบบกลุ่ม

เทคนิคการเรียนรู้แบบกลุ่ม (Ensemble Method) เป็นเทคนิคของการเรียนรู้ของเครื่องขั้นสูงที่ช่วยเพิ่มประสิทธิภาพการพยากรณ์ผล (Prediction Performance) โดยการใช้เทคนิคที่ใช้แบบจำลอง Classification หลายๆ แบบจำลอง มาช่วยในการหาคำตอบ ซึ่งเป็นเทคนิคที่มีประสิทธิภาพสูง (Dietterich, 2000) (Manish K., 2012) โดยมีอัลกอริทึมที่พัฒนาบนแนวคิดเทคนิค Ensemble ที่ถูกใช้กันมาก มีดังนี้

1. Vote Ensemble เป็นการใช้ Training Data ชุดเดียวกันแต่สร้างแบบจำลองด้วยเทคนิคหลากหลายต่าง ๆ จึงเป็นการเลือกที่จะสร้างแบบจำลองด้วยเทคนิคการจำแนกข้อมูล (Classification) ที่ต่างกัน คือหลังจากได้แบบจำลองมาชุดหนึ่งแล้วจะทำการนำไปพยากรณ์ข้อมูลและนำคำตอบมารวมกันเพื่อดูว่าคำตอบไหนเหมาะสมที่สุด โดยใช้วิธีการโหวต (Vote) เลือกคำตอบที่ตอบตรงกันมากที่สุด ซึ่งเป็นเทคนิคที่ใช้ในงานวิจัยครั้งนี้ แสดงตามภาพประกอบ 15

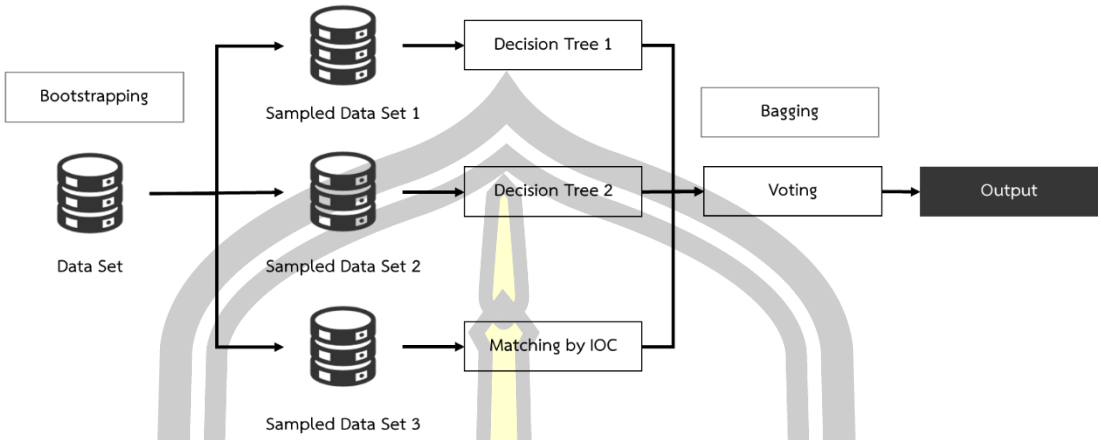


ที่มา: https://medium.com/

ภาพประกอบ 15 เทคนิคการเรียนรู้แบบกลุ่มด้วย Vote

2. Bootstrap Aggregating (Bagging) เป็นการสุ่ม Training Data ให้เป็นหลายชุด แต่สร้างแบบจำลองด้วยเทคนิคเดียวกันทั้งหมด เช่น ใช้เทคนิค ต้นไม้ตัดสินใจ หรือ เทคนิค โครงข่ายประสาทเทียมทั้งหมด ซึ่งมีข้อดีคือ ค่าความแปรปรวน (Variance) จะลดลง ค่าความเที่ยงตรง (Precision) เพิ่มขึ้น เนื่องจากค่าพยากรณ์ที่ได้จากแต่ละต้นไม้ จะถูกเฉลี่ย อาจสรุปได้ว่ายิ่งสร้างต้นไม้มาก ค่าความแปรปรวนก็ยิ่งลดลง ข้อเสียคือยังคงมีค่า Bias อยู่

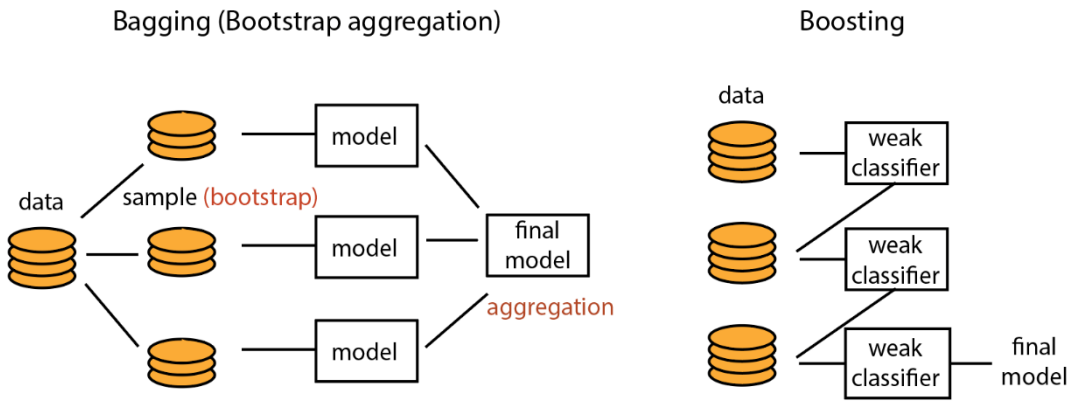
ยกตัวอย่าง Random Forest เป็นการสุ่มเลือก Attribute ต่าง ๆ ออกมาเป็นหลายชุด และสร้างแบบจำลองด้วยเทคนิค ต้นไม้ตัดสินใจ หลายๆ แบบจำลอง ตั้งแต่ 10 แบบจำลองถึงมากกว่า 1,000 แบบจำลอง มีวิธีการที่คล้ายกับ Bagging แต่เพิ่มการสร้างความหลากหลายของแบบจำลองด้วยการสุ่ม Attribute แทนที่จะเป็นการสุ่มเฉพาะข้อมูลตัวอย่างเพียงอย่างเดียว และเทคนิคที่ใช้ในการสร้างแบบจำลองก็เป็นเพียงแค่ ต้นไม้ตัดสินใจอย่างเดียว ซึ่งมีข้อดีคือลดค่า Correlation ระหว่างต้นไม้ อธิบายหลักการตามภาพประกอบ 16 ดังนี้



ภาพประกอบ 16 แสดงหลักการทำ Random Forest

เริ่มจากการสุ่มตัวอย่างชุดใหม่จากจำนวนข้อมูลทั้งหมด (Bootstrapping) ใช้วิธีสุ่มแบบแทนที่ให้ได้ออกมาชุดจำนวนที่ต้องการโดยมีลักษณะไม่เหมือนกันมาสร้างแบบจำลองต้นไม้ตัดสินใจ (Breiman, 2001) สำหรับแต่ละชุดขอมูลเพื่อพยากรณ์ชุดข้อมูลใหม่ โดยการสร้างต้นไม้ตัดสินใจ โดยนิยาม Entropy หรือปริมาณที่บอกถึงความไม่เป็นระเบียบของข้อมูลในเซต หลักจากนั้น ทำการหาผลลัพธ์ (Aggregation) จากแต่ละแบบจำลอง (Bagging) เช่น การโหวต (Voting) ในกรณีเทคนิคการจำแนกหรือค่าเฉลี่ย (Mean) สำหรับการวิเคราะห์การถดถอย (Regression Analysis)

3. Boosting คือการนำ Weak Classifier หรือ Classifier ที่มีความแม่นยำต่ำ มาพยากรณ์ข้อมูลที่มี จากนั้น ให้ Weak Classifier ตัวใหม่มาแก้ไข Error ที่มี โดยผลรวมของ Classifier จะเกิดเป็น Classifier ใหม่ขึ้นมา โดยจะทำแบบนี้ไปเรื่อย ๆ จนได้แบบจำลองที่ดีที่สุดจากผลรวมของ Classifier อันมีข้อเสียคือ ต้องประมวลผลข้อมูลหลายครั้ง และเป็นลำดับกว่าจะได้แบบจำลองที่ต้องการ ต่างจาก Bagging ที่สามาถสุ่มข้อมูลได้แล้วฝึกแบบจำลองได้พร้อมกัน โดยแสดงเทคนิคการเรียนรู้แบบกลุ่มด้วย Bagging และ Boosting ตามภาพประกอบ 17



ที่มา: https://tupleblog.github.io/bagging-boosting/

ภาพประกอบ 17 เทคนิคการเรียนรู้แบบกลุ่มด้วย Bagging และ Boosting

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยดำเนินการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ในด้านการลงทุนและการใช้กระบวนการทางด้านการทำเหมืองข้อมูล และการใช้การเรียนรู้ของเครื่อง โดยมีรายละเอียดและแสดงผลสรุปตามตาราง 2 โดยสามารถอธิบายแนวทางการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้

ปัจจุบันจำนวนข้อมูลมหาศาลในฐานข้อมูลของธนาคารหรือสถาบันการเงิน ได้เพิ่มมากขึ้นอย่างรวดเร็ว ข้อมูลขนาดใหญ่นี้ส่วนใหญ่เป็นข้อมูลที่สำคัญและมีมูลค่ามหาศาลสำหรับธุรกิจ เช่น ข้อมูลประชากรของลูกค้า (Customer's Demographic) และข้อมูลพฤติกรรมของลูกค้า (Customer's Behavior) (Zentut website, 2016) เป็นต้น การทำเหมืองข้อมูล หรือการใช้เทคนิคการเรียนรู้ของเครื่อง ได้ถูกนำมาใช้ในหลายภาคธุรกิจ โดยฉพาะธุรกิจทางด้านการเงินหรือการลงทุน โดยการนำข้อมูลมาสำรวจและวิเคราะห์ สกัดและจัดเก็บข้อมูลที่เป็นประโยชน์เพื่อนำไปใช้ในการตัดสินใจและการพยากรณ์ในเชิงธุรกิจสำหรับองค์กร

เริ่มต้นจากงานด้านการศึกษา วิจัยเกี่ยวกับพฤติกรรม ความต้องการของลูกค้าโดยทฤษฎีที่หลากหลาย จากงานวิจัยของ Reichstein, C. (2018) ทำการพัฒนาแบบจำลองที่สามารถสำรวจความต้องการของลูกค้าในอุตสาหกรรมการท่องเที่ยว โดยการเก็บรวบรวมข้อมูลเชิงประจักษ์และเชิงคุณภาพโดยการสัมภาษณ์ผู้เชี่ยวชาญ เป็นการสร้างพื้นฐานข้อมูล ทฤษฎีฐานราก (Grounded Theory) ใส่อ้างอิง ใช้ในการประเมินบทสัมภาษณ์ ที่มีความพิเศษเหมาะสำหรับการระบุความสัมพันธ์ใหม่หรือความสัมพันธ์ที่ไม่รู้จัก โดยมีวัตถุประสงค์ของการศึกษานี้คือเพื่อให้คำแนะนำแก่บริษัทท่องเที่ยวที่ต้องเผชิญกับความท้าทายและสภาพแวดล้อมของตลาดที่เปลี่ยนแปลงผ่านการ

เปลี่ยนแปลงในยุคดิจิทัลดิจิทัล ปัจจัยหลักที่มีอิทธิพลต่อศักยภาพในการเปลี่ยนแปลงความต้องการของลูกค้า (บริการดิจิทัล การตลาดดิจิทัล การขุดข้อมูล และชุมชนการท่องเที่ยวออนไลน์) ก่อให้เกิดรูปแบบแนวคิดเพื่อนำเสนอคำแนะนำสำหรับการดำเนินการ ผลการศึกษาแสดงให้เห็นว่าแนวทางส่วนบุคคลสำหรับลูกค้าในช่องทางการสื่อสารดิจิทัลถือเป็นข้อกำหนดที่สำคัญในการให้บริการในอนาคต ในขณะที่ Tanizaki et al. (2020) ใช้การคาดการณ์ความต้องการของลูกค้ามาใช้ในการจัดการร้านค้า โดยเสนอการจัดการร้านค้าร้านอาหารตามการคาดการณ์ความต้องการมาจัดการร้านค้าในร้านอาหาร เช่น การจัดหาพนักงาน การสั่งอาหาร เป็นต้น โดยอิงจากวิธีการการเรียนรู้ของเครื่อง ด้วยข้อมูลภายใน เช่น ข้อมูล POS และข้อมูลภายนอกที่ออกไปอย่างทั่วถึง อาทิ สภาพอากาศและเหตุการณ์ จะเห็นได้ว่า พฤติกรรมและความต้องการของลูกค้าหรือผู้บริโภคในยุคที่เทคโนโลยีมีความก้าวหน้าไปอย่างรวดเร็ว มีการเปลี่ยนแปลงไปอย่างมากจากความสะดวกสบายที่เพิ่มขึ้น รวมถึงข้อมูลจำนวนมากที่มีอยู่แล้วภายในองค์กร หรือการสนับสนุนทางเทคโนโลยีในการเก็บรวบรวมข้อมูลที่รวดเร็วยิ่งขึ้น

จากจำนวนข้อมูลลูกค้า ข้อมูลที่ต้องการคาดการณ์หรือจำนวนตัวแปรที่หลากหลาย ส่งผลให้มีการนำเทคนิคการทำเหมืองข้อมูล และการเรียนรู้ของเครื่อง หรือทฤษฎีอื่นๆ มาใช้ในการแบ่งกลุ่มข้อมูล (Clustering) เพื่อให้สามารถทำการวิเคราะห์ข้อมูลที่มีความแม่นยำมากยิ่งขึ้น เริ่มจาก Murray, Paul, et al. (2015) การพยากรณ์ความต้องการของห่วงโซ่อุปทาน (Supply Chain) โดยการจัดกลุ่มลูกค้า โดยปัญหาที่ว่าจะสร้างแบบจำลองที่สามารถคาดการณ์ขึ้น เป็นสิ่งที่ยากเมื่อไม่มีข้อมูลการทำงานร่วมกัน มีเครื่องมือพยากรณ์แบบดั้งเดิมและขั้นสูง แต่ไม่สามารถนำไปใช้กับลูกค้าจำนวนมากได้ และได้ใช้เทคนิคการทำเหมืองข้อมูลเพื่อระบุกลุ่มลูกค้าที่มีพฤติกรรมความต้องการที่คล้ายคลึงกัน การใช้งานในอดีตใช้เพื่อจัดกลุ่มลูกค้าที่มีความต้องการคล้ายกัน เช่นเดียวกันกับ Zhang et al. (2018) ใช้วิธีการแบ่งกลุ่มด้วยเทคนิค K-means มาแนะนำร้านอาหารจากการจัดกลุ่มความสัมพันธ์และความชอบของลูกค้า ซึ่งรายการคำแนะนำจัดทำขึ้นโดยมองหากลุ่มที่คล้ายกันมากที่สุดที่เป็นลูกค้าเป้าหมาย และตรวจสอบความถูกต้องของแบบจำลองโดยการนำกรณีศึกษาของ TripAdvisor.com มาเปรียบเทียบประสิทธิภาพ และในด้านการลงทุน (อรรถพร สะพะพันธุ์, 2559) ได้ใช้อัลกอริทึมอย่างต้นไม้ตัดสินใจ ในการสร้างแบบจำลองพยากรณ์ แต่พบว่าการจัดเตรียมข้อมูลในช่วงต้น ได้ใช้วิธีการแบ่งระดับข้อมูลเป็นลำดับขั้นที่เท่ากัน

หากจะกล่าวถึงการแบ่งกลุ่มลูกค้า หรือการแบ่งกลุ่มข้อมูลอันเป็นกระบวนการหนึ่งของการวิเคราะห์ข้อมูลเชิงพรรณนาหรือเป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยให้

ผลลัพธ์และตีความหมายข้อมูลได้ในรูปแบบความเข้าใจที่ใช้ประสบการณ์ทางด้านธุรกิจเข้ามาช่วยประกอบการตัดสินใจนั้นๆ จึงได้มีการพัฒนาแบบจำลองหรือต้นแบบสำหรับการพยากรณ์ขึ้นมาด้วยวิธีการจำแนกข้อมูล (Classification) หรือเป็นการเรียนรู้แบบมีผู้สอน (Supervised learning) ที่มีการนำเข้าข้อมูลจำนวนมากมาใช้ในการสร้างแบบจำลองพยากรณ์ต่างๆ และเพื่ออ้างอิงหลักการทางคณิตศาสตร์ที่มีความแม่นยำ และความน่าเชื่อถือมากยิ่งขึ้น โดย Li, Wang, et al. (2010) ได้พัฒนาแบบจำลองในการพยากรณ์เพื่อจำแนกประเภทลูกค้าตามทฤษฎีรัฟเซต (Rough Set Theory) ที่สามารถทำการจัดประเภทลูกค้าตามคุณสมบัติ ช่วยลดความซับซ้อนของผู้มีอำนาจตัดสินใจ ช่วยบริษัทคาดการณ์ล่วงหน้าเกี่ยวกับลูกค้าใหม่หรือระดับมูลค่าลูกค้าที่มีศักยภาพ ในด้านการเงิน การลงทุน (He, Shi, et al., 2014) ศึกษาการคาดคะเนการลาออกของลูกค้าของธนาคารพาณิชย์ด้วยแบบจำลอง ซัพพอร์ต เวคเตอร์ แมชชีน (Support Vector Machine) กล่าวถึงปัญหาธนาคารพาณิชย์ของจีนกำลังเผชิญ ในขณะที่ความต้องการทางการเงินที่เพิ่มขึ้นของลูกค้าทำให้การแข่งขันระหว่างธนาคารพาณิชย์ทวีความรุนแรง ธนาคารพาณิชย์จึงต้องหลีกเลี่ยงการสูญเสียลูกค้า พบว่าวิธีการนี้สามารถปรับปรุงความแม่นยำในการพยากรณ์ของแบบจำลองที่เลือกได้อย่างมีประสิทธิภาพ ในส่วนของ He, Changzheng, et al., (2015) นำเสนอการคาดคะเนประเภทความเสี่ยงของลูกค้าโดยอิงจาก Analog Complexing (AC) กล่าวถึงพฤติกรรมของผู้ถือบัตรเครดิตที่การเปลี่ยนแปลง ซึ่งนำไปสู่การเปลี่ยนแปลงประเภทความเสี่ยง โดยแนะนำวิธีการ AC ที่ใช้กระบวนการทางประวัติศาสตร์ที่สังเกตได้เองสำหรับการพยากรณ์ซึ่งไม่ต้องการข้อมูลใดๆ ของตัวแปรอินพุตในช่วงเวลาการพยากรณ์ที่ไม่รู้จัก ผู้เขียนได้นำวิธีการ AC มาใช้กับชุดข้อมูลลูกค้าของธนาคารจากเมืองทางตะวันตกของจีน และการศึกษาเชิงประจักษ์แสดงให้เห็นว่า AC ดีกว่าโครงข่ายประสาทเทียมที่ใช้กันอย่างแพร่หลายในแง่ของความแม่นยำในการพยากรณ์อย่างมีนัยสำคัญ

เมื่อได้มาซึ่งแบบจำลองในการพยากรณ์แล้ว จะเป็นการนำแบบจำลองนั้นไปใช้เพื่อแนะนำผลิตภัณฑ์ หรือใช้ประกอบการตัดใจในทางธุรกิจ Chatterjee (2019) ได้อธิบายการให้คะแนนและคำแนะนำของลูกค้าโดยที่ผู้ใช้งานสร้างขึ้นในเชิงคุณภาพและเชิงปริมาณ โดยการให้คะแนนและคำแนะนำของลูกค้าไม่ได้เป็นเพียงตัวบ่งชี้ประสิทธิภาพที่สำคัญสำหรับธุรกิจเท่านั้น แต่ยังทำหน้าที่เป็นแหล่งข้อมูลที่สำคัญสำหรับลูกค้าที่ไม่ได้รับข้อมูลที่อาจเกิดขึ้น โดยผลลัพธ์ที่ได้ให้ผลสนับสนุนทางทฤษฎีที่สำคัญ และข้อมูลเชิงลึกที่ได้สามารถช่วยให้นักท่องเที่ยวมีมุมมองโดยรวมเกี่ยวกับคุณภาพการบริการ ซึ่งจะช่วยในการตัดสินใจเข้ารับบริการ ในด้านการเงิน Yong, Xu, et al. (2018) ทำการปรับปรุงระบบคำแนะนำเฉพาะบุคคลตามข้อมูลผลิตภัณฑ์ทางการเงินที่กระจัดกระจาย ระบุปัญหา

เกี่ยวกับปัญหาของการขยายตัวของตลาดการเงินทางอินเทอร์เน็ต โดยงานวิจัยได้ใช้อัลกอริทึมการแนะนำแบบ กรองร่วม (Collaborative filtering) ในขณะที่ Hernández-Nieves, Elena, et al. (2020) ที่มีวัตถุประสงค์เพื่อสร้างระบบคำแนะนำการลงทุนโดยอาศัยการสกัดสัญญาณซื้อ ขาย จากผลการวิเคราะห์ทางเทคนิคและการคาดการณ์ มีการศึกษาเทคนิคต่างๆ ในการดึงข้อมูลและวิเคราะห์ข้อมูล มีการเสนอการพัฒนาแพลตฟอร์มการแสดงภาพสำหรับการโต้ตอบในระดับสูงระหว่างผู้ใช้และระบบการแนะนำ

ในทางเทคนิคหรือการนำแบบจำลองพยากรณ์ไปใช้ในการแนะนำ หรือพยากรณ์ผลต่างๆ อาจมีความแม่นยำที่ไม่แน่นอน มีการไหลของข้อมูลใหม่เข้ามาตลอดเวลา พฤติกรรมของลูกค้าที่เปลี่ยนไป หรือแม้กระทั่ง ความหลากหลายของตัวแปร ความผิดพลาดของข้อมูล ก็ส่งผลต่อแบบจำลองพยากรณ์ทั้งสิ้น โดย Xia, Jing, et al. (2017) นำเสนออัลกอริทึม Adjusted Weight Voting Algorithm สำหรับ Random Forests (RF) เรียกว่าอัลกอริทึมที่มีประสิทธิภาพในการจำแนกประเภท วิธีการทั่วไปในการจัดการกับค่าที่หายไปมักจะใช้วิธีประมาณค่า และการใส่ข้อมูลที่มีประสิทธิภาพซึ่งเชื่อมโยงกับสมมติฐานของคุณลักษณะข้อมูล อัลกอริทึม AWVRF ถูกนำมาเปรียบเทียบกับวิธีการใส่ค่าเฉลี่ย LeoFill, knnimput, BPCAfill และ RF แบบเดิมด้วยการตัดสินใจแทน (surrRF) โดยใช้การตรวจสอบความถูกต้อง 50 ครั้งในชุดข้อมูล 10 ชุดข้อมูล ด้วยการตั้งค่าการทดลองทั้งหมด 22 แบบ วิธีการของ AWVRF จะมีความแม่นยำสูงสุดในการตั้งค่า 14 แบบ Mashlakov, Aleksei, et al., 2021) ทำการประเมินประสิทธิภาพของแบบจำลองการเรียนรู้เชิงลึกสำหรับตัวแปรหลายตัวในการพยากรณ์ โดยแบบจำลองการเรียนรู้เชิงลึก (Deep Learning) มีศักยภาพที่จะพัฒนาการตัดสินใจในระยะสั้นโดยจับการพึ่งพาที่ซับซ้อนและความไม่แน่นอนของการทำงาน เพื่อทบทวนการพัฒนาล่าสุดในด้านการคาดการณ์อนุกรมเวลาความน่าจะเป็น

อย่างไรก็ตาม ในงานวิจัยนี้ เป็นการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน โดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Ensemble Learning) โดยใช้รูปแบบการโหวต (Majority Vote) จากอัลกอริทึมการจำแนกข้อมูล (Classification) ผู้วิจัยจึงเริ่มศึกษาค้นคว้าการเปรียบเทียบอัลกอริทึมและกระบวนการพัฒนาแบบจำลองพยากรณ์ ซึ่งThongchai Kaewkiriya (2017) นำเสนอกรอบการพยากรณ์ลูกค้าประกันชีวิตตาม Multi-Algorithm โดยกรอบการพยากรณ์ประกอบด้วยสามโมดูล ได้แก่ โมดูลการเตรียมข้อมูล โมดูลการทำความสะอาดข้อมูล และโมดูลการแยกข้อมูล ขั้นตอนการดึงข้อมูลแบ่งออกเป็น 3 ขั้นตอน ได้แก่ การเลือกคุณสมบัติ การจัดกลุ่มข้อมูลโดยใช้อัลกอริทึม K-mean และการดึงข้อมูลตามอัลกอริทึมโครงข่ายประสาทเทียมกับต้นไม้ตัดสินใจ เพื่อสร้างแบบจำลอง

คำแนะนำ ผลการเปรียบเทียบแสดงให้เห็นว่าการใช้หลายอัลกอริทึมมีความแม่นยำในการพยากรณ์สูงสุดที่ 92.83% Jaiswal, Devendra Prakash, et al. (2020) นำเสนอระบบการพยากรณ์ธุรกรรมของลูกค้าโดยรวมเอา Deep learning ร่วมกับอัลกอริทึมเพื่อเปรียบเทียบอย่าง XGBoost และ Logistic Regression ซึ่งจะช่วยในการระบุแนวทางที่ดีที่สุดสำหรับปัญหาให้สถาบันการเงินบรรลุความพึงพอใจของลูกค้า และในการระบุคุณลักษณะที่สำคัญจากชุดข้อมูล โดยผลลัพธ์ความแม่นยำพบว่า Deep Learning มีประสิทธิภาพมากที่สุด ตามด้วย XGBoost และ Logistic Regression ตามลำดับ Kong, Kun, et al. (2020) ศึกษาการพยากรณ์อัตราส่วนสภาพคล่องของกองทุนรวมผ่านกระบวนการ การเรียนรู้แบบกลุ่ม โดยการคาดการณ์รายวันของการถือครองกองทุนรวมสามารถเป็นเครื่องมือที่มีประโยชน์มากหากสามารถพยากรณ์ตำแหน่งการถือครองรายวันของกองทุนรวมขนาดใหญ่ได้ และนำเสนอแบบจำลอง Ensemble Learning ด้วย Majority Vote เพื่อพยากรณ์อัตราส่วนสภาพคล่องของกองทุนรวม แบบจำลองนี้มีการตีความที่ดี ซึ่งเป็นประโยชน์ต่อผู้ใช้ นักพัฒนา และหน่วยงานกำกับดูแล และทุกฝ่ายที่เกี่ยวข้อง เมื่อเทียบกับข้อมูลตำแหน่งกองทุนจริงที่เปิดเผยเพียงไตรมาสละครั้ง Tao, Tao, et al. (2019) นำเสนอการจำแนกประเภท (Classification) การลงทุนในกองทุนรวมด้วยแบบจำลอง Machine learning จากปัญหาที่จะกำหนดประเภทของกองทุนรวมได้อย่างถูกต้อง เนื่องจากกองทุนรวมจะประกอบด้วยเครื่องมือทางการเงินต่าง ๆ เช่น หุ้น พันธบัตร ส่วนเสริมเป็นต้น กองทุนถูกรวบรวมโดยผู้จัดการกองทุนที่มีประสบการณ์และจำแนกประเภทกองทุนรวมโดยใช้แบบจำลองทางสถิติ ในงานวิจัยนี้ได้ใช้ข้อมูลที่ใช้คัดลอกมาจาก Yahoo Finance ซึ่งประกอบด้วยกองทุนต่างๆ จำนวน 25,393 กองทุน กับ 54 ตัวแปร และทำการเปรียบเทียบประสิทธิภาพอัลกอริทึม 4 อัลกอริทึม ได้แก่ K-Nearest Neighbors Algorithm, Neural Network Model, XGBoost และ Random Forest โดยอธิบายไว้ว่า XGboost มีประสิทธิภาพที่ดีที่สุด Random Forest จำนวนต้นไม้มีผลเพียงเล็กน้อยต่อการตรวจสอบความถูกต้อง โครงข่ายประสาทเทียมทำงานได้แย่ที่สุดในทั้งสี่กรณี Mashayekhi, Morteza, et al. (2018) นำเสนอรูปแบบพยากรณ์การสมัครและถอนในกองทุนรวมโดยใช้ Machine learning ได้ทำการศึกษาปัจจัยที่เกี่ยวข้องกับรูปแบบการลงทุน และพฤติกรรมของนักลงทุน เป็นประโยชน์อย่างมากสำหรับผู้จัดการกองทุนที่จะรู้ว่าใครคือนักลงทุนที่มีแนวโน้มว่าจะสมัครรับหรือไถ่ถอนจากกองทุนใดกองทุนหนึ่งในอนาคต นอกจากนี้ อีกทั้งยังช่วยให้ผู้จัดการกองทุนสามารถวางแผนเพื่อเพิ่มประสิทธิภาพการทำงานของกองทุนได้ โดยการทดลองกับข้อมูลการทำธุรกรรมในอดีตของกองทุนรวมประมาณ 400 กองทุน และทำการเปรียบเทียบประสิทธิภาพอัลกอริทึม 3 อัลกอริทึม ได้แก่ XGBoost, Random Forest

และ Logistic Regression โดยให้ผลลัพธ์ประสิทธิภาพความแม่นยำในการพยากรณ์การสมัครและการถอนกองทุนมากที่สุดคือ Logistic Regression และ Saloni Kumari (2021) ใช้วิธีการเรียนรู้แบบกลุ่มสำหรับการจำแนก และการพยากรณ์โรคเบาหวานโดยใช้การโหวต ด้วยอัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ AdaBoost, Logistic Regression, Support Vector machine, Random Forest, Naïve Bayes, Bagging, GradientBoost, XGBoost, CatBoost และดำเนินการนำผลลัพธ์มาพยากรณ์ในรูปแบบการโหวตที่มีความแม่นยำถึง 97.02%

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องแล้ว ผู้วิจัยเล็งเห็นและให้ความสนใจในการศึกษาใน 2 ประเด็น ได้แก่ รูปแบบการจัดการข้อมูลก่อนนำไปพัฒนาแบบจำลองพยากรณ์ โดยตั้งสมมติฐานว่าการแบ่งกลุ่มของข้อมูลด้วยวิธีการทางการเรียนรู้ของเครื่องมีความเหมาะสมกว่าการแบ่งกลุ่มข้อมูลออกเป็นลำดับแบบกำหนดระดับขั้นแน่นอน เพื่อช่วยในเรื่องของปริมาณข้อมูลที่มีแนวโน้มกลุ่มใดกลุ่มหนึ่งที่มากเกินไป และเปรียบเทียบประสิทธิภาพอัลกอริทึมการจำแนกข้อมูล (Classification) สำหรับวิเคราะห์ข้อสังเกตอัลกอริทึมต่าง ๆ ที่จะมาใช้เทคนิคการโหวต (Majority Vote) เพื่อให้ได้ผลการพยากรณ์ที่จะนำไปใช้ในการแนะนำที่ดีที่สุด

ตาราง 2 สรุปงานวิจัยที่เกี่ยวข้องเกี่ยวข้อง

งานวิจัย	**วัตถุประสงค์**	**ทฤษฎี, อัลกอริทึมที่ใช้หรือเปรียบเทียบ**	**ผลของอัลกอริทึมที่ดีที่สุด**
Customer demand, Customer behavior			
Reichstein, C. (2018)	สำรวจความต้องการของลูกค้าในอุตสาหกรรมการท่องเที่ยวและแนะนำบริษัทท่องเที่ยว	Grounded Theory	
Tanizaki, Takashi, et al. (2020)	จัดการร้านค้าร้านอาหารตามการคาดการณ์อุปสงค์	Random forest	
Clustering			
Murray, Paul, et al. (2015)	พยากรณ์ความต้องการของห่วงโซ่อุปทาน โดยการจัดกลุ่มลูกค้า	K-means	

ตาราง 2 สรุปงานวิจัยที่เกี่ยวข้องเกี่ยวข้อง (ต่อ)

งานวิจัย	วัตถุประสงค์	ทฤษฎี, อัลกอริทึมที่ใช้หรือเปรียบเทียบ	ผลของอัลกอริทึมที่ดีที่สุด
Zhang, Chenbin (2018)	แนะนำร้านอาหารจากการจัดกลุ่มความสัมพันธ์และความชอบของลูกค้า	K-means	
Atthaporn Sapaphan (2016)	ออกแบบระบบช่วยตัดสินใจเลือกซื้อผลิตภัณฑ์กองทุน	Scale, Decision Tree	
Recommendation			
Chatterjee, Swagato (2019)	อธิบายการให้คะแนนและคำแนะนำของลูกค้า	Text mining	
Yong, Xu, et al. (2018)	ปรับปรุงระบบคำแนะนำเฉพาะบุคคลตามข้อมูลผลิตภัณฑ์ทางการเงิน	collaborative filtering	
Hernández-Nieves, Elena, et al. (2020)	แพลตฟอร์ม Machine learning สำหรับระบบแนะนำการลงทุนหุ้น	RF, GB, SMV-LinearSVR, MLP และ K-NN	
Classification			
Mashlakov, Aleksei, et al. (2021)	ประเมินประสิทธิภาพของแบบจำลองการเรียนรู้เชิงลึกสำหรับตัวแปรหลายตัว	DeepAR, DSANet, DeepAR และ DeepTCN	DeepAR และ DeepTCN
Li, Ju, et al (2020)	พยากรณ์การจำแนกประเภทลูกค้า	Rough Set Theory	
HE, Benlan, et al. (2014)	คาดคะเนการลาออกของลูกค้าของธนาคารพาณิชย์	Support Vector Machine	SVM
HE, Changzheng, et al. (2015)	การคาดคะเนประเภทความเสี่ยงของลูกค้า	Analog Complexing, Neural Netwok	Analog Complexing

ตาราง 2 สรุปงานวิจัยที่เกี่ยวข้องเกี่ยวข้อง (ต่อ)

งานวิจัย	วัตถุประสงค์	ทฤษฎี, อัลกอริทึมที่ใช้หรือเปรียบเทียบ	ผลของอัลกอริทึมที่ดีที่สุด
Xia, Jing, et al. (2017)	การใส่ค่า Missing Value เพื่อรองรับการจำแนกขอมูล	AWVRF, LeoFill, knnimput, BPCAfill และ surrRF	Adjusted Weight Voting Algorithm
Ensemble learning			
Thongchai Kaewkiriya (2017)	กรอบการพยากรณ์ลูกค้าประกันชีวิตตาม Multi-Algorithm	Decision Tree, Neural Network	Multi-Algorithm
Jaiswal, Devendra Prakash, et al. (2020)	ระบบการพยากรณ์ธุรกรรมของลูกค้า	Deep Learning, XGBoost และ LR	Deep Learning
Kong, Kun, et al. (2020)	พยากรณ์อัตราส่วนสภาพคล่องของกองทุนรวม	Lasso regression, multiple linear regression	Ensemble Learning ด้วย Majority Vote
Tao, Tao, et al. (2019, May)	จำแนกประเภทการลงทุนในกองทุนรวม	K-NN, Neural Networkl, XGBoost และ RF	XGboost
Mashayekhi, Morteza, et al. (2018)	พยากรณ์การสมัครและถอนในกองทุนรวม	XGBoost, RF และ Logistic Regression	Logistic Regression
Saloni Kumari (2021)	จำแนกและการพยากรณ์โรคเบาหวาน	Majority Vote, AdaBoost, LR, SVM, RF, Naïve Bayes, Bagging, GB, XGBoost, DT, K-NN	Majority Vote

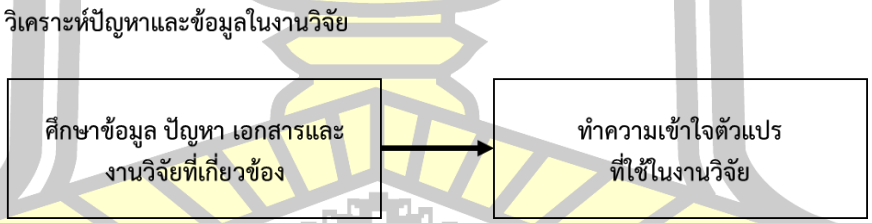
บทที่ 3
วิธีการดำเนินการวิจัย

การวิจัยเรื่อง การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Development of Customer Predictive Model for Investment using Ensemble Learning Technic) โดยมีรายละเอียดวิธีการดำเนินการวิจัยตามขั้นตอน ดังนี้

1. วิเคราะห์ปัญหาและข้อมูลในงานวิจัย
2. ออกแบบกรอบแนวคิด
3. พัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน
4. ทดสอบเปรียบเทียบประสิทธิภาพ

วิเคราะห์ปัญหาและข้อมูลในงานวิจัย

ผู้วิจัยได้ดำเนินการวิเคราะห์ปัญหา ข้อมูล ปัจจัย และรูปแบบการนำเสนอผลิตภัณฑ์จากบริษัทหลักทรัพย์จัดการกองทุนที่ทำการขาย ผู้วิจัยได้ดำเนินการศึกษารูปแบบการนำเสนอผลิตภัณฑ์จากบริษัทหลักทรัพย์จัดการกองทุน โดยมีขั้นตอนการดำเนินการวิจัย และอธิบายได้ตามภาพประกอบ 18 ดังต่อไปนี้



ภาพประกอบ 18 กรอบการดำเนินงานวิจัยเกี่ยวกับการวิเคราะห์ปัญหาและข้อมูลในงานวิจัย

ศึกษาข้อมูล ปัญหา เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาข้อมูล ปัญหา เอกสารและงานวิจัยที่เกี่ยวข้อง มาสรุปเป็นแนวทางในการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม โดยทำการศึกษาเรื่องต่าง ๆ ตลอดการดำเนินการขายกองทุน ตั้งแต่ การนำเสนอผลิตภัณฑ์ผ่านพนักงานสาขา, การอธิบายคุณสมบัติกองทุนต่าง ๆ ให้แก้ผู้สนใจสมัครเข้าซื้อกองทุน และการปิดการขายกองทุนให้แก้ผู้สนใจสมัครเข้าซื้อกองทุน

ทำความเข้าใจตัวแปรที่ใช้ในงานวิจัย

ผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลจากลูกค้าที่ทำรายการซื้อกองทุน โดยมีรายละเอียด ดังนี้

1. ดำเนินการรวบข้อมูลกลุ่มประชากรลูกค้ากองทุนของธนาคารแห่งหนึ่ง จำนวนทั้งสิ้น 19,577 ราย ที่มีการซื้อกองทุนระหว่างวันที่ 2 มกราคม – 31 กรกฎาคม 2562 มูลค่าการเลือกซื้อตั้งแต่ 500 บาทขึ้นไป โดยจัดหมวดหมู่กองทุนออกเป็น 4 ประเภท โดยมีรายละเอียดทางสถิติ ดังนี้

ตาราง 3 ภาพรวมข้อมูลประเภทกองทุน

Class	N	Average	Min	Max	S.D.
LTF/RMF (LR)	2,992	39,911	500	900,000	69,195.33
Money Market (MM)	11,562	794,790	500	9,850,000	1,171,852.40
Mutual Fund (OE)	6,441	932,044	1,085	9,472,406	659,896.77
Team Fund (TF)	3,485	270,921	500	8,000,000	1,041,179.52

2. ดำเนินการเก็บรวบรวมตัวแปรอื่น ๆ ของกลุ่มประชากรลูกค้ากองทุนสำหรับเป็นตัวแปรต้นในการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม โดยมีรายละเอียดทางสถิติ ดังนี้

ตาราง 4 ภาพรวมข้อมูลตัวแปรต้น

ตัวแปร	**ความหมาย**
Branch_Provice	ภูมิลำเนา
Gender	เพศ
Education	การศึกษา
Occupation	อาชีพ
Income	รายได้
Seaving A/C	บัญชีเงินฝากออมทรัพย์
Fixed A/C	บัญชีเงินฝากประจำ
SPA A/C	บัญชีเงินฝากประจำประเภทพิเศษ
Frequency	ความถี่ในการซื้อกองทุน

3. ดำเนินการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ด้วยเครื่องมือทางด้านการวิเคราะห์ข้อมูลอย่างแรพพิทไมเนอร์ สตูดิโอให้อยู่ในรูปแบบตัวเลข (numeric) ตามตาราง 5 โดยผู้วิจัยสนใจนำเสนอรูปแบบการแปลงข้อมูลที่เป็นตัวแปรตามออกเป็น 2 รูปแบบ ดังนี้

ตาราง 5 การแปลงข้อมูลตัวแปรต้นประเภท Nominal

ตัวแปร	**แทนค่าข้อมูลนำไปประมวลผล**
Branch_Provice	1 = กรุงเทพและปริมณฑล 2 = ภาคกลาง 3 = ภาคตะวันออกเฉียงเหนือ 4 = ภาคเหนือ 5 = ภาคใต้
Gender	1 = ชาย 2 = หญิง
Education	1 = ต่ำกว่าปริญญาตรี 2 = ปริญญาตรี 3 = ปริญญาโท 4 = สูงกว่าปริญญาโท
Occupation	1 = รับราชการ 2 = บริษัทเอกชน 3 = ธุรกิจส่วนตัว 4 = อาชีพอิสระ
Income	1 = น้อยกว่า 20,000 2 = ระหว่าง 20,001 – 40,000 3 = ระหว่าง 40,001 – 60,000 4 = ระหว่าง 60,001 – 80,000 5 = ระหว่าง 80,001 – 100,000 6 = 100,000 บาท ขึ้นไป

พหุน ปณ ฑิต

ตาราง 5 การแปลงข้อมูลตัวแปรต้นประเภท Nominal (ต่อ)

ตัวแปร	แทนค่าข้อมูลนำไปประมวลผล
Frequency	1 = ซื้อครั้งเดียว 2 = 2-4 ครั้งต่อเดือน 3 = 5-7 ครั้งต่อเดือน 4 = 8-10 ครั้งต่อเดือน 5 = 11-13 ครั้งต่อเดือน 6 = มากกว่า 14 ครั้งต่อเดือน

3.1 รูปแบบที่ 1 แปลงข้อมูลตัวแปรตามออกเป็นกลุ่มตามระยะของตัวแปรตามเท่า ๆ กัน เช่นเดียวกับตัวแปรต้นก่อนนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม ตามตาราง 6

ตาราง 6 การแปลงข้อมูลตัวแปรตามและตัวแปลต้นประเภท numeric รูปแบบที่ 1

Class	แทนค่าข้อมูลนำไปประมวลผล รูปแบบที่ 1
LTF/RMF (LR) class label	"1"=1-200,000 "2"=200,001-400,000 "3"400,001-600,000 "4"=600,001-800,000 "5"=800,001-1,000,000 "6"=1,000,001 ขึ้นไป
Money Market (MM) class label	"1"=1-200,000 "2"=200,001-400,000 "3"400,001-600,000 "4"=600,001-800,000 "5"=800,001-1,000,000 "6"=1,000,001 ขึ้นไป

ตาราง 6 การแปลงข้อมูลตัวแปรตามและตัวแปลต้นประเภท numeric รูปแบบที่ 1 (ต่อ)

Class	**แทนค่าข้อมูลนำไปประมวลผล รูปแบบที่ 1**
Mutual Fund (MF) class label	"1"=1-200,000 "2"=200,001-400,000 "3"400,001-600,000 "4"=600,001-800,000 "5"=800,001-1,000,000 "6"=1,000,001 ขึ้นไป
Team Fund (TF) class label	"1"=1-200,000 "2"=200,001-400,000 "3"400,001-600,000 "4"=600,001-800,000 "5"=800,001-1,000,000 "6"=1,000,001 ขึ้นไป
Saving A/C	1 = น้อยกว่า 20,000 2 = ระหว่าง 20,001 – 40,000 3 = ระหว่าง 40,001 – 60,000 4 = ระหว่าง 60,001 – 80,000 5 = ระหว่าง 80,001 – 100,000 6 = 100,000 บาท ขึ้นไป
Fixed A/C	1 = น้อยกว่า 20,000 2 = ระหว่าง 20,001 – 40,000 3 = ระหว่าง 40,001 – 60,000 4 = ระหว่าง 60,001 – 80,000 5 = ระหว่าง 80,001 – 100,000 6 = 100,000 บาท ขึ้นไป

ตาราง 6 การแปลงข้อมูลตัวแปรตามและตัวแปลต้นประเภท numeric รูปแบบที่ 1 (ต่อ)

Class	แทนค่าข้อมูลนำไปประมวลผล รูปแบบที่ 1
SPA A/C	1 = น้อยกว่า 20,000 2 = ระหว่าง 20,001 – 40,000 3 = ระหว่าง 40,001 – 60,000 4 = ระหว่าง 60,001 – 80,000 5 = ระหว่าง 80,001 – 100,000 6 = 100,000 บาท ขึ้นไป

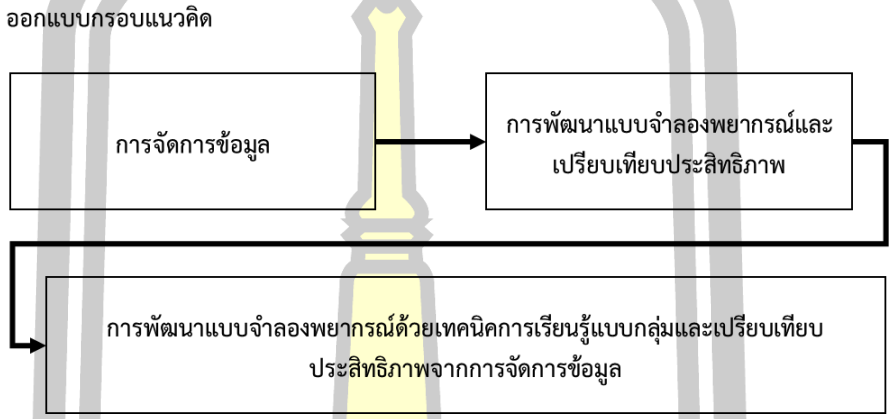
3.2 รูปแบบที่ 2 แปลงข้อมูลตัวแปรตามและตัวแปรต้นด้วยเทคนิคการแบ่งกลุ่ม (Clustering) ด้วย K-Means algorithm โดยนำเสนอจำนวนกลุ่มข้อมูลหรือค่า K ที่หลากหลายเพื่อเปรียบเทียบจำนวนกลุ่มที่เหมาะสมก่อนนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม

หลังจากดำเนินการแปลงข้อมูลเสร็จแล้ว จึงตรวจสอบข้อมูลดังกล่าวว่ามีข้อมูลใดที่ผิดพลาดหรือไม่ เช่น บางรายการมีค่าว่าง ซึ่งจะถือว่าเป็นข้อมูลที่ผิดพลาดไม่สมบูรณ์ทำให้ไม่สามารถนำข้อมูลชุดนั้นมาแสดงผลได้ จึงได้นำข้อมูลที่ผิดพลาดดังกล่าวนั้น ผ่านกระบวนการคัดกรองข้อมูลนำเข้าด้วยโปรแกรมแรพพิทไมเนอร์ สตูดิโอ จากนั้นจึงเลือกอัลกอริทึมในการวิเคราะห์ ซึ่งอัลกอริทึมที่เลือกใช้ในการประมวลผลมี 5 อัลกอริทึม ประกอบด้วย K Nearest Neighbors, Naive Bayes, Decision Tree, Rule Induction และ Neural Network โดยเปรียบเทียบผลความถูกต้อง และความแม่นยำในการพยากรณ์ของอัลกอริทึมแต่ละตัว แล้วจึงเลือกรูปแบบที่ให้ความแม่นยำสูงสุด เพื่อใช้เป็นกฎพื้นฐานเพื่อใช้ในการพยากรณ์ในระบบแนะนำเลือกซื้อผลิตภัณฑ์กองทุนในขั้นตอนต่อถัดไป โดยสามารถแสดงได้จากกรอบแนวคิดการวิจัย

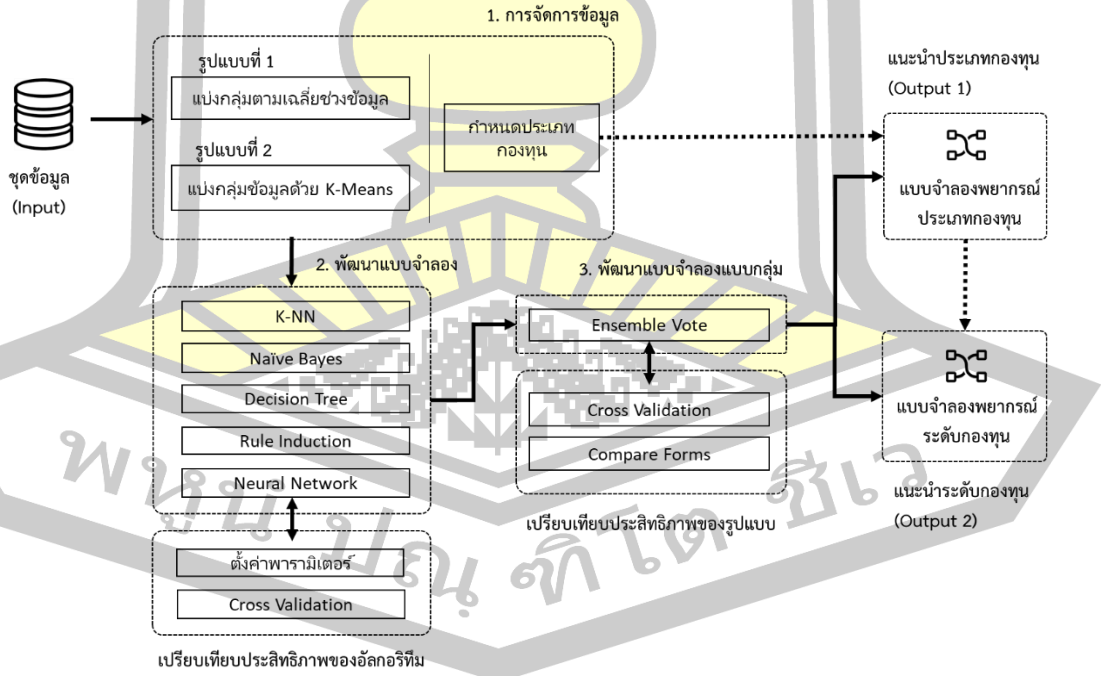
ออกแบบกรอบแนวคิด

หลังจากผู้วิจัยได้ดำเนินการศึกษาข้อมูล ปัญหา เอกสารและงานวิจัยที่เกี่ยวข้อง จึงได้ออกแบบกรอบแนวคิดเกี่ยวกับการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม ตามภาพประกอบ 19 และ 20 ประกอบด้วยส่วนการดำเนินงานหลัก 3 ส่วน และสามารถอธิบายส่วนต่าง ๆ ได้ ดังนี้

1. การจัดการข้อมูล
2. การพัฒนาแบบจำลองพยากรณ์และเปรียบเทียบประสิทธิภาพ
3. การพัฒนาแบบจำลองพยากรณ์ด้วยเทคนิคการเรียนรู้แบบกลุ่มและเปรียบเทียบประสิทธิภาพจากการจัดการข้อมูล



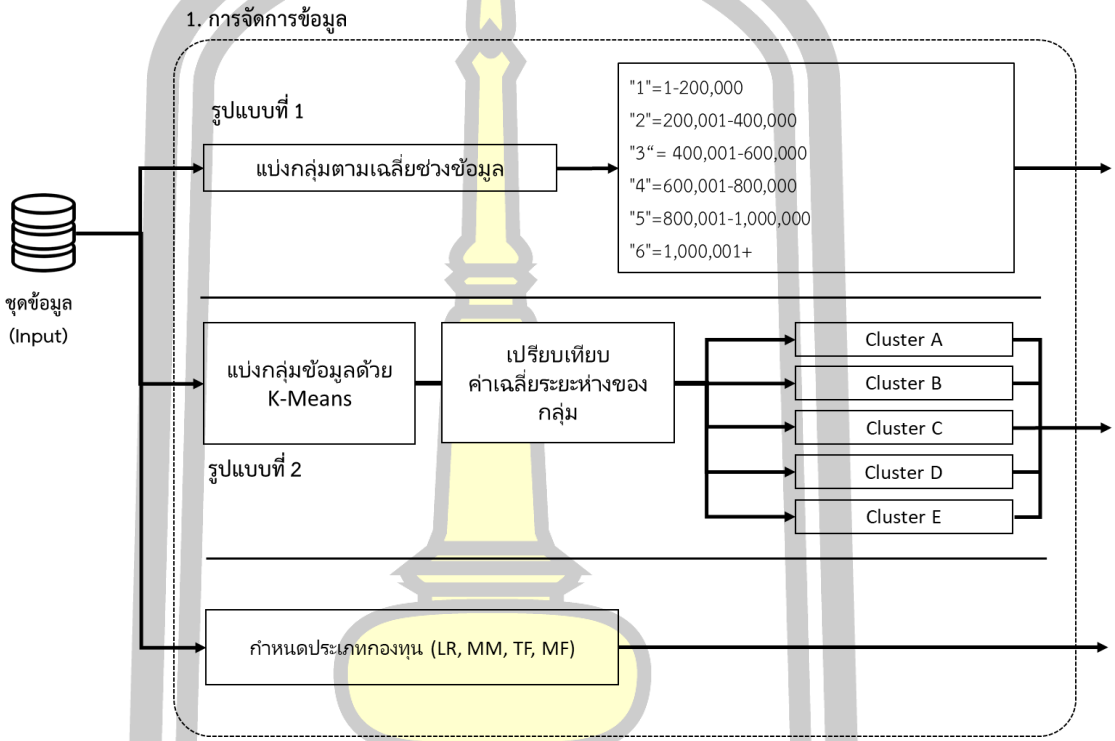
ภาพประกอบ 19 กรอบการดำเนินงานวิจัยเกี่ยวกับการวิเคราะห์ปัญหาและข้อมูลในงานวิจัย



ภาพประกอบ 20 การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

การจัดการข้อมูล

ดำเนินการจัดการข้อมูลรับเข้าออกเป็น 2 รูปแบบ ตามภาพประกอบ 21 โดยสามารถอธิบายรายละเอียดได้ ดังนี้



ภาพประกอบ 21 ส่วนการจัดการข้อมูล

รูปแบบที่ 1 จัดการข้อมูลตัวแปรตามหรือผลิตภัณฑ์กองทุนทั้ง 4 ผลิตภัณฑ์ ได้แก่ LTF/RMF (LR), Money Market (MM), Mutual Fund (OE) และ Team Fund (TF) ให้อยู่ในรูปแบบตัวเลข (numeric) ตามเกณฑ์กำหนด เช่นเดียวกับตัวแปรอิสระอื่น ๆ ตามตาราง 7

ตาราง 7 เกณฑ์กำหนดการแปลงข้อมูลตัวแปรตามรูปแบบที่ 1

เกณฑ์กำหนด	แปลงเป็น
1 – 200,000	1
200,001 – 400,000	2
400,001 – 600,000	3
600,001 – 800,000	4
800,001 – 1,000,000	5
> 1,000,000	6

รูปแบบที่ 2 จัดการข้อมูลตัวแปรตามหรือผลิตภัณฑ์กองทุนทั้ง 4 ผลิตภัณฑ์ ได้แก่ LTF/RMF (LR), Money Market (MM), Mutual Fund (OE) และ Team Fund (TF) ให้อยู่ในรูปแบบตัวเลข (numeric) จากวิธีการแบ่งกลุ่มข้อมูล (Clustering) ด้วย K-Means Algorithm โดยผู้วิจัยทำการทดลองกำหนดค่า K ในหลายๆ ค่า เพื่อให้ได้จำนวนกลุ่มที่ดีที่สุด ก่อนนำผลในการแบ่งกลุ่มไปใช้ในขั้นตอนถัดไป

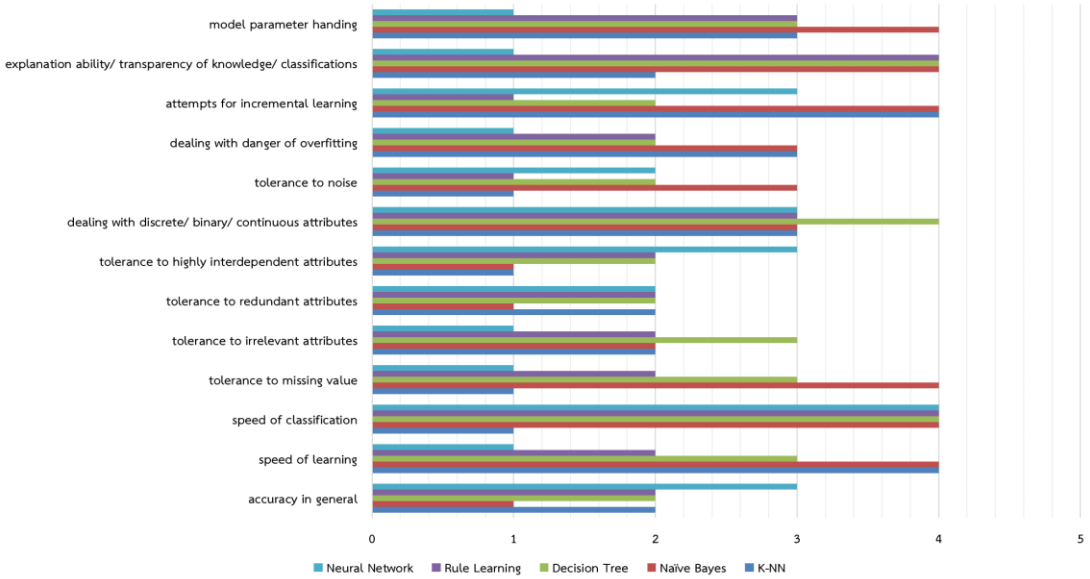
ต่อไป เป็นการกำหนดประเภทกองทุนจากข้อมูลมูลค่ากองทุนที่มากที่สุดของผู้ถือกองทุนนั้นๆ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

1. การพัฒนาแบบจำลองพยากรณ์และเปรียบเทียบประสิทธิภาพ

ทำการนำชุดข้อมูลที่ได้จากส่วนการจัดการข้อมูลทั้ง 2 รูปแบบ มาผ่านกระบวนการเรียนรู้ Machine learning ด้วยเทคนิคทั้งหมด 5 อัลกอริทึม ได้แก่ K-NN, Naïve Bayes, Decision Tree, Rule Induction และ Neural Network โดยแสดงเหตุผลในการคัดเลือกอัลกอริทึม ได้แก่

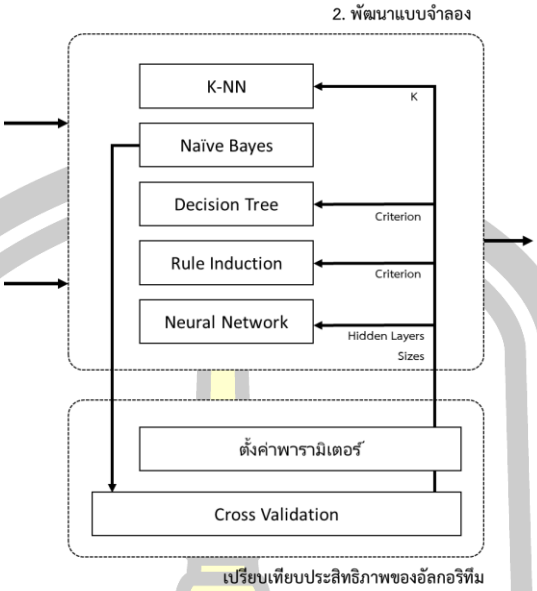
1. เป็นอัลกอริทึมสำหรับพยากรณ์ประเภท Lazy (K-NN), Bayesian (Naïve Bayes), Trees, Rules และ Neural Nets โดยจะเห็นว่าผู้วิจัยทำการเลือกอัลกอริทึมที่แตกต่างกัน เพื่อให้สามารถเปรียบเทียบผลประสิทธิภาพได้อย่างชัดเจน
2. ผู้วิจัยทำการเลือกอัลกอริทึมที่มีความต้องการประเภทข้อมูลรับเข้าสำหรับประมวลผลที่เหมือนกัน เพื่อลดความคลาดเคลื่อน และความยุ่งยากในการจัดเตรียมข้อมูลที่ไม่เหมือนกัน
3. เลือกอัลกอริทึมอันเป็นที่นิยม และเป็นพื้นฐานในการเรียนรู้อัลกอริทึมอื่นๆ ได้

อย่างเช่น Decision Tree ที่เป็นส่วนหนึ่งอย่างอัลกอริทึม Random Forest หรือ Neural Network ที่เป็นส่วนหนึ่งอย่างอัลกอริทึม Deep learning และทำการเปรียบเทียบประสิทธิภาพในด้านต่างๆ ของอัลกอริทึมที่ใช้ในงานวิจัย (Kotsiantis et al., 2007) ตามภาพประกอบ 22 จะพบว่า หากเปรียบเทียบด้านความแม่นยำ Neural Network จะมีประสิทธิภาพมากที่สุด ในขณะที่หากเปรียบเทียบด้านความเร็วในการเรียนรู้ข้อมูล พบว่า Neural Network จะมีประสิทธิภาพน้อยที่สุด



ภาพประกอบ 22 เปรียบเทียบอัลกอริทึมการเรียนรู้

โดยในการเปรียบเทียบ ทำการบันทึกการตั้งค่าพารามิเตอร์ไว้ แล้วทำการวัดผลประสิทธิภาพด้วยเทคนิคการแบ่งข้อมูลแบบ Cross Validation เพื่อประเมินค่าความแม่นยำ (Accuracy) ของรูปแบบแต่ละแบบจำลอง ตามภาพประกอบ 23

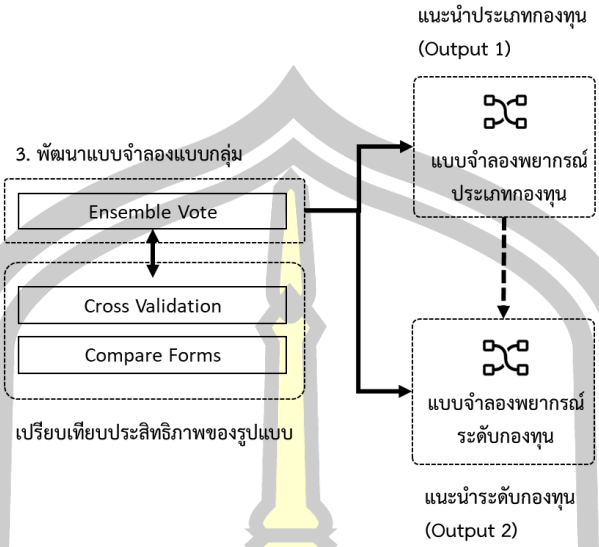


ภาพประกอบ 23 ส่วนการพัฒนารูปแบบพยากรณ์และเปรียบเทียบประสิทธิภาพ

เมื่อได้ค่าความแม่นยำแล้ว ทำการปรับค่าพารามิเตอร์ของแต่ละอัลกอริทึม และทำการเปรียบเทียบค่าความแม่นยำทุกครั้ง จนได้ค่าพารามิเตอร์ที่ส่งผลต่อประสิทธิภาพของอัลกอริทึมมากที่สุด อย่างไรก็ตาม อัลกอริทึมนาอีฟเบย์ ไม่มีค่าพารามิเตอร์ที่สามารถปรับตั้งค่าได้ โดยค่าพารามิเตอร์ของอัลกอริทึมอื่นๆ จะอธิบายในหัวข้อพัฒนาแบบจำลองการพยากรณ์

ส่วนการพัฒนาแบบจำลองพยากรณ์ด้วยเทคนิคการเรียนรู้แบบกลุ่มและเปรียบเทียบประสิทธิภาพจากการจัดการข้อมูล

ทำการนำรูปแบบการตั้งค่าพารามิเตอร์ของแต่ละอัลกอริทึมที่ได้ทำการเปรียบเทียบในส่วนของการพัฒนารูปแบบพยากรณ์และเปรียบเทียบประสิทธิภาพ มาเข้าสู่การพัฒนารูปแบบพยากรณ์ด้วยเทคนิคการเรียนรู้แบบกลุ่ม (Ensemble learning) ด้วยเทคนิคการ Vote แสดงตามภาพประกอบ 24



ภาพประกอบ 24 ส่วนการพัฒนาแบบจำลองพยากรณ์และเปรียบเทียบประสิทธิภาพ

หลังจากนั้น ทำการวัดประสิทธิภาพของผลลัพธ์ด้วยเทคนิคการแบ่งข้อมูลแบบ Cross Validation เพื่อประเมินค่าความแม่นยำ (Accuracy) เช่นเดียวกับส่วนการพัฒนารูปแบบพยากรณ์และเปรียบเทียบประสิทธิภาพ เพียงในส่วนนี้เป็นการเปรียบเทียบระหว่างรูปแบบ 2 รูปแบบ ที่ได้จากส่วนการจัดการข้อมูล เพื่อให้ได้รูปแบบพยากรณ์ที่มีความแม่นยำมากที่สุดแล้วทำการสรุปผลก่อนนำไปใช้

ในส่วนนี้ จะแบ่งการพัฒนาแบบจำลองเป็น 2 แบบจำลองได้แก่ แบบจำลองพยากรณ์ประเภทกองทุนและ แบบจำลองพยากรณ์ระดับกองทุน ซึ่งให้แบบจำลองพยากรณ์ประเภทกองทุนมีหน้าที่ในการแนะนำประเภทกองทุนก่อนเพื่อจำกัดจำนวนแบบจำลองพยากรณ์ระดับกองทุนสำหรับการแนะนำกองทุนที่มีความละเอียดยิ่งขึ้น

พัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

ทำการสร้างแบบจำลองสำหรับพยากรณ์เพื่อนำผลลัพธ์ไปแนะนำลูกค้าในการลงทุน โดยมีขั้นตอนการดำเนินงานและอธิบายได้ ดังต่อไปนี้

กำหนดค่าพารามิเตอร์และตัวชี้วัดประสิทธิภาพ

ผู้วิจัยได้ใช้เทคนิคการเรียนรู้แบบกลุ่ม (Ensemble Learning) ประกอบไปด้วย กลุ่มอัลกอริทึมสำหรับพยากรณ์ประเภท Lazy, Bayesian, Trees, Rules และ Neural Nets ซึ่งมีการตั้ง

ค่าพารามิเตอร์และตัวชี้วัดประสิทธิภาพที่ใช้ในการวิจัย (Sachdev, 2020) ตามตาราง 8

ตาราง 8 ค่าพารามิเตอร์ที่สามารถปรับสำหรับการเปรียบเทียบประสิทธิภาพของแบบจำลอง

อัลกอริทึม	**ประเภท**	**พารามิเตอร์**
1.K Nearest Neighbors	Lazy	ค่า K
2.Naive Bayes	Bayesian	-
3.Decision Tree	Trees	Criterion = gain_ratio, information_gain, gini_index และ accuracy
4.Rule Induction	Rules	Criterion = information_gain และ accuracy
5.Neural Network	Neural Nets	hidden layers sizes

โดยอัลกอริทึมที่ใช้จะเป็นอัลกอริทึมประเภทการจำแนกข้อมูล (Classification) สามารถวัดค่าประสิทธิภาพของแบบจำลอง ซึ่งได้กำหนดตัวชี้วัดที่ได้ดำเนินการตรวจสอบเพื่อประเมินประสิทธิภาพทางสถิติของรูปแบบการเรียนรู้ ด้วยวิธีการแบ่งข้อมูลเพื่อทำการทดสอบ Cross Validation วิธีนี้เป็นวิธีที่นิยมในการทำงานวิจัย นำมาใช้ในการทดสอบประสิทธิภาพของแบบจำลอง เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพจะทำโดยการแบ่งข้อมูลออกเป็นหลายส่วน โดยในงานวิจัยได้กำหนด 5-Fold Cross-Validation คือการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของแบบจำลอง ทำวนไปเช่นนี้จนครบจำนวนที่แบ่งไว้ ซึ่งได้เลือกรายการค่าเกณฑ์ประสิทธิภาพของงานการจำแนกประเภท ดังนี้

1. ความแม่นยำ (Accuracy) ของผลการพยากรณ์ อธิบายได้ว่าแบบจำลองพยากรณ์ที่พัฒนามีความถูกต้องทั้งหมดกี่เปอร์เซ็นต์ ด้วยตาราง Confusion Matrix ซึ่งใช้กันอย่างแพร่หลายในงาน Machine Learning ระหว่างผลการพยากรณ์ของแบบจำลอง (Prediction) และค่าจริง (Actual Label) การคำนวณ Metrics คือค่าที่ได้จากการคำนวณเปรียบเทียบว่าแบบจำลองพยากรณ์ได้มีความถูกต้องแค่ไหนพิจารณาตามสมการ 5

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative}) / N \quad (5)$$

2. ค่าเฉลี่ยความเที่ยง (Weighted mean precision) จำนวนที่พยากรณ์ถูกจากข้อมูลที่พยากรณ์ว่าเป็นคลาสที่พิจารณาอยู่ตามสมการที่ 6

Precision = True Positive / (True Positive + False Positive) (6)

3. ค่าเฉลี่ยความระลึก (Weighted mean recall) จำนวนข้อมูลที่พยากรณ์ถูกพิจารณาตามสมการที่ 7

Recall = True Positive / (True Positive + False Negative) (7)

เมื่อ True Positive คือ จำนวนที่พยากรณ์ตรงในคลาสที่กำลังพิจารณา

True Negative คือ จำนวนที่พยากรณ์ตรงในคลาสที่ไม่ได้พิจารณา

False Positive คือ จำนวนที่พยากรณ์ผิดในคลาสที่กำลังพิจารณา

False Negative คือ จำนวนที่พยากรณ์ผิดในคลาสที่ไม่ได้พิจารณา

หลักจากได้กำหนดการตั้งค่าพารามิเตอร์และตัวชี้วัดประสิทธิภาพต่าง ๆ แล้วได้ทำการประมวลผลด้วยโปรแกรมแรพพิทไมเนอร์ สตูดิโอ แล้วจึงนำผลที่ได้นำเสนอกับอาจารย์ที่ปรึกษาก่อนการวิเคราะห์และปรับปรุงรูปแบบการแนะนำงานต่อไป

ทดสอบเปรียบเทียบประสิทธิภาพ

ทำการวิเคราะห์และเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับพยากรณ์ลูกค้าในการลงทุนสำหรับการแนะนำกองทุน โดยมีขั้นตอนการดำเนินงาน และอธิบายได้ ดังต่อไปนี้

1. ทำการวิเคราะห์ผลลัพธ์ที่ได้ของแต่ละอัลกอริทึม สำหรับพิจารณาการตั้งค่าพารามิเตอร์ที่ได้ประสิทธิภาพที่ดีที่สุด จากชุดข้อมูลที่ใช้ในการวิจัย โดยหลังจากได้ทำการวิเคราะห์และปรับปรุงพารามิเตอร์ แล้วทำการประมวลผลด้วยโปรแกรมแรพพิทไมเนอร์ สตูดิโอ อีกครั้ง

2. ทำการเปรียบเทียบผลลัพธ์ที่ได้ของแต่ละอัลกอริทึม สำหรับการนำไปใช้ในการเรียนรู้แบบกลุ่ม (Ensemble Learning) เพื่อให้ได้ผลลัพธ์ที่ดีที่สุดของแต่ละอัลกอริทึมที่ใช้ในการวิจัย

3. ทำการเปรียบเทียบประสิทธิภาพจากรูปแบบการจัดการข้อมูลทั้ง 2 รูปแบบ เพื่อให้ได้รูปพยากรณ์ลูกค้าในการลงทุนสำหรับการแนะนำกองทุนที่ดีที่สุด

หลังจากนั้นจึงทำการสรุปผลลัพธ์ที่ได้จากการดำเนินการวิจัยและวิเคราะห์ผลเปรียบเทียบประสิทธิภาพของแบบจำลองในลำดับถัดไป

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยเรื่อง การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Development of Customer Predictive Model for Investment using Ensemble Learning Technic) โดยมีรายละเอียดผลการวิเคราะห์ข้อมูลตามความมุ่งหมายของการวิจัย ดังนี้

1. ผลการจัดการข้อมูลตัวแปรที่ใช้ในการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน
2. ผลการเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึม
3. ผลการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน
4. ผลการเปรียบเทียบการจัดการข้อมูลที่ส่งผลต่อประสิทธิภาพของรูปแบบพยากรณ์
5. ผลการเปรียบเทียบประสิทธิภาพของวิธีการใหม่ที่นาเสนอกับงานวิจัยที่ผ่านมาเพื่อยืนยันผลลัพธ์ในการนาไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

ผลการจัดการข้อมูลตัวแปรที่ใช้ในการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

ในการศึกษาข้อมูลตัวแปรที่ใช้ในการพัฒนารูปแบบพยากรณ์ ได้แก่ เพศ ระดับการศึกษา อาชีพ ภูมิลำเนา ระดับรายได้ และผลการศึกษา ซึ่งเป็นข้อมูลประเภท Nominal โดยในการพัฒนาจะใช้โปรแกรมแรพพิทไมเนอร์ สตูดิโอซึ่งมีรายละเอียดดังนี้

ตาราง 9 ข้อมูลตัวแปรที่ใช้ในการพัฒนารูปแบบพยากรณ์ประเภท Nominal

	Value	Count	Fraction
	Branch_Provice		
1	กรุงเทพและปริมณฑล	8666	0.44
2	ภาคเหนือ	2459	0.13
3	ภาคกลาง	2893	0.15
4	ภาคใต้	3427	0.18
5	ภาคตะวันออกเฉียงเหนือ	2131	0.11

ตาราง 9 ข้อมูลตัวแปรที่ใช้ในการพัฒนารูปแบบพยากรณ์ประเภท Nominal (ต่อ)

	Value	Count	Fraction
	Gender		
1	M	13411	0.69
2	F	6165	0.31
	Education		
1	ต่ำกว่าปริญญาตรี	11308	0.58
2	ปริญญาโท	3009	0.15
3	ปริญญาตรี	4860	0.25
4	สูงกว่าปริญญาโท	399	0.02
	Occupation		
1	เอกชน	7726	0.39
2	รับราชการ	1761	0.09
3	อาชีพอิสระ	7894	0.40
4	ธุรกิจส่วนตัว	2195	0.11
	Income		
1	20,001-40,000	8004	0.42
2	60,001-80,000	6927	0.37
3	1-20,000	2306	0.12
4	40,001-60,000	1405	0.07
5	100,000 บาทขึ้นไป	291	0.02
6	80,001-100,000	643	0.03

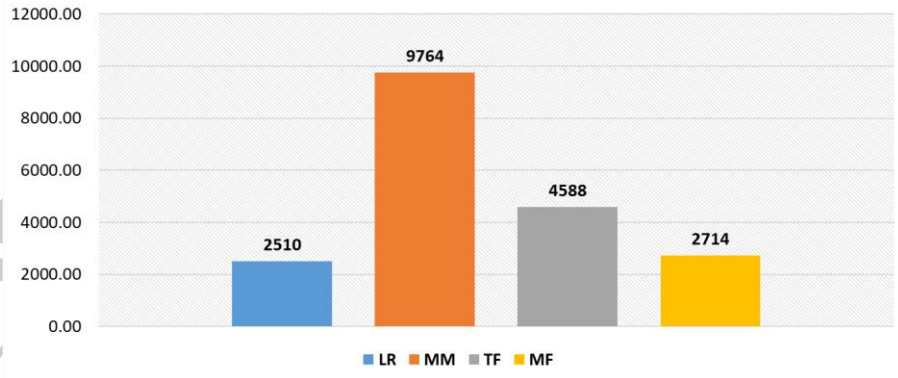
ในส่วนนี้จะเป็นข้อมูลตัวแปรเงินในบัญชีฝากประจำ เงินในบัญชีออมทรัพย์ เงินในบัญชีฝากประจำ เงินในบัญชีฝากประจำประเภทพิเศษ และตัวแปรตามที่กำหนดให้เป็นตัวแปรพยากรณ์ ได้แก่ กองทุนรวม กองทุนตลาดเงิน กองทุนที่มีกำหนดอายุโครงการ และกองทุน RTF/LTF โดยมีรายละเอียดทางสถิติตามตาราง 10

ตาราง 10 ข้อมูลตัวแปรที่ใช้ในการพัฒนารูปแบบพยากรณ์ประเภท Real

	Value	N	Max	Average	Deviation
1	LTF/RMF (LR)	2992	900000.00	39911.17	69195.33
2	Money Market (MM)	11562	9850000.00	794790.31	1171852.40
3	Team Fund (TF)	6441	9472406.41	932043.58	1041179.52
4	Mutual Fund (MF)	3485	8000000.00	270921.09	659896.77
5	Fixed A/C	4664	4873236.58	1179798.62	1128949.89
6	SPA A/C	3986	4572914.77	503593.17	599174.78
7	Saving A/C	19563	4992329.23	747393.16	959785.81

ผลการจัดการข้อมูลตัวแปรสำหรับแบบจำลองการพยากรณ์ประเภทกองทุน

โดยทำดำเนินการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ด้วยเครื่องมือทางด้านการวิเคราะห์ข้อมูลอย่างแรพพิทไมเนอร์ สตูดิโอ โดยกำหนดตัวแปรพยากรณ์เป็นประเภทกองทุน โดยมีรายละเอียดตามภาพประกอบ 25

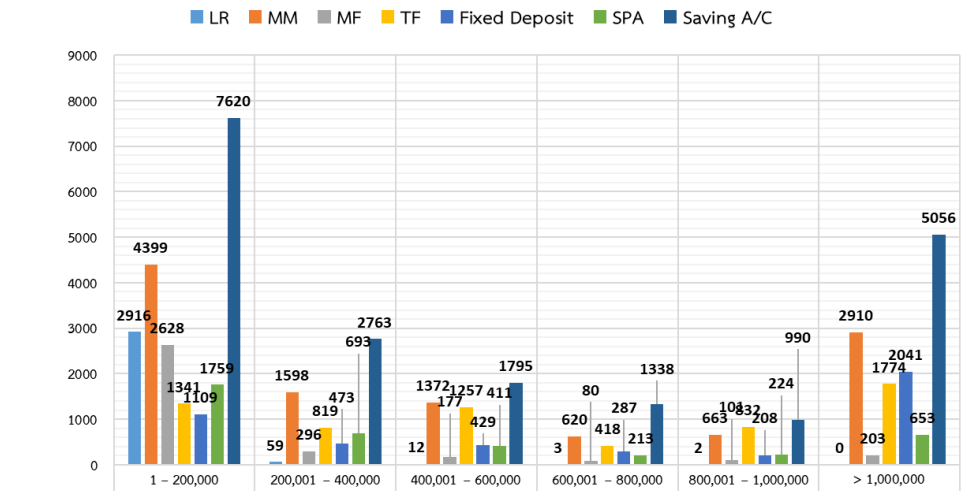


ภาพประกอบ 25 จำนวนของข้อมูลประเภทกองทุน

อธิบายได้ว่า จำนวนข้อมูลลูกค้ากองทุนประเภท MM มีจำนวนมากที่สุด 10,685 คน คิดเป็น 55% กองทุนประเภท TF จำนวน 3,769 คน (19%) กองทุนประเภท LR จำนวน 2,779 คน (14%) และกองทุนประเภท MF จำนวน 2,343 คน (12%) ตามลำดับ

ผลการจัดการข้อมูลตัวแปรรูปแบบที่ 1 สำหรับแบบจำลองพยากรณ์ระดับกองทุน

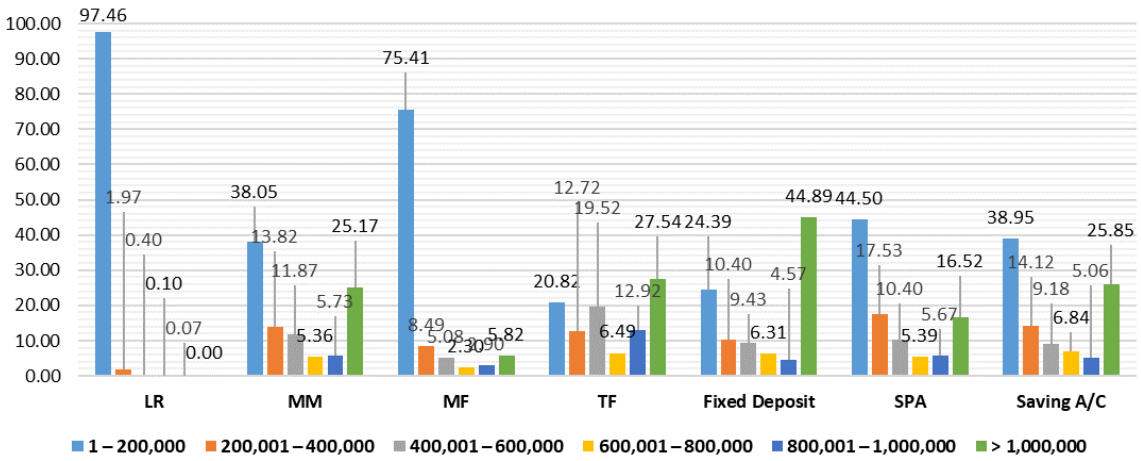
โดยดำเนินการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ด้วยเครื่องมือทางด้านการวิเคราะห์ข้อมูลอย่างแรพพิทไมเนอร์ สตูดิโอให้อยู่ในรูปแบบตัวเลข (Numeric) โดยอธิบายไว้ในตาราง 3.3 และ 3.4 เมื่อทำการแปลงแล้ว ในส่วนของข้อมูลประเภท Nominal จะยังคงมีรูปแบบของข้อมูลไม่เปลี่ยนแปลง ในขณะที่ข้อมูลประเภท Real เป็นการรวบรวมข้อมูลให้อยู่ในลักษณะค่าเฉลี่ยช่วงข้อมูล โดยมีรายละเอียดตามภาพประกอบ 26



	1 – 200,000	200,001 – 400,000	400,001 – 600,000	600,001 – 800,000	800,001 – 1,000,000	> 1,000,000
LR	2916	59	12	3	2	0
MM	4399	1598	1372	620	663	2910
MF	2628	296	177	80	101	203
TF	1341	819	1257	418	832	1774
Fixed Deposit	1109	473	429	287	208	2041
SPA	1759	693	411	213	224	653
Saving A/C	7620	2763	1795	1338	990	5056

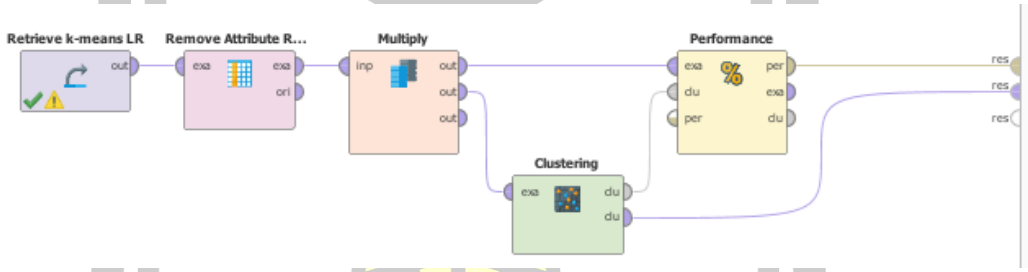
ภาพประกอบ 26 จำนวนของข้อมูลกลุ่มต่างๆ ที่ได้จากการจัดการข้อมูลตามการเฉลี่ยช่วงข้อมูล

จากภาพประกอบ 26 พบว่า ข้อมูลกองทุน LTF/RMF (LR), Mutual Fund (MF), Team Fund (TF), บัญชีเงินฝากประจำประเภทพิเศษ (SPA) และ บัญชีเงินฝากออมทรัพย์ (Saving A/C) จะมีข้อมูลส่วนใหญ่อยู่ในช่วง 1-200,000 คิดเป็นร้อยละ 97.46, 38.05, 75.41, 44.50 และ 38.95 ตามลำดับ ในขณะที่ ข้อมูลกองทุน Mutual Fund (MF) และ บัญชีเงินฝากประจำ (Fixed Deposit) จะมีข้อมูลส่วนใหญ่อยู่ในช่วงมากกว่า 1,000,000 คิดเป็นร้อยละ 27.54 และ 44.89 ตามลำดับ และข้อมูลกองทุน LTF/RMF (LR) ไม่มีข้อมูลในช่วงมากกว่า 1,000,000 ตามภาพประกอบ 27



ภาพประกอบ 27 ร้อยละของปริมาณข้อมูลกลุ่มต่างๆ ที่ได้จากการจัดการข้อมูลตามเฉลี่ยช่วงข้อมูล

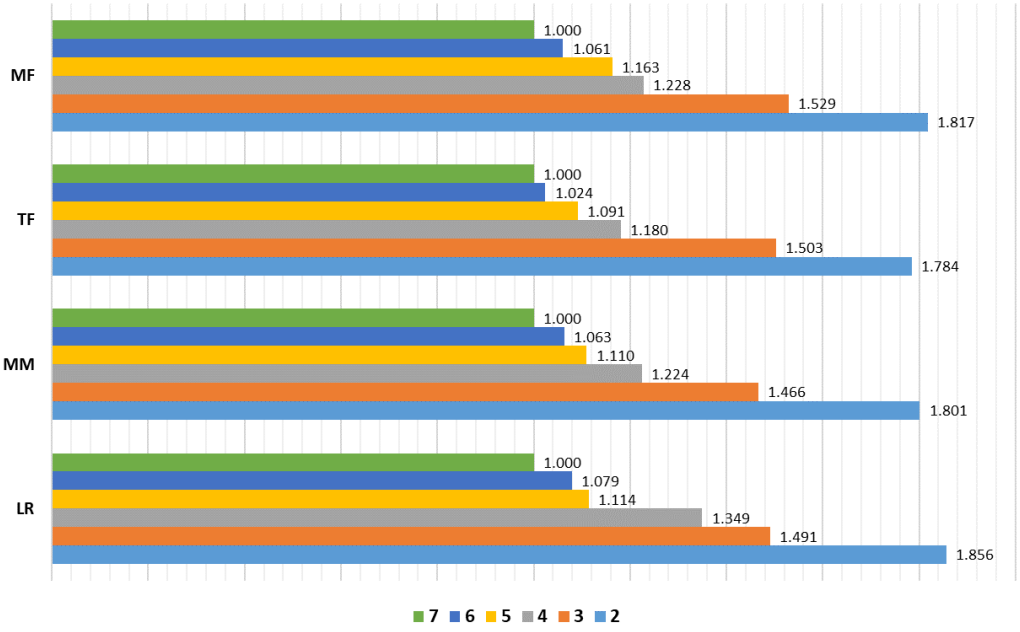
ผลการจัดการข้อมูลตัวแปรรูปแบบที่ 2 สำหรับแบบจำลองการพยากรณ์ระดับกองทุน ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอ เพื่อทำการแบ่งกลุ่มข้อมูลด้วย เคมีนอัลกอริทึมโดยมีกระบวนแบ่งกลุ่มข้อมูลตามภาพประกอบ 28 อธิบายได้ ดังนี้



ภาพประกอบ 28 กระบวนการแบ่งกลุ่มข้อมูลด้วยเคมีนอัลกอริทึม

1. Retrieve เป็นการนำเข้าข้อมูล
2. Remove Attributes Range ใช้สำหรับเลือกแอททริบิวต์ประเภทค่าตัวเลข
3. Multiply สำหรับการนำข้อมูลก่อนหน้าไปใช้ในหลายๆ กระบวนการถัดไป
4. Clustering เป็นโอเปอเรเตอร์ในการแบ่งกลุ่มข้อมูลด้วยเคมีนอัลกอริทึม
5. Performance สำหรับการหาค่าระยะทางเฉลี่ยของแต่ละกลุ่ม (Avg. within Centroid Distance) เพื่อใช้ในการกำหนดค่า K ที่ดีที่สุด โดยการเปรียบเทียบระยะทางเฉลี่ยของแต่ละกลุ่มที่มีค่าเฉลี่ยมากที่สุดของแต่ละการกำหนดค่า K เมื่อพบว่า ระยะทางเฉลี่ยของแต่ละกลุ่มมีการเปลี่ยนแปลงเฉลี่ยไม่มาก ให้ยึดจำนวน K จำนวนนั้น

ผู้วิจัยได้ทำการตั้งค่า K แล้วทำการเปรียบเทียบผลลัพธ์ของจำนวนข้อมูลภายในกลุ่ม สำหรับการหาค่า K ที่ดีที่สุดของแต่ละผลิตภัณฑ์กองทุน (กำหนดจำนวนรอบในการเรียนรู้ = 10) ตามภาพประกอบ 4.3 สามารถระบุได้ว่า K=5 สำหรับกองทุน LR เป็นการกำหนดค่า K ที่ดีที่สุด ลำดับถัดมาคือค่า K=4 สำหรับกองทุนประเภท MF, TF และ MM จากการเปรียบเทียบระยะห่างเฉลี่ยจากการกำหนดค่า K ที่แตกต่างกัน ซึ่งมีรายละเอียดสมาชิกภายในกลุ่มตามตาราง 11 และภาพประกอบ 29 ดังนี้ (Kodinariya & Makwana, 2013)



ภาพประกอบ 29 ค่าระยะทางเฉลี่ยของแต่ละกลุ่ม (Avg. Within Centroid Distance)

ตาราง 11 ค่า Centroid ของแต่ละกลุ่ม และจำนวนสมาชิกภายในกลุ่ม

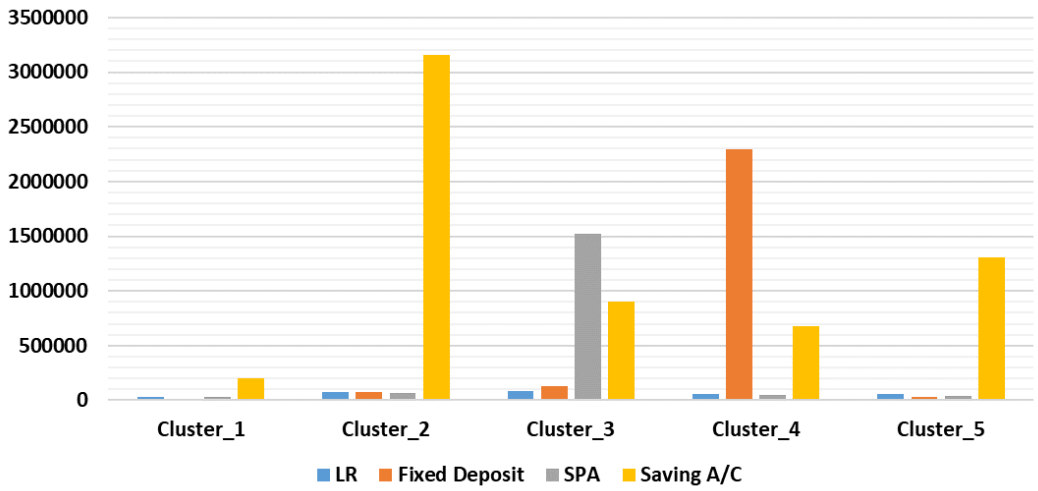
	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
LR					
Count	2093	197	98	103	501
LR	28676.89	77325.82	80907.04	61865.24	59599.40
Fixed Deposit	17370.73	79802.97	132906.60	2292568.34	31803.04
SPA	35731.67	65674.00	1521542.13	46177.72	38970.33

ตาราง 11 ค่า Centroid ของแต่ละกลุ่ม และจำนวนสมาชิกภายในกลุ่ม (ต่อ)

	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
Saving A/C	199843.77	3160726.37	904155.21	676226.34	1309265.93
MM					
Count	6939	713	830	1518	
MM	424414.49	4218739.24	725811.22	948270.25	
Fixed Deposit	71920.97	142461.96	2447713.61	132552.44	
SPA	105381.45	160033.96	111179.36	152948.42	
Saving A/C	362517.24	1130091.15	724981.10	2603007.62	
TF					
Count	3973	474	1186	808	
TF	601159.75	3732733.11	996557.69	821351.63	
Fixed Deposit	104744.09	181795.90	161050.23	2490336.30	
SPA	121476.72	165111.63	133555.73	90400.67	
Saving A/C	459973.64	1121135.21	2692045.88	766678.76	
MF					
Count	2914	112	313	146	
MF	131312.30	3165625.77	474597.92	400114.34	
Fixed Deposit	30914.72	88858.51	102320.64	2524858.27	
SPA	64163.10	110673.89	173321.72	80645.95	
Saving A/C	220805.18	795544.95	2449072.83	518329.30	

จากภาพประกอบ 29 จะเห็นได้ว่า จำนวนสมาชิกภายใน Cluster_1 ของกองทุน LR มีจำนวนมากที่สุด โดยหากเมื่อวิเคราะห์แสดงได้ว่า สมาชิกภายในกลุ่ม Cluster_1 เป็นกลุ่มลูกค้าที่มีจำนวนมากที่สุด สมาชิกภายในกลุ่ม Cluster_2 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากออมทรัพย์มากที่สุด สมาชิกภายในกลุ่ม Cluster_3 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากประจำประเภทพิเศษมากที่สุดและ

บัญชีเงินฝากออมทรัพย์ปานกลาง สมาชิกภายในกลุ่ม Cluster_4 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากประจำมากที่สุดและบัญชีเงินฝากออมทรัพย์ปานกลาง และสมาชิกภายในกลุ่ม Cluster_5 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากออมทรัพย์ปานกลาง

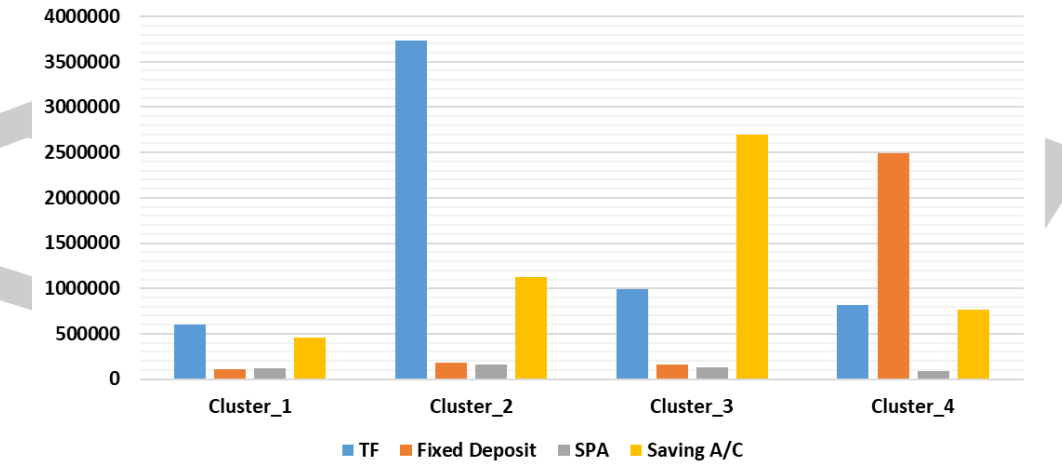


ภาพประกอบ 30 เปรียบเทียบค่า Centroid ของกองทุนประเภท LR

จากภาพประกอบ 30 จะเห็นได้ว่า จำนวนสมาชิกภายใน Cluster_2 ของกองทุน MM มีจำนวนมากที่สุด โดยหากเมื่อวิเคราะห์แสดงได้ว่า สมาชิกภายในกลุ่ม Cluster_1 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากออมทรัพย์มากที่สุดและมูลค่ากองทุนปานกลาง สมาชิกภายในกลุ่ม Cluster_3 เป็นกลุ่มลูกค้าที่มีมูลค่ากองทุนมากที่สุด และสมาชิกภายในกลุ่ม Cluster_4 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากประจำมากที่สุด บัญชีเงินฝากออมทรัพย์ปานกลาง และมูลค่ากองทุนปานกลาง

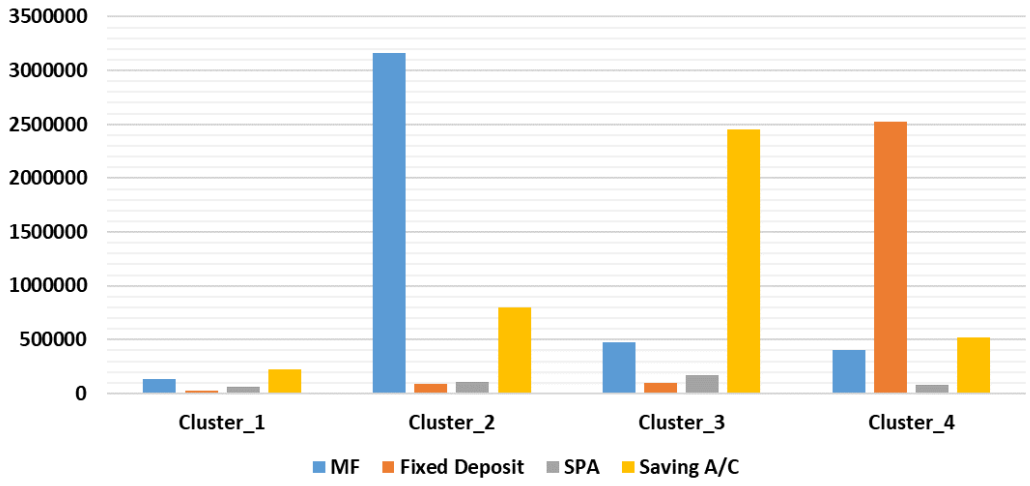
ภาพประกอบ 31 เปรียบเทียบค่า Centroid ของกองทุนประเภท MM

จากภาพประกอบ 31 จะเห็นได้ว่า จำนวนสมาชิกภายใน Cluster_1 ของกองทุน TF มีจำนวนมากที่สุด โดยหากเมื่อวิเคราะห์แสดงได้ว่า สมาชิกภายในกลุ่ม Cluster_2 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากออมทรัพย์ปานกลางและมูลค่ากองทุนมากที่สุด สมาชิกภายในกลุ่ม Cluster_3 เป็นกลุ่มลูกค้าที่มีมูลค่ากองทุนปานกลางและบัญชีเงินฝากออมทรัพย์มากที่สุด และสมาชิกภายในกลุ่ม Cluster_4 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากประจำมากที่สุด บัญชีเงินฝากออมทรัพย์ปานกลาง และมูลค่ากองทุนปานกลาง



ภาพประกอบ 32 เปรียบเทียบค่า Centroid ของกองทุนประเภท TF

จากภาพประกอบ 32 จะเห็นได้ว่า จำนวนสมาชิกภายใน Cluster_1 ของกองทุน MF มีจำนวนมากที่สุด โดยหากเมื่อวิเคราะห์แสดงได้ว่า สมาชิกภายในกลุ่ม Cluster_2 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากออมทรัพย์ปานกลางและมูลค่ากองทุนมากที่สุด สมาชิกภายในกลุ่ม Cluster_3 เป็นกลุ่มลูกค้าที่มีมูลค่ากองทุนปานกลางและบัญชีเงินฝากออมทรัพย์มากที่สุด และสมาชิกภายในกลุ่ม Cluster_4 เป็นกลุ่มลูกค้าที่มีบัญชีเงินฝากประจำมากที่สุด บัญชีเงินฝากออมทรัพย์ปานกลาง และมูลค่ากองทุนปานกลาง



ภาพประกอบ 33 เปรียบเทียบค่า Centroid ของกองทุนประเภท MF

ผลการเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึม

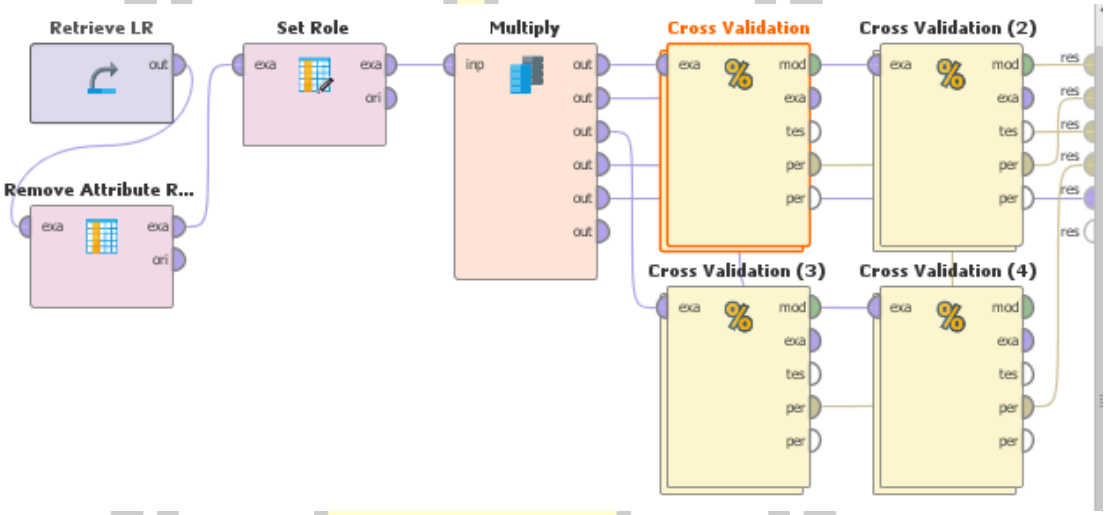
หลังจากได้ทำการแปลงและจัดกลุ่มข้อมูลเรียบร้อยแล้ว ต่อจากนั้นได้ทำการนำข้อมูลดังกล่าวไปผ่านกระบวนการเรียนรู้ด้วยอัลกอริทึมการเรียนรู้ของเครื่อง โดยกำหนดและปรับแต่งค่าพารามิเตอร์ต่าง ๆ สำหรับการเปรียบเทียบประสิทธิภาพ ด้วยวิธีการแบ่งข้อมูล 5 Fold Cross-Validation โดยแสดงผลลัพธ์ทั้ง 2 รูปแบบ ดังนี้

เคเนียร์เรสเนเบอร์ โดยรูปแบบที่ 1

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอ เพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 34 อธิบายได้ ดังนี้

1. Retrieve เป็นการนำเข้าข้อมูลโดยใช้ข้อมูลผลิตภัณฑ์อย่างละ 1 ครั้ง ประมวลผล

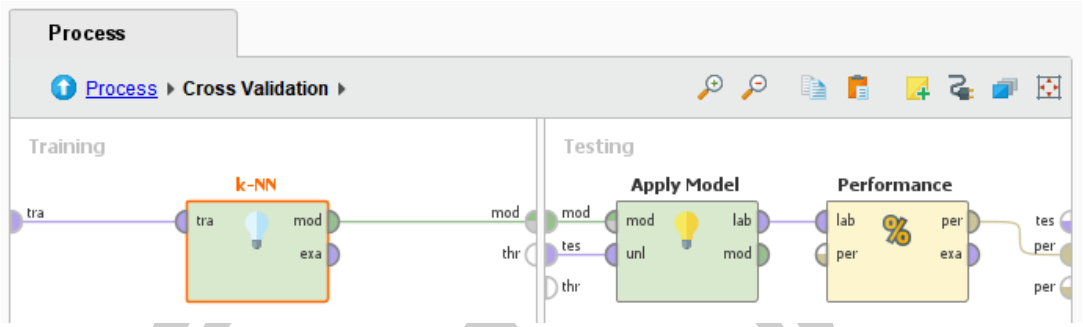
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set Role สำหรับกำหนดตัวแปรที่ต้องการพยากรณ์ หรือกำหนดเป็น Label
4. Multiply สำหรับการนำข้อมูลก่อนหน้าไปใช้ในหลาย ๆ กระบวนการถัดไป
5. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ โดยแบ่งออกเป็น 4 โอเปอเรเตอร์ ด้วยกัน ตามจำนวนพารามิเตอร์ที่ต้องการเปรียบเทียบ ที่นำมาใช้ในงานวิจัย (Number of Folds = 5)



ภาพประกอบ 34 กระบวนการแบ่งกลุ่มข้อมูลด้วยเคเนียร์เรสเนเบอร์

ทำการจัดการส่วยภายใน Cross Validation ทุกโอเปอเรเตอร์โดยมีกระบวนแบ่งกลุ่มตามภาพ 34 อธิบายได้ ดังนี้

1. K-NN สำหรับใช้อัลกอริทึมเคเนียร์เรสเนเบอร์
2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป
3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision



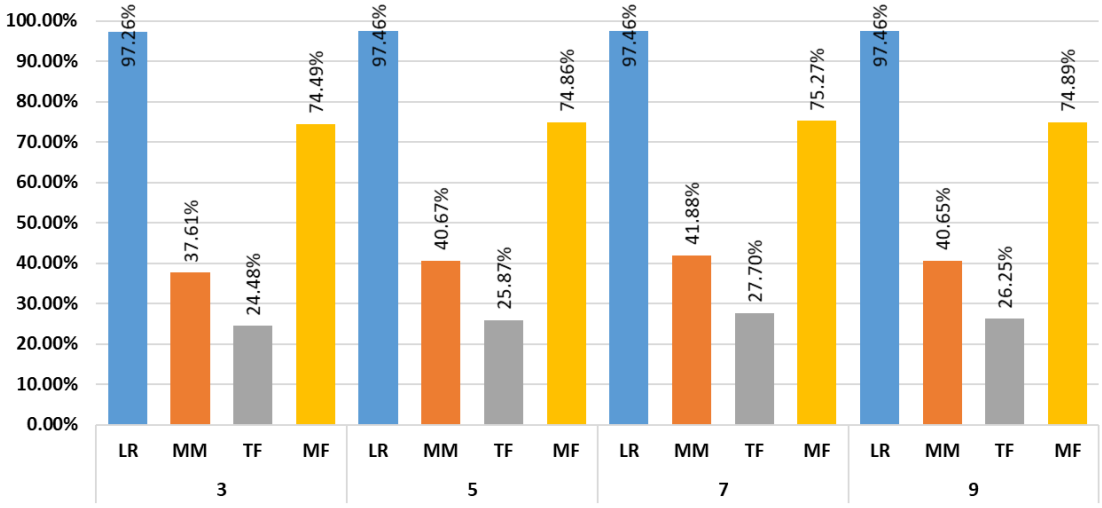
ภาพประกอบ 35 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของเคเนียร์เรสเนเบอร์

หลักจากนั้นทำการประมวนผล โดยสามารถเปรียบเทียบการตั้งค่าพารามิเตอร์ของอัลกอริทึมเคเนียร์เรสเนเบอร์ ได้ตามตาราง 12 ดังนี้

ตาราง 12 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมเคเนียร์เรสเนเบอร์ รูปแบบที่ 1

K	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
3	LR	97.26%	19.96%	19.50%
	MM	37.61%	22.16%	22.83%
	TF	24.48%	20.68%	20.74%
	MF	74.49%	17.38%	21.60%
5	LR	97.46%	20.00%	19.49%
	MM	40.67%	22.71%	23.33%
	TF	25.87%	20.89%	20.59%
	MF	74.86%	16.85%	16.67%
7	LR	97.46%	20.00%	19.49%
	MM	41.88%	22.49%	22.73%
	TF	27.70%	22.06%	22.42%
	MF	75.27%	16.91%	17.44%
9	LR	97.46%	20.00%	19.49%
	MM	40.65%	22.68%	23.22%
	TF	26.25%	21.30%	21.79%
	MF	74.89%	16.80%	16.49%

พบว่า แบบจำลองเคเนียร์เรสเนเบอร์ จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K = 7, 5, 9 และ 3 ตามลำดับ โดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีค่าความแม่นยำน้อยที่สุดโดยมีค่าความแม่นยำเฉลี่ยอยู่ที่ 26.08% แสดงตามภาพประกอบ 36



ภาพประกอบ 36 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์ รูปแบบที่ 1

เคเนียร์เรสเนเบอร์ โดยรูปแบบที่ 2

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับเคเนียร์เรสเนเบอร์ โดยรูปแบบที่ 1 ซึ่งสามารถแสดงผลได้ตามตาราง 13

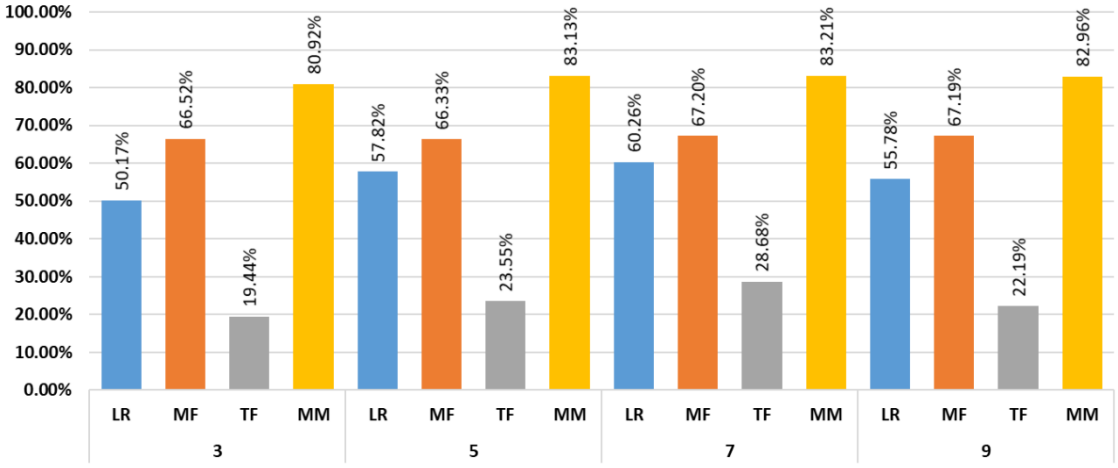
ตาราง 13 เปรียบเทียบประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์ รูปแบบที่ 2

K	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
3	LR	50.17%	22.75%	22.91%
	MM	66.52%	25.88%	27.41%
	TF	19.44%	26.21%	27.10%
	MF	80.92%	26.32%	31.28%

ตาราง 13 เปรียบเทียบประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์ รูปแบบที่ 2 (ต่อ)

K	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
5	LR	57.82%	21.85%	24.53%
	MM	66.33%	25.53%	27.27%
	TF	23.55%	26.05%	26.49%
	MF	83.13%	25.86%	29.01%
7	LR	60.26%	20.07%	19.03%
	MM	67.20%	25.99%	31.11%
	TF	28.68%	26.45%	27.56%
	MF	83.21%	25.24%	25.38%
9	LR	55.78%	22.49%	23.81%
	MM	67.19%	25.74%	28.64%
	TF	22.19%	26.11%	25.34%
	MF	82.96%	25.58%	27.24%

พบว่า แบบจำลองเคเนียร์เรสเนเบอร์ จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K = 7, 5, 9 และ 3 ตามลำดับ โดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีค่าความแม่นยำน้อยที่สุดโดยมีค่าความแม่นยำเฉลี่ยอยู่ที่ 23.47% แสดงตามภาพประกอบ 37



ภาพประกอบ 37 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์ รูปแบบที่ 2

เคเนียร์เรสเนเบอร์ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

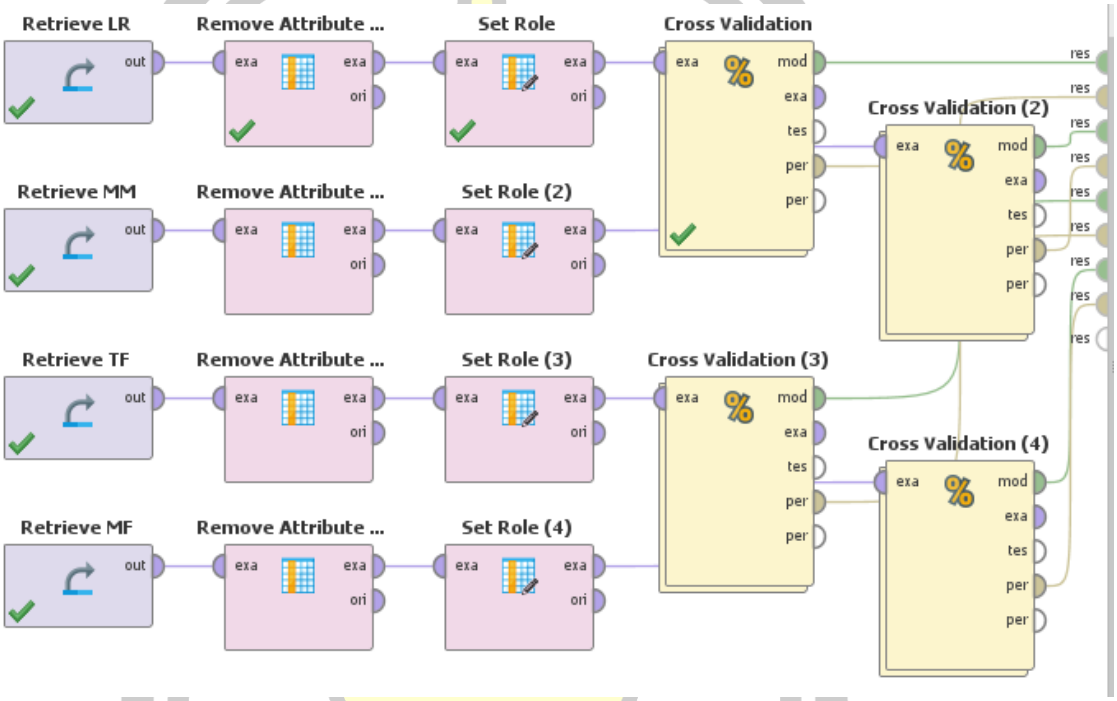
ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับข้างต้น ซึ่งสามารถแสดงผลได้ตามตาราง 4.6 พบว่า แบบจำลองเคเนียร์เรสเนเบอร์ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำมากที่สุดเมื่อตั้งค่า K = 3, 9, 5 และ 7

ตาราง 14 เปรียบเทียบประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

K	Accuracy	Weight Mean Recall	Weight Mean Precision
3	90.94%	89.71%	91.89%
5	75.85%	85.17%	77.74%
7	75.52%	84.87%	77.54%
9	75.92%	85.11%	77.76%

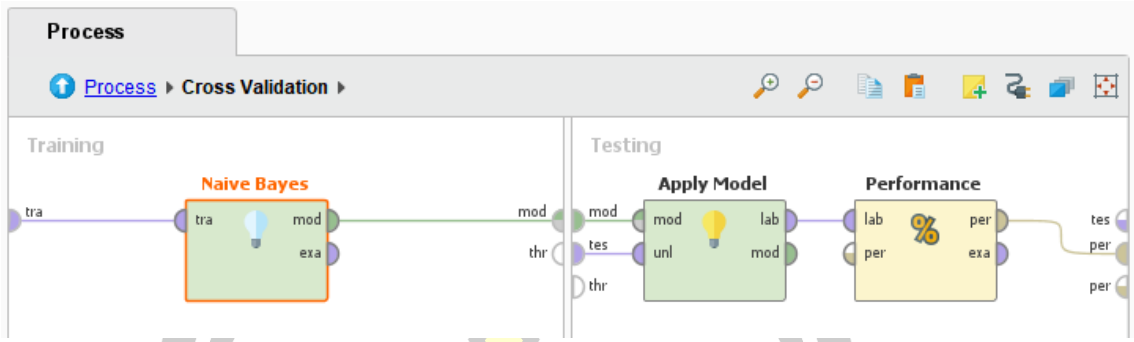
นาอีฟเบย์ โดยรูปแบบที่ 1

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 38 อธิบายได้ ดังนี้



ภาพประกอบ 38 กระบวนการแบ่งกลุ่มข้อมูลด้วยนาอีฟเบย์

1. Retrieve เป็นการนำเข้าข้อมูลโดยแบ่งตามประเภทผลิตภัณฑ์กองทุน
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set role สำหรับกำหนดแอตทริบิวต์ที่ต้องการพยากรณ์
4. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ

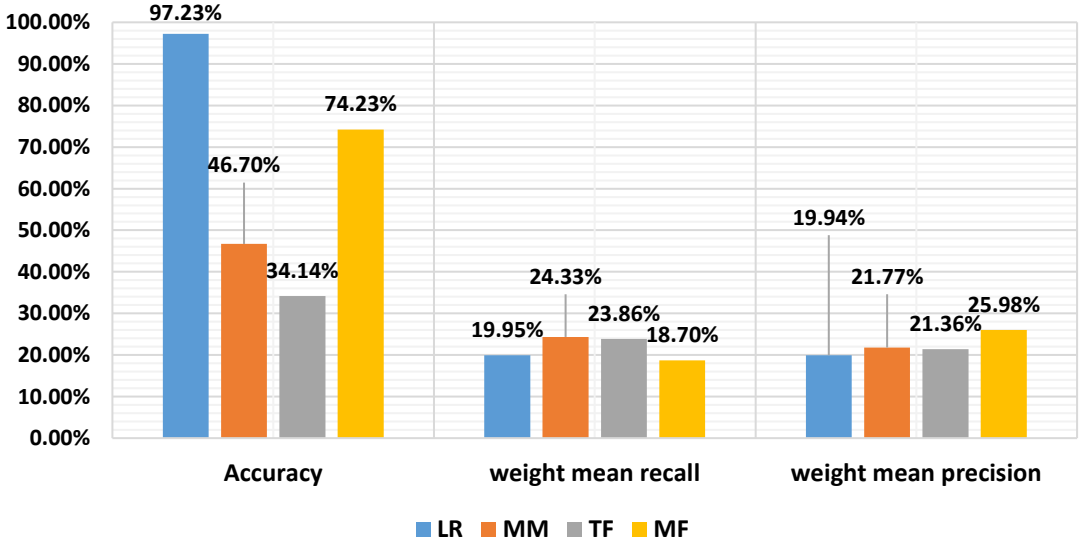


ภาพประกอบ 39 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของนาอีฟเบย์

ทำการจัดการส่วยภายใน Cross Validation ทุกโอเปอเรเตอร์โดยมีกระบวนแบ่งกลุ่มตามภาพ 39 อธิบายได้ ดังนี้

1. Naïve Bayes สำหรับใช้อัลกอริทึมนาอีฟเบย์
2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป
3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision

หลักจากนั้นทำการประมวนผล โดยสามารถแสดงประสิทธิภาพที่ได้จากอัลกอริทึมนาอีฟเบย์ ได้ภาพประกอบ 40 ดังนี้

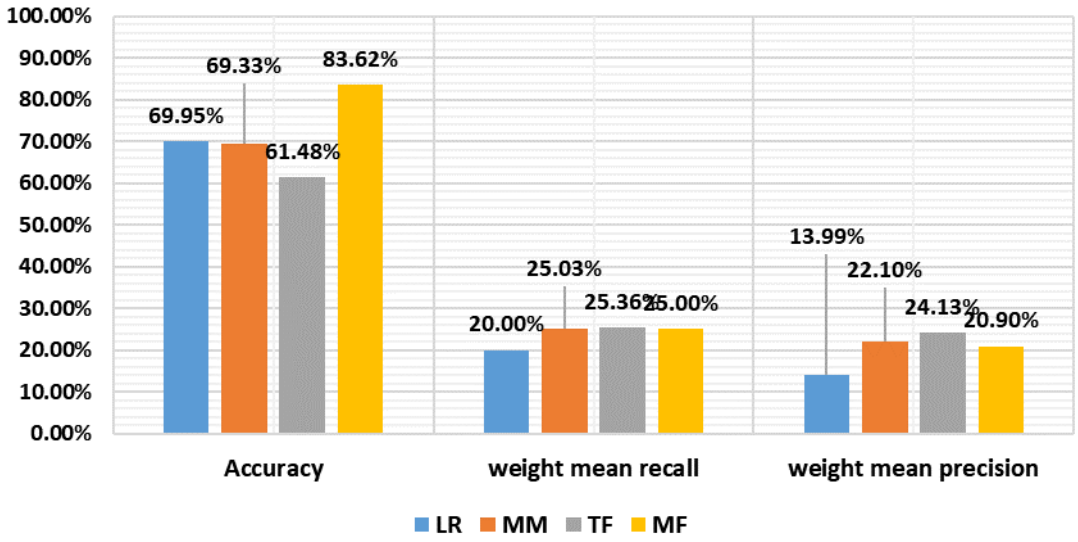


ภาพประกอบ 40 ประสิทธิภาพที่ได้จากอัลกอริทึมนาอีฟเบย์ รูปแบบที่ 1

พบว่า แบบจำลองนาอีฟเบย์จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท LR มากที่สุด (97.23%) และผลิตภัณฑ์กองทุนประเภท MF (74.23%) ตามลำดับ ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีค่าความแม่นยำน้อยที่สุด (34.14%)

นาอีฟเบย์ โดยรูปแบบที่ 2

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอ เพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับนาอีฟเบย์โดยรูปแบบที่ 1 ซึ่งสามารถแสดงผลได้ตามภาพประกอบ 41



ภาพประกอบ 41 ประสิทธิภาพที่ได้จากอัลกอริทึมนาอีฟเบย์ รูปแบบที่ 2

พบว่า แบบจำลองนาอีฟเบย์จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำโดยมี Label เป็นผลิตภัณฑ์กองทุนประเภท MF มากที่สุด (83.62%) และผลิตภัณฑ์กองทุนประเภท LR (69.95%) ตามลำดับ ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีค่าความแม่นยำน้อยที่สุด (61.48%) โดยจะสังเกตได้ว่าค่าความแม่นยำของแต่ละกองทุนจะมีความใกล้เคียงกันมากกว่ารูปแบบที่ 1

นาอีฟเบย์ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับข้างต้น ซึ่งสามารถแสดงผลได้ตามตาราง 15 พบว่า แบบจำลองพยากรณ์ประเภทกองทุนนาอีฟเบย์ มีค่าความแม่นยำเท่ากับ 76.09%

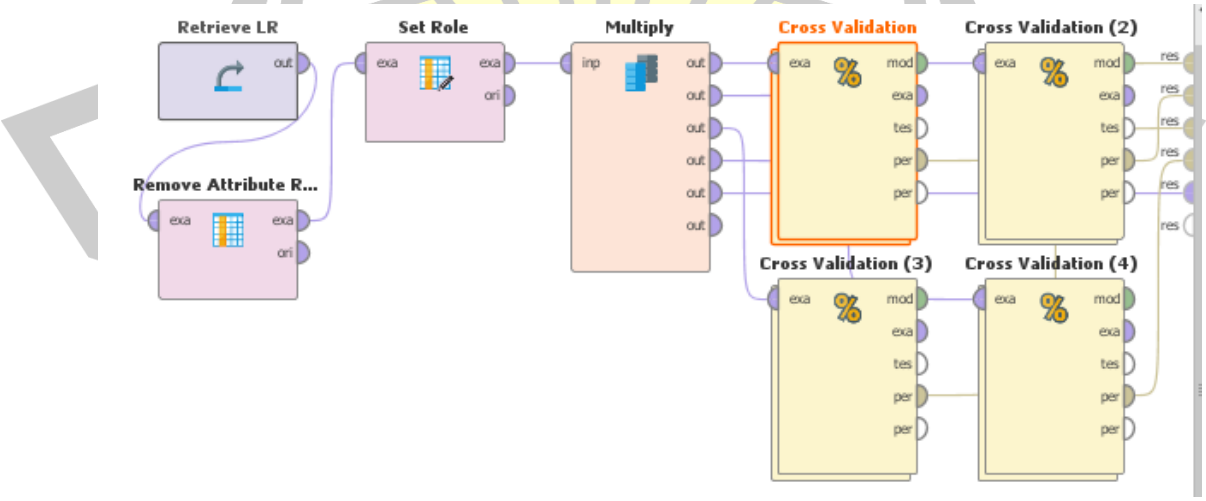
ตาราง 15 ประสิทธิของแบบจำลองพยากรณ์ประเภทกองทุนนาอีฟเบย์

Accuracy	Weight Mean Recall	Weight Mean Precision
76.09%	85.27%	77.93%

ต้นไม้ตัดสินใจ โดยรูปแบบที่ 1

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอ เพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพ 42 อธิบายได้ ดังนี้

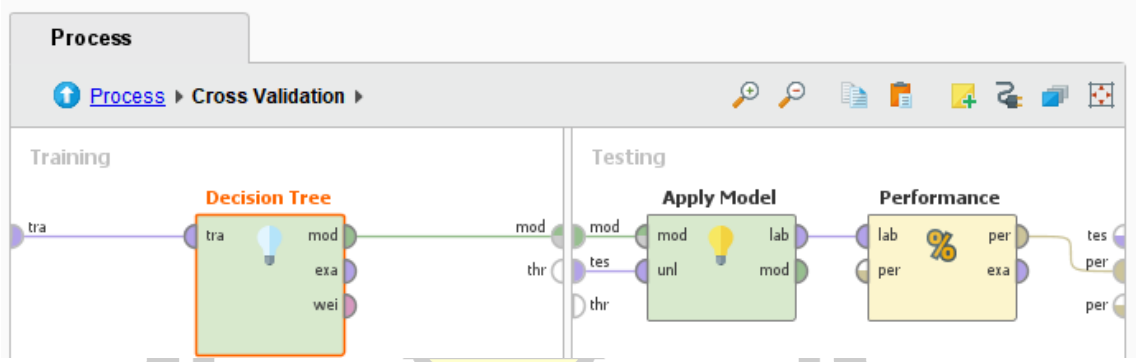
1. Retrieve เป็นการนำเข้าข้อมูลโดยใช้ข้อมูลผลิตภัณฑ์อย่างละ 1 ครั้งประมวลผล
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set role สำหรับกำหนดแอตทริบิวต์ที่ต้องการพยากรณ์
4. Multiply สำหรับการนำข้อมูลก่อนหน้าไปใช้ในหลาย ๆ กระบวนการถัดไป
5. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ โดยแบ่งออกเป็น 4 โอเปอเรเตอร์ด้วยกัน ตามจำนวนพารามิเตอร์ที่ต้องการเปรียบเทียบ ที่นำมาใช้ในงานวิจัย (number of folds = 5)



ภาพประกอบ 42 กระบวนการแบ่งกลุ่มข้อมูลด้วยต้นไม้ตัดสินใจ

ทำการจัดการส่วยภายใน Cross Validation ทุกโอเปอเรเตอร์โดยมีกระบวนแบ่งกลุ่มตามภาพ 43 อธิบายได้ ดังนี้

1. Decision Tree สำหรับใช้อัลกอริทึมต้นไม้ตัดสินใจ โดยกำหนดค่าพารามิเตอร์ที่ต้องการเปรียบเทียบคือ Gain_Ratio, Information_Gain, Gini_Index และ Accuracy
2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป
3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision



ภาพประกอบ 43 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของต้นไม้ตัดสินใจ

หลักจากนั้นทำการประมวนผล โดยสามารถเปรียบเทียบการตั้งค่าพารามิเตอร์ของอัลกอริทึมต้นไม้ตัดสินใจ ได้ตามตาราง 16 ดังนี้

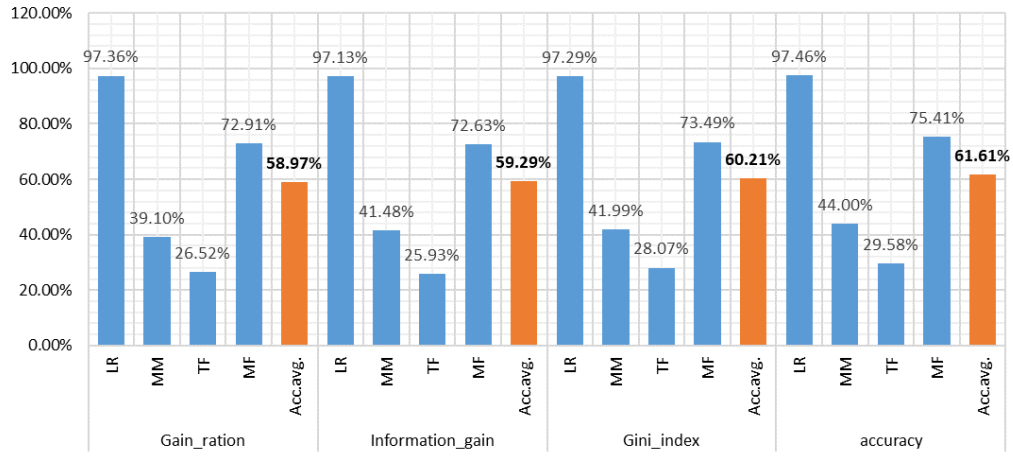
ตาราง 16 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมต้นไม้ตัดสินใจ รูปแบบที่ 1

Criterion	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
Gain_Ration	LR	97.36%	19.98%	19.94%
	MM	39.10%	22.14%	21.85%
	TF	26.52%	21.20%	21.01%
	MF	72.91%	18.32%	20.59%
Information_Gain	LR	97.13%	19.93%	19.94%
	MM	41.48%	23.13%	23.47%
	TF	25.93%	20.73%	20.65%
	MF	72.63%	17.89%	18.29%

ตาราง 16 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมต้นไม้ตัดสินใจ รูปแบบที่ 1 (ต่อ)

Criterion	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
Gini_Index	LR	97.29%	20.29%	20.30%
	MM	41.99%	23.30%	22.68%
	TF	28.07%	21.12%	20.09%
	MF	73.49%	17.72%	19.24%
Accuracy	LR	97.46%	20.00%	19.49%
	MM	44.00%	22.98%	21.03%
	TF	29.58%	21.06%	18.62%
	MF	75.41%	16.67%	12.57%

พบว่า แบบจำลองต้นไม้ตัดสินใจ จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Criterion = Accuracy (61.61%), Gini_Index (60.21%), Information_Gain (59.29%) และ Gain_Ratio (58.97%) ตามลำดับ โดยผลิตภัณฑ์กองทุนประเภท LR มีภาพรวมประสิทธิภาพมากที่สุด ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีประสิทธิภาพแบบจำลองน้อยที่สุด แสดงตามภาพประกอบ 44



ภาพประกอบ 44 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของต้นไม้ตัดสินใจ รูปแบบที่ 1

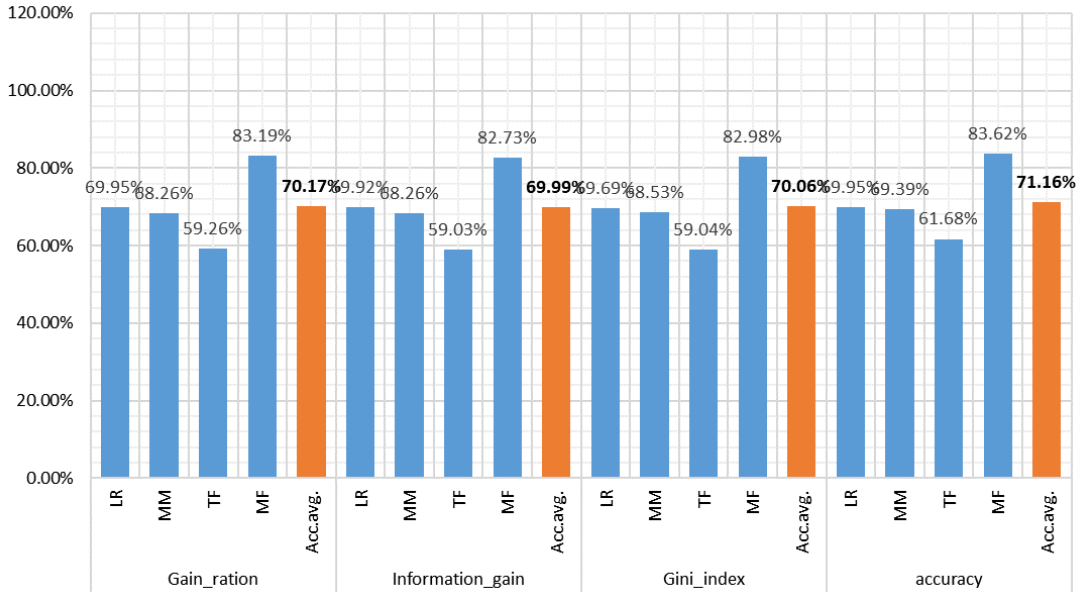
ต้นไม้ตัดสินใจ โดยรูปแบบที่ 2

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอ เพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับต้นไม้ตัดสินใจ โดยรูปแบบที่ 1 ซึ่งสามารถแสดงผลได้ตามตาราง 17

ตาราง 17 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมต้นไม้ตัดสินใจ รูปแบบที่ 2

Criterion	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
Gain_Ration	LR	69.95%	20.00%	13.99%
	MM	68.26%	25.19%	26.31%
	TF	59.26%	25.49%	27.85%
	MF	83.19%	25.90%	37.38%
Information_Gain	LR	69.92%	20.11%	16.34%
	MM	68.26%	25.25%	29.41%
	TF	59.03%	25.72%	27.29%
	MF	82.73%	25.31%	24.80%
Gini_Index	LR	69.69%	19.95%	14.98%
	MM	68.53%	25.43%	30.03%
	TF	59.04%	25.64%	25.83%
	MF	82.98%	25.33%	26.11%
Accuracy	LR	69.95%	20.00%	13.99%
	MM	69.39%	25.00%	17.35%
	TF	61.68%	25.00%	15.42%
	MF	83.62%	25.00%	20.90%

พบว่า แบบจำลองต้นไม้ตัดสินใจ จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Criterion = Accuracy (71.16%), Gain_Ratio (70.17%), Gini_Index (70.06%) และ Information_Gain (69.99%) ตามลำดับ โดยผลิตภันฑ์กองทุนประเภท MF มีภาพรวมประสิทธิภาพมากที่สุด ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีประสิทธิภาพแบบจำลองน้อยที่สุด แสดงตามภาพประกอบ 45



ภาพประกอบ 45 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของต้นไม้ตัดสินใจ รูปแบบที่ 2

ต้นไม้ตัดสินใจ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับข้างต้น ซึ่งสามารถแสดงผลได้ตามตาราง 4.10 พบว่า แบบจำลองต้นไม้ตัดสินใจ สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำมากที่สุด เมื่อตั้งค่า Criterion = Gini_Index (93.28%), Information_Gain (93.23%), Gain_Ratio (92.62%) และ Accuracy (91.90%) ตามลำดับ

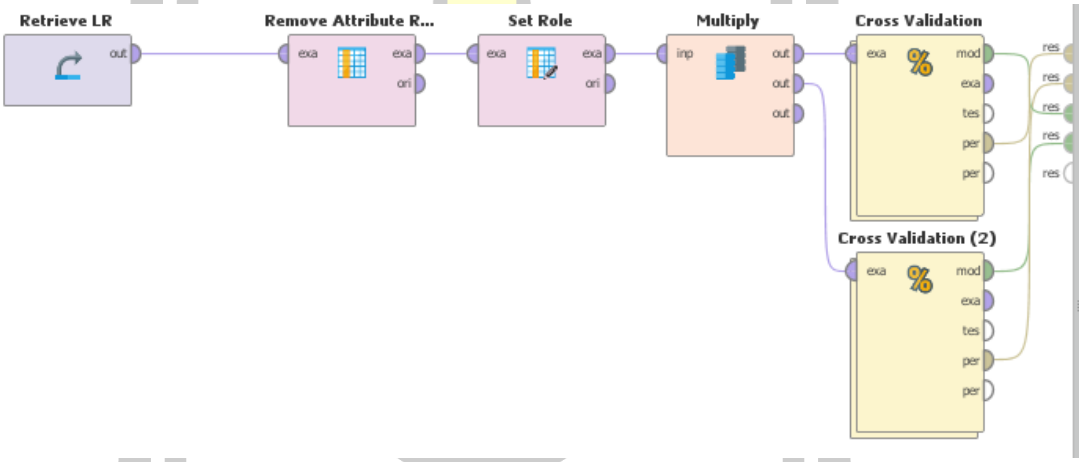
ตาราง 18 ประสิทธิของแบบจำลองพยากรณ์ประเภทกองทุนต้นไม้ตัดสินใจ

Criterion	Accuracy	Weight Mean Recall	Weight Mean Precision
Gain_Ration	92.62%	92.64%	93.71%
Information_Gain	93.23%	92.03%	94.65%
Gini_Index	93.28%	92.50%	94.08%
Accuracy	91.90%	89.64%	95.61%

กฎการอุปนัย โดยรูปแบบที่ 1

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 46 อธิบายได้ ดังนี้

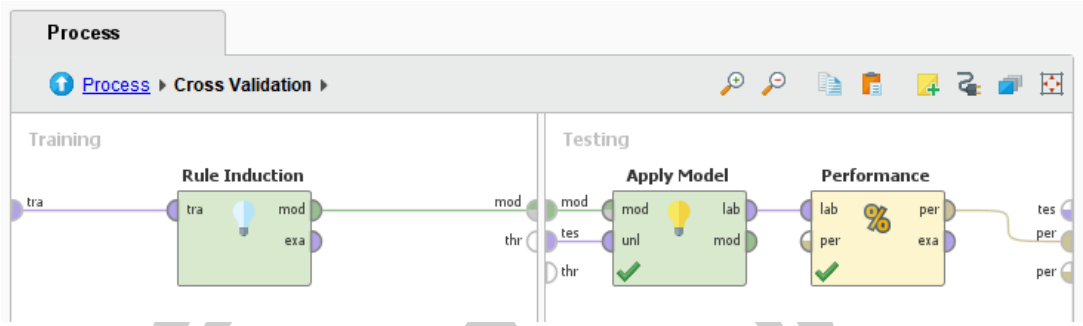
1. Retrieve เป็นการนำเข้าข้อมูลโดยใช้ข้อมูลผลิตภัณฑ์อย่างละ 1 ครั้งประมวลผล
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set role สำหรับกำหนดแอตทริบิวต์ที่ต้องการพยากรณ์
4. Multiply สำหรับการนำข้อมูลก่อนหน้าไปใช้ในหลาย ๆ กระบวนการถัดไป
5. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ โดยแบ่งออกเป็น 2 โอเปอเรเตอร์ตามจำนวนพารามิเตอร์ที่ต้องการเปรียบเทียบ ที่นำมาใช้ในงานวิจัย



ภาพประกอบ 46 กระบวนการแบ่งกลุ่มข้อมูลด้วยกฎการอุปนัย

ทำการจัดการส่วยภายใน Cross Validation ทุกโอเปอเรเตอร์โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 47 อธิบายได้ ดังนี้

1. Rule Induction สำหรับใช้อัลกอริทึมกฎการอุปนัย โดยกำหนดค่าพารามิเตอร์ที่ต้องการเปรียบเทียบคือ Information_Gain และ Accuracy
2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป
3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision



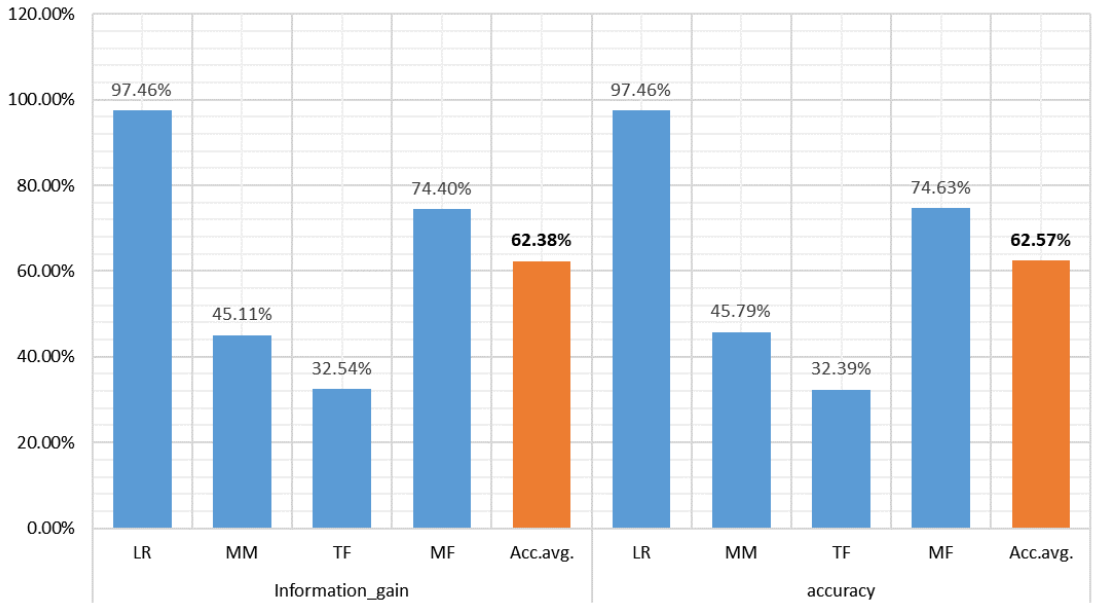
ภาพประกอบ 47 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของกฎการอุปนัย

หลักจากนั้นทำการประมวนผล โดยสามารถเปรียบเทียบการตั้งค่าพารามิเตอร์ของอัลกอริทึม กฎการอุปนัย ได้ตามตาราง 19 ดังนี้

ตาราง 19 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมกฎการอุปนัย รูปแบบที่ 1

Criterion	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
Information_Gain	LR	97.46%	20.00%	19.49%
	MM	45.11%	23.47%	19.59%
	TF	32.54%	22.40%	19.12%
	MF	74.40%	17.28%	18.87%
Accuracy	LR	97.46%	20.00%	19.49%
	MM	45.79%	23.43%	21.61%
	TF	32.39%	22.33%	21.48%
	MF	74.63%	17.46%	17.71%

พบว่า แบบจำลองกฎการอุปนัย จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Criterion = Accuracy (62.57%) และ Information_Gain (62.38%) ตามลำดับ โดยผลิตภัณฑ์กองทุนประเภท LR มีภาพรวมประสิทธิภาพมากที่สุด ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีประสิทธิภาพแบบจำลองน้อยที่สุด แสดงตามภาพประกอบ 48



ภาพประกอบ 48 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของกฎการอุปนัย รูปแบบที่ 1

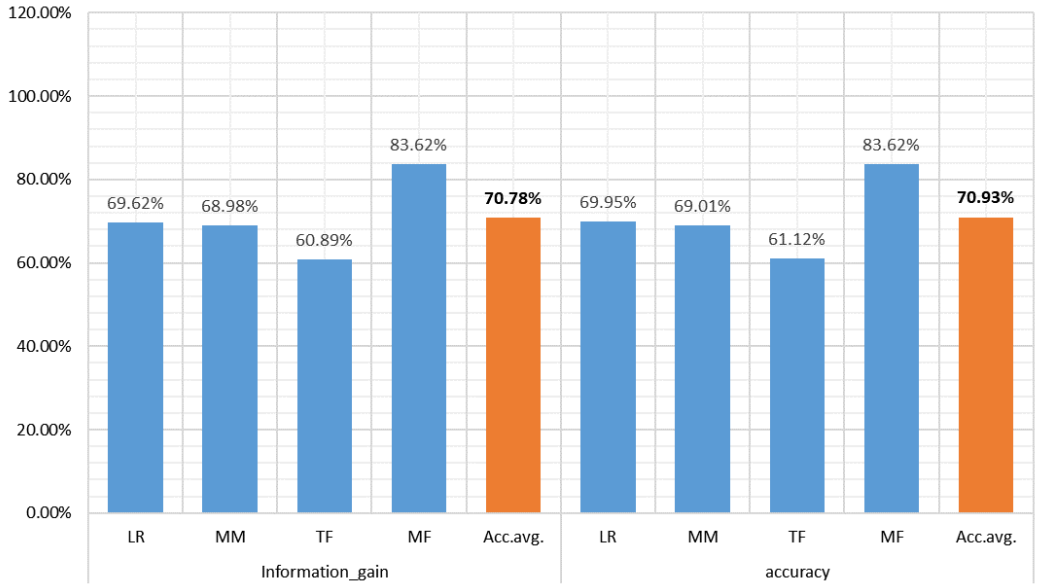
กฎการอุปนัย โดยรูปแบบที่ 2

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับกฎการอุปนัย โดยรูปแบบที่ 1 ซึ่งสามารถแสดงผลได้ตามตาราง 20

ตาราง 20 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมกฎการอุปนัย รูปแบบที่ 2

Criterion	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
Information_Gain	LR	69.62%	19.99%	14.92%
	MM	68.98%	25.20%	28.52%
	TF	60.89%	25.35%	26.38%
	MF	83.62%	25.00%	20.90%
Accuracy	LR	69.95%	20.00%	13.99%
	MM	69.01%	25.15%	26.77%
	TF	61.12%	25.38%	26.65%
	MF	83.62%	25.00%	20.90%

พบว่า แบบจำลองกฎการอุปนัย จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Criterion = Accuracy (70.93%) และ Information_Gain (70.78%) ตามลำดับ โดยผลิตภันฑ์กองทุนประเภท MF มีภาพรวมประสิทธิภาพมากที่สุด ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF มีประสิทธิภาพแบบจำลองน้อยที่สุด แสดงตามภาพประกอบ 49



ภาพประกอบ 49 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของกฎการอุปนัย รูปแบบที่ 2

กฎการอุปนัย สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

โดยดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับข้างต้น ซึ่งสามารถแสดงผลได้ตามตาราง 4.13 พบว่า แบบจำลองกฎการอุปนัย สำหรับแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำมากที่สุด เมื่อตั้งค่า Criterion = Information_Gain (92.16%) และ Accuracy (90.52%) ตามลำดับ

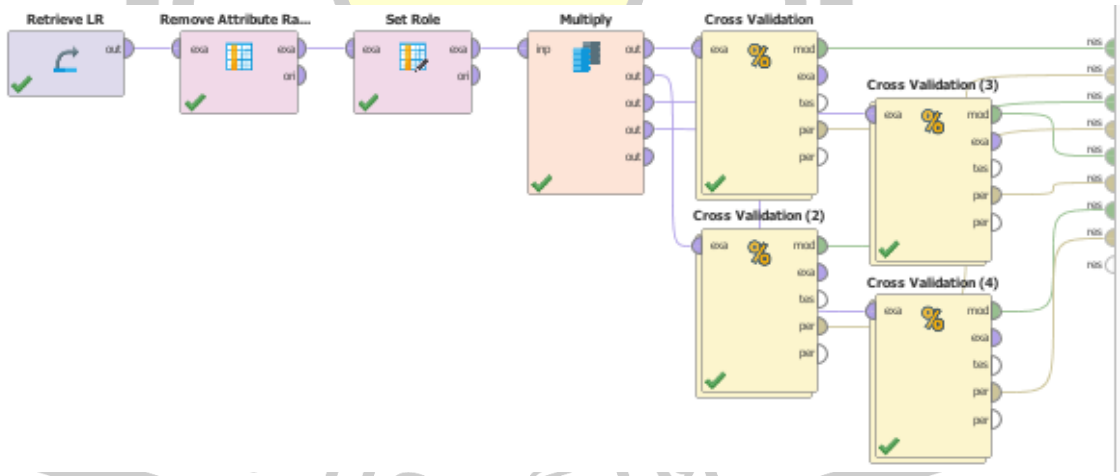
ตาราง 21 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมกฎการอุปนัย รูปแบบที่ 2

Criterion	Accuracy	Weight Mean Recall	Weight Mean Precision
Information_Gain	92.16%	90.11%	94.58%
Accuracy	90.52%	88.20%	94.50%

โครงข่ายประสาทเทียม โดยรูปแบบที่ 1

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 4.26 อธิบายได้ ดังนี้

1. Retrieve เป็นการนำเข้าข้อมูลโดยใช้ข้อมูลผลิตภัณฑ์อย่างละ 1 ครั้งประมวลผล
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set role สำหรับกำหนดแอตทริบิวต์ที่ต้องการพยากรณ์
4. Multiply สำหรับการนำข้อมูลก่อนหน้าไปใช้ในหลาย ๆ กระบวนการถัดไป
5. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ โดยแบ่งออกเป็น 4 โอเปอเรเตอร์ตามจำนวนพารามิเตอร์ที่ต้องการเปรียบเทียบ ที่นำมาใช้ในงานวิจัย



ภาพประกอบ 50 กระบวนการแบ่งกลุ่มข้อมูลด้วยโครงข่ายประสาทเทียม

ทำการจัดการส่วยภายใน Cross Validation ทุกโอเปอเรเตอร์ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 50 อธิบายได้ ดังนี้

1. Deep Learning สำหรับใช้อัลกอริทึมโครงข่ายประสาทเทียม โดยกำหนดค่าพารามิเตอร์ที่ต้องการเปรียบเทียบ จากข้อมูลตามตาราง 22 ได้แก่

ตาราง 22 รายละเอียดค่าสำหรับพิจารณากำหนดค่าพารามิเตอร์

	Label	Input	Output	Number of Hidden Neurons Dize	Number of Hidden Neurons (Input*Output)
Form 1	All	9	5	5-9	45
Form 2	LR	6	5	5-6	30
	MM, TF, MF	6	4	4-6	30
Fund Type Prediction	All	12	1	1-12	12

รูปแบบที่ 1

1.1 Hidden Layer Sizes = 1 Nodes = 9

1.2 Hidden Layer Sizes = 2 Nodes = 9 / 8

1.3 Hidden Layer Sizes = 3 Nodes = 9 / 8 / 7

1.4 Hidden Layer Sizes = 4 Nodes = 9 / 8 / 7 / 6

รูปแบบที่ 2 Label = LR

1.5 Hidden Layer Sizes = 1 Nodes = 6

1.6 Hidden Layer Sizes = 2 Nodes = 6 / 5

1.7 Hidden Layer Sizes = 3 Nodes = 6 / 6 / 5

1.8 Hidden Layer Sizes = 4 Nodes = 6 / 6 / 5 / 5

รูปแบบที่ 2 Label = MM, TF, MF

1.9 Hidden Layer Sizes = 1 Nodes = 6

1.10 Hidden Layer Sizes = 2 Nodes = 6 / 5

1.11 Hidden Layer Sizes = 3 Nodes = 6 / 5 / 4

1.12 Hidden Layer Sizes = 4 Nodes = 6 / 5 / 4 / 4

รูปแบบสำหรับพยากรณ์ประเภทกองทุน

1.13 Hidden Layer Sizes = 1 Nodes = 12

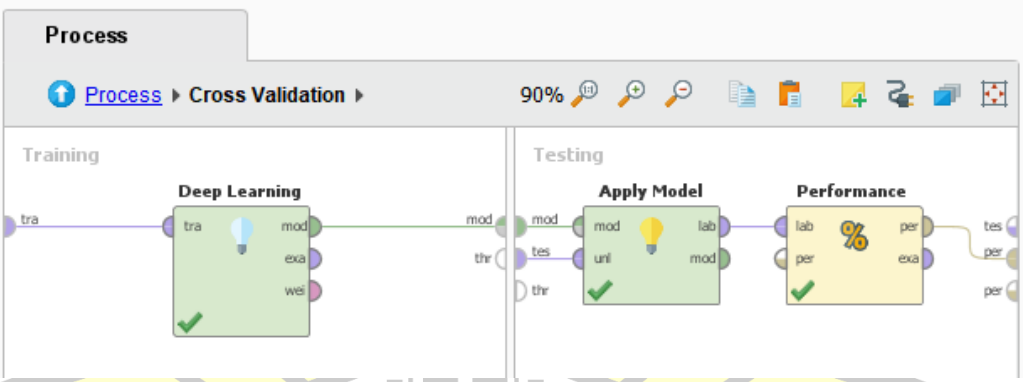
1.14 Hidden Layer Sizes = 2 Nodes = 10 / 2

1.15 Hidden Layer Sizes = 3 Nodes = 6 / 4 / 2

1.16 Hidden Layer Sizes = 4 Nodes = 4 / 3 / 3 / 2

2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป

3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision



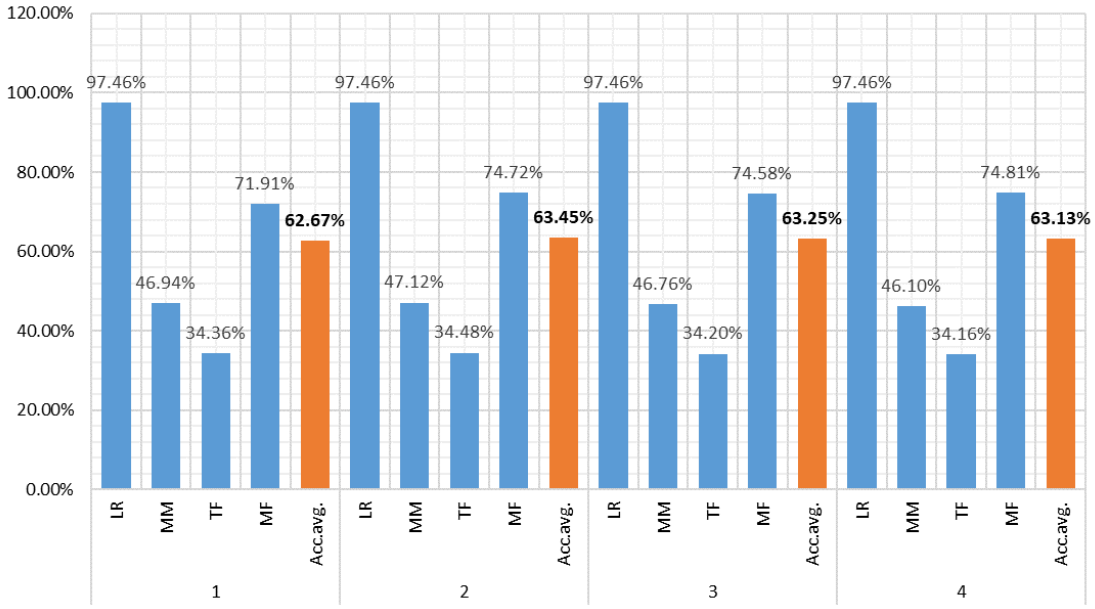
ภาพประกอบ 51 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของโครงข่ายประสาทเทียม

หลักจากนั้นทำการประมวนผล โดยสามารถเปรียบเทียบการตั้งค่าพารามิเตอร์ของอัลกอริทึม โครงข่ายประสาทเทียมได้ตามตาราง 23 ดังนี้

ตาราง 23 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมโครงข่ายประสาทเทียม รูปแบบที่ 1

Hidden layer sizes	Label	Accuracy	weight mean recall	weight mean precision
1	LR	97.46%	20.00%	19.49%
	MM	46.94%	24.09%	20.59%
	TF	34.36%	23.59%	17.27%
	MF	71.91%	17.32%	14.22%
2	LR	97.46%	20.00%	19.49%
	MM	47.12%	24.26%	19.57%
	TF	34.48%	23.41%	17.15%
	MF	74.72%	17.62%	15.22%
3	LR	97.46%	20.00%	19.49%
	MM	46.76%	24.40%	19.54%
	TF	34.20%	23.48%	15.78%
	MF	74.58%	17.32%	13.89%
4	LR	97.46%	20.00%	19.49%
	MM	46.10%	23.77%	19.10%
	TF	34.16%	23.37%	16.72%
	MF	74.81%	17.24%	13.31%

พบว่า แบบจำลองโครงข่ายประสาทเทียมจากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Hidden Layers Size = 2 (63.45%) โดยเมื่อพิจารณาประสิทธิภาพแยกผลิตภันฑ์กองทุน โดย LR มีภาพรวมประสิทธิภาพมากที่สุดและไม่มีการเปลี่ยนแปลงเมื่อเปลี่ยนค่าพารามิเตอร์ ผลิตภัณฑ์กองทุน MM มีค่าพารามิเตอร์ที่ดีที่สุด = 1 ผลิตภัณฑ์กองทุน TF มีภาพรวมประสิทธิภาพน้อยที่สุดและมีค่าพารามิเตอร์ที่ดีที่สุด = 2 และผลิตภัณฑ์กองทุน MF มีค่าพารามิเตอร์ที่ดีที่สุด = 4 แสดงตามภาพประกอบ 52



ภาพประกอบ 52 ความแม่นยำการพยากรณ์และค่าพารามิเตอร์โครงข่ายประสาทเทียมรูปแบบที่ 1

โครงข่ายประสาทเทียม โดยรูปแบบที่ 2

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับโครงข่ายประสาทเทียม โดยรูปแบบที่ 1 ซึ่งสามารถแสดงผลได้ตามตาราง 24

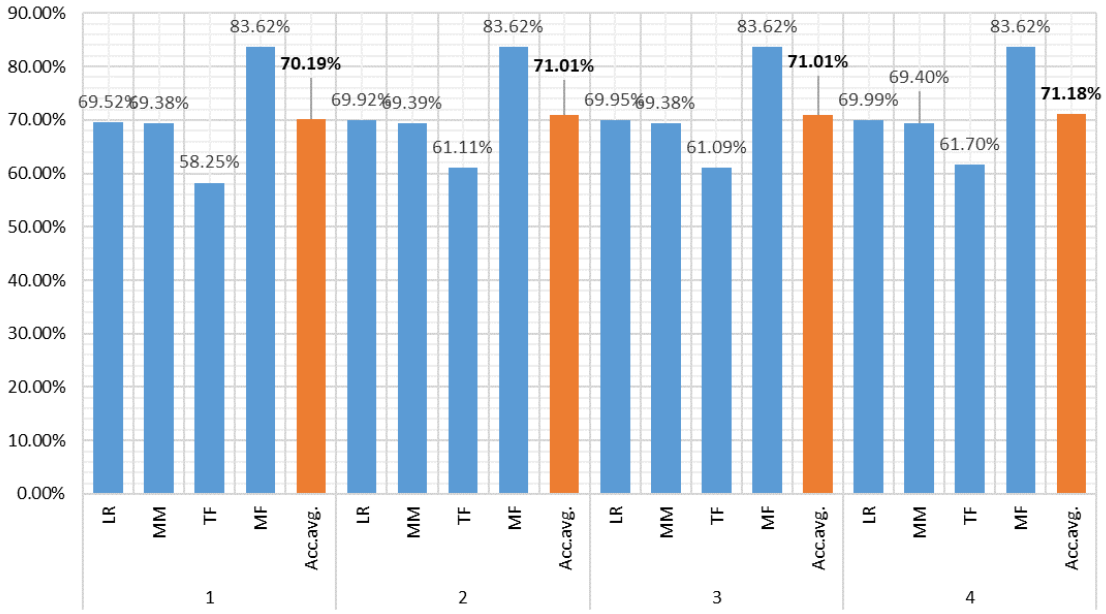
ตาราง 24 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมโครงข่ายประสาทเทียม รูปแบบที่ 2

Hidden Layer Sizes	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
1	LR	69.52%	20.00%	15.32%
	MM	69.38%	25.03%	24.02%
	TF	58.25%	26.68%	24.69%
	MF	83.62%	25.00%	20.90%
2	LR	69.92%	19.99%	13.99%
	MM	69.39%	25.00%	17.35%
	TF	61.11%	25.48%	18.43%
	MF	83.62%	25.00%	20.90%

ตาราง 24 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมโครงข่ายประสาทเทียม รูปแบบที่ 2 (ต่อ)

Hidden Layer Sizes	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
3	LR	69.95%	20.00%	13.99%
	MM	69.38%	25.00%	17.35%
	TF	61.09%	25.15%	22.95%
	MF	83.62%	25.00%	20.90%
4	LR	69.99%	20.07%	16.67%
	MM	69.40%	25.04%	20.35%
	TF	61.70%	25.02%	20.42%
	MF	83.62%	25.00%	20.90%

พบว่า แบบจำลองโครงข่ายประสาทเทียม จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำมากที่สุดเฉลี่ยของทุกผลิตภัณฑ์กองทุน เมื่อตั้งค่า Hidden Layers Size = 4 (71.18%) โดยเมื่อพิจารณาประสิทธิภาพแยกผลิตภันฑ์กองทุน โดย MF มีภาพรวมประสิทธิภาพมากที่สุดและไม่มีการเปลี่ยนแปลงเมื่อเปลี่ยนค่าพารามิเตอร์ ผลิตภัณฑ์กองทุน LR มีค่าพารามิเตอร์ที่ดีที่สุด = 4 ผลิตภัณฑ์กองทุน MM มีค่าพารามิเตอร์ที่ดีที่สุด = 3 และผลิตภัณฑ์กองทุน TF มีภาพรวมประสิทธิภาพน้อยที่สุดและมีค่าพารามิเตอร์ที่ดีที่สุด = 2 แสดงตามภาพประกอบ 53



ภาพประกอบ 53 ความแม่นยำในการพยากรณ์และค่าพารามิเตอร์ของโครงข่ายประสาทเทียม รูปแบบที่ 2

โครงข่ายประสาทเทียม สำหรับแบบจำลองพยากรณ์ประเภทกองทุน

ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มเช่นเดียวกันกับข้างต้น ซึ่งสามารถแสดงผลได้ตามตาราง 4.7 พบว่า แบบจำลองโครงข่ายประสาทเทียม สำหรับแบบจำลองพยากรณ์ประเภทกองทุน เมื่อตั้งค่า Hidden Layers Size = 1 (93.43%) 2, 3 และ 4 ตามลำดับ โดยสังเกตได้ว่า เมื่อมีการเพิ่มขนาดของ Hidden Layers Size แล้วจะทำให้ค่าความแม่นยำลดลง

ตาราง 25 ประสิทธิภาพกับการตั้งค่าพารามิเตอร์ของอัลกอริทึมโครงข่ายประสาทเทียมสำหรับแบบจำลองพยากรณ์ประเภทกองทุน

Hidden Layer Sizes	Accuracy	Weight Mean Recall	Weight Mean Precision
1	93.43%	93.25%	94.09%
2	93.00%	92.31%	93.67%
3	92.60%	92.74%	92.73%
4	92.41	92.21%	92.86%

ผลการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

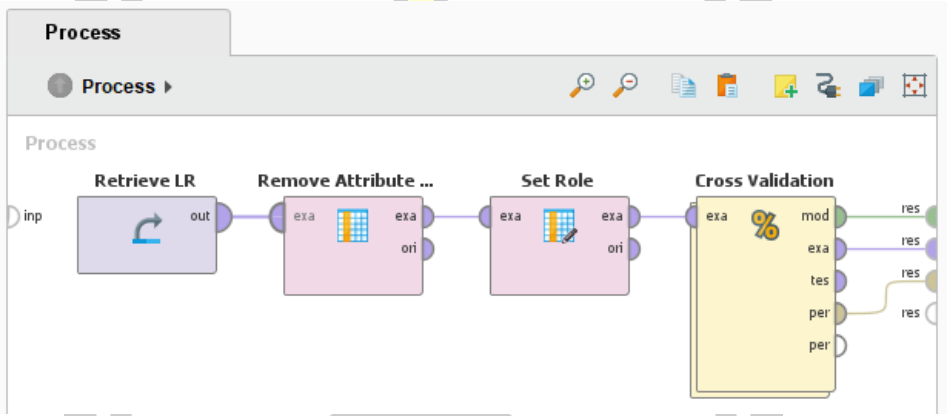
หลังจากทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทั้ง 5 อัลกอริทึม ได้แก่ เคเนียร์เรสเนเบอร์ นาอีฟเบย์ ต้นไม้ตัดสินใจ กฎการอุปนัยและ โครงข่ายประสาทเทียม โดยมีรายละเอียดดังต่อไปนี้ สำหรับเลือกการตั้งค่าพารามิเตอร์ที่ดีที่สุดแล้ว โดยมีรายละเอียดตามตาราง26 มาเป็นการกำหนดพารามิเตอร์ภายใน การเรียนรู้แบบกลุ่มอีกครั้ง ซึ่งผลการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนจะสอดคล้องกับวัตถุประสงค์การวิจัยข้อที่ 1 ซึ่งการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนพิจารณาจากภาพประกอบที่ 20 ในบทที่ 3

ตาราง 26 ตัวเลือกการตั้งค่าพารามิเตอร์การเปรียบเทียบของแต่ละอัลกอริทึม

Algorithm	Label	Parameters			
		Form 1	Form 2	Fund Type Prediction	
K-NN		K			
	LR	5	7	3	
	MM	7	7		
	TF	7	7		
	MF	7	9		
Naïve Bayes		No parameter setting			
Decision Tree		Criterion			
	All	Accuracy	Accuracy	Gini_Index	
Rule Induction		Criterion			
	LR	Accuracy	Accuracy	Information_Gain	
	MM	Accuracy	Accuracy		
	TF	Information_gain	Accuracy		
	MF	Accuracy	Accuracy		
Neural Network		Hidden Layer Size	Nodes	Hidden Layer Size	Nodes	Hidden Layer Size	Nodes
	LR	1	9	4	6/6/5/5	1	12
	MM	1	9	3	6/5/4		
	TF	2	9/8	2	6/5		
	MF	4	9/8/7/6	1	6		

หลังจากนั้น ดำเนินการนำเข้าข้อมูลไปยังโปรแกรมแรพพิทไมเนอร์ สตูดิโอเพื่อทำการพัฒนาและประเมินประสิทธิภาพ โดยมีกระบวนแบ่งกลุ่มตามภาพที่ 54 อธิบายได้ ดังนี้

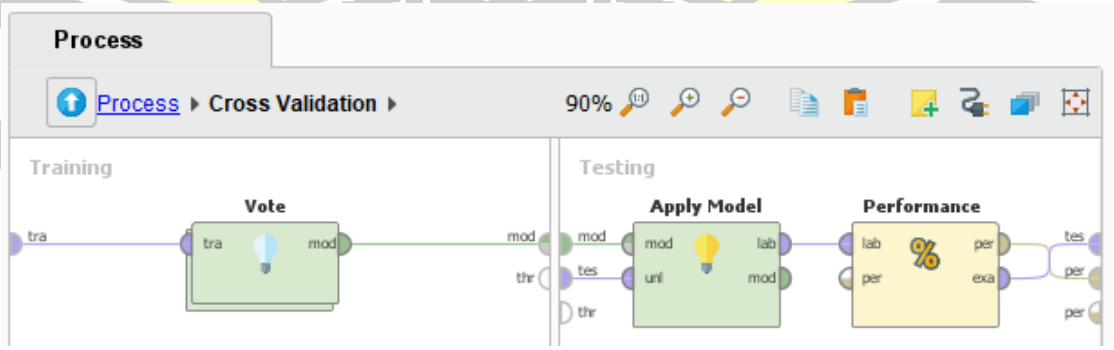
1. Retrieve เป็นการนำเข้าข้อมูลโดยใช้ข้อมูลผลิตภัณฑ์อย่างละ 1 ครั้งประมวลผล
2. Remove Attributes ใช้ตัดข้อมูล Label ที่ไม่ได้ต้องการพยากรณ์
3. Set role สำหรับกำหนดแอตทริบิวต์ที่ต้องการพยากรณ์
4. Cross Validation เป็นโอเปอเรเตอร์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพ



ภาพประกอบ 54 กระบวนการโดยรวมของการเรียนรู้แบบกลุ่ม

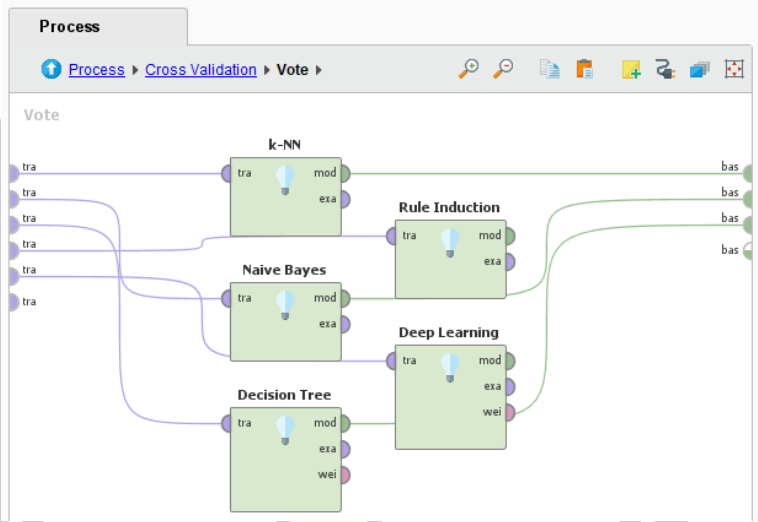
ทำการจัดการส่วยภายใน Cross Validation โดยมีกระบวนตามภาพที่ 55 อธิบายได้ ดังนี้

1. Vote สำหรับใช้อัลกอริทึมการเรียนรู้แบบกลุ่ม
2. Apply Model สำหรับนำผลลัพธ์ที่ได้จากอัลกอริทึมไปใช้งานถัดไป
3. Performance สำหรับการหาค่าประสิทธิภาพของแบบจำลอง โดยต้องการค่าพารามิเตอร์ = Accuracy, Weight Mean Recall และ Weight Mean Precision



ภาพประกอบ 55 กระบวนการแบ่งข้อมูลด้วย Cross Validation ของการเรียนรู้แบบกลุ่ม

ทำการจัดการส่วนภายใน Vote โดยมีอัลกอริทึมที่ใช้ทั้ง 5 อันกอริทึม ได้แก่ เคเนียร์เรสเนเบอร์, นาอีฟเบย์, ต้นไม้ตัดสินใจ, กฎการอุปนัย และโครงข่ายประสาทเทียม โดยทำการกำหนดค่าที่ได้ประสิทธิภาพที่ดีที่สุดจากการเปรียบเทียบในขั้นตอนก่อนหน้า ตามภาพประกอบ 56



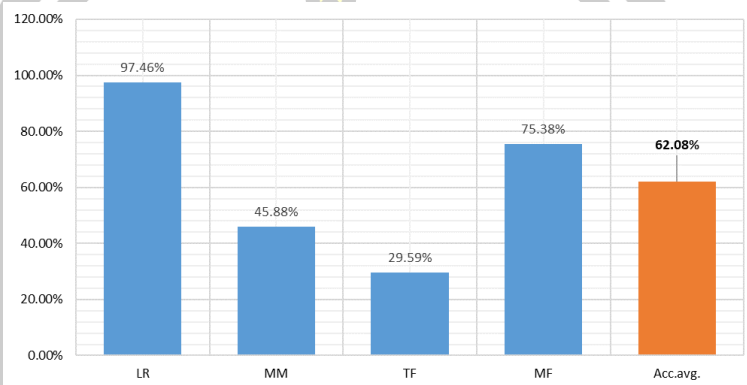
ภาพประกอบ 56 อัลกอริทึมภายใน Vote

หลักจากนั้น ประมวลผลข้อมูลตามจำนวน Label ที่ใช้ในการพยากรณ์ โดยสามารถแสดงผลได้ ตามตาราง 27 ดังนี้

ตาราง 27 แสดงประสิทธิภาพแบบจำลองของการเรียนรู้แบบกลุ่ม

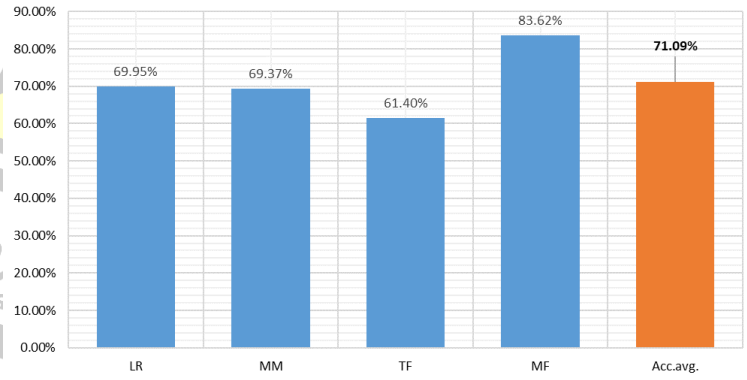
Form	Label	Accuracy	Weight Mean Recall	Weight Mean Precision
1	LR	97.46%	20.00%	19.49%
	MM	45.88%	23.62%	22.26%
	TF	29.59%	22.86%	27.02%
	MF	75.38%	16.66%	12.57%
2	LR	69.95%	20.00%	13.99%
	MM	69.37%	25.02%	20.68%
	TF	61.40%	25.11%	24.95%
	MF	83.62%	25.00%	20.90%
Fund Type Prediction	All	92.38%	92.59%	92.78%

พบว่า แบบจำลองการเรียนรู้แบบกลุ่ม จากการจัดการข้อมูลในรูปแบบที่ 1 มีค่าความแม่นยำเฉลี่ยของทุกผลิตภัณฑ์กองทุนเท่ากับ 62.08% โดยเมื่อพิจารณาประสิทธิภาพแยกผลิตภันฑ์กองทุน โดย LR มีค่าความแม่นยำมากที่สุด (97.46%) และผลิตภัณฑ์กองทุน MF, MM และ TF ตามลำดับ แสดงตามภาพประกอบ 57



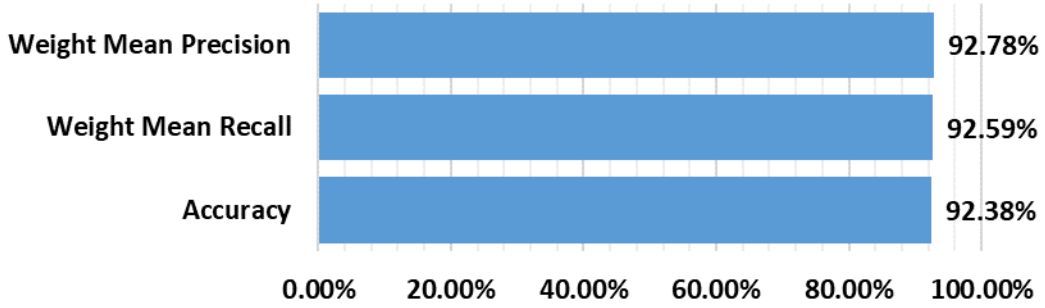
ภาพประกอบ 57 ความแม่นยำในการพยากรณ์ของการเรียนรู้แบบกลุ่ม รูปแบบที่ 1

และแบบจำลองการเรียนรู้แบบกลุ่ม จากการจัดการข้อมูลในรูปแบบที่ 2 มีค่าความแม่นยำเฉลี่ยของทุกผลิตภัณฑ์กองทุนเท่ากับ 71.09% โดยเมื่อพิจารณาประสิทธิภาพแยกผลิตภันฑ์กองทุน โดย MF มีค่าความแม่นยำมากที่สุด (83.62%) และผลิตภัณฑ์กองทุน LR, MM และ TF ตามลำดับ แสดงตามภาพประกอบ 58



ภาพประกอบ 58 ความแม่นยำในการพยากรณ์ของการเรียนรู้แบบกลุ่ม รูปแบบที่ 2

สุดท้าย แบบจำลองการเรียนรู้แบบกลุ่มสำหรับการพยากรณ์ประเภทกองทุน มีค่าความแม่นยำเท่ากับ 92.38% ซึ่งจะใช้สำหรับพยากรณ์ประเภทกองทุน หรือเป็นแบบจำลองพยากรณ์ที่ใช้สำหรับกำหนดแบบจำลองระดับของกองทุนต่อไป แสดงตามภาพประกอบ 59



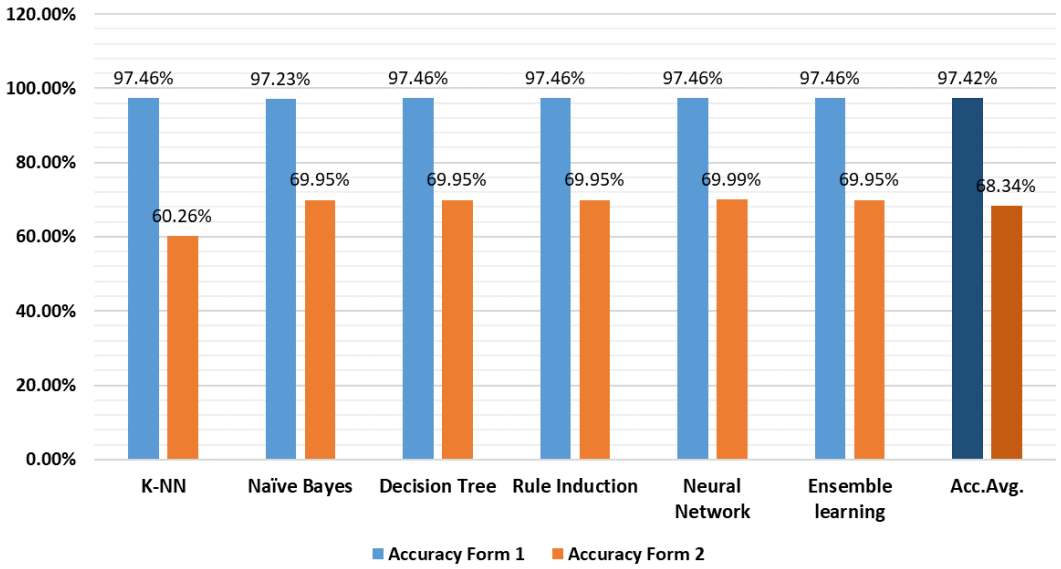
ภาพประกอบ 59 ความแม่นยำในการพยากรณ์ของการเรียนรู้แบบกลุ่ม สำหรับการพยากรณ์ประเภทกองทุน

ผลการเปรียบเทียบการจัดการข้อมูลที่ส่งผลต่อประสิทธิภาพของรูปแบบพยากรณ์

ผู้วิจัยทำการเปรียบเทียบค่าประสิทธิภาพ ในทุกขั้นตอนการการพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม จากการกำหนดหรือปรับแต่งค่าพารามิเตอร์ สามารถแสดงผลโดยผู้วิจัยนำเสนอแยกตามประเภทของผลิตภัณฑ์กองทุนได้ซึ่งผลการเปรียบเทียบการจัดการข้อมูลที่ส่งผลต่อประสิทธิภาพของรูปแบบพยากรณ์จะสอดคล้องกับวัตถุประสงค์การวิจัยข้อที่ 2 โดยมีรายละเอียดดังต่อไปนี้

ผลิตภัณฑ์กองทุนประเภท LR

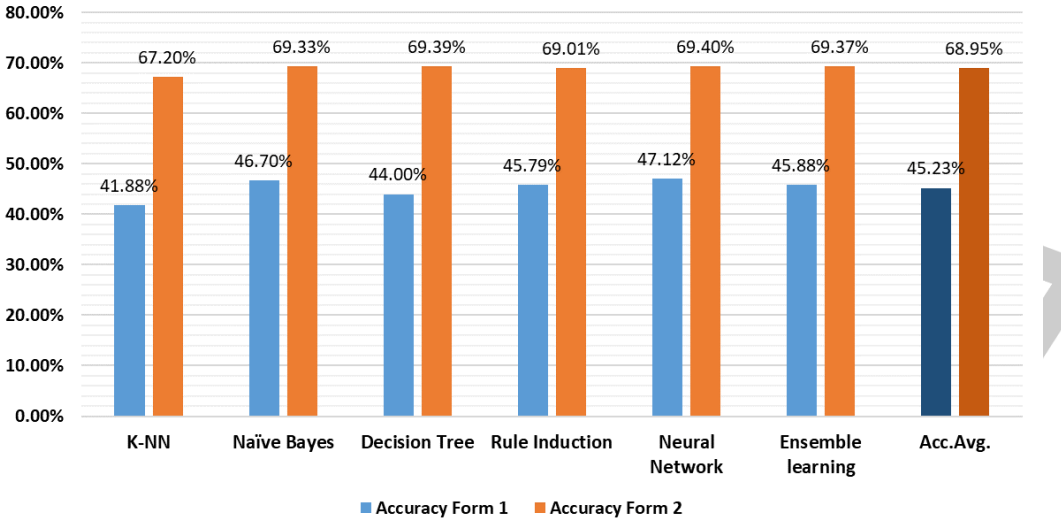
อธิบายได้ว่า ผลิตภัณฑ์กองทุนประเภท LR จากการจัดการข้อมูลรูปแบบที่ 1 มีประสิทธิภาพเฉลี่ย 97.42% มากกว่ารูปแบบที่ 2 ที่มีค่าเฉลี่ยอยู่ที่ 68.34% ในทุกอัลกอริทึม โดยมีค่าความต่างเฉลี่ยอยู่ที่ 29.08% ตามภาพประกอบ 60



ภาพประกอบ 60 ความแม่นยำในการพยากรณ์ผลิตภัณฑ์กองทุนประเภท LR จากการจัดการข้อมูล

ผลิตภัณฑ์กองทุนประเภท MM

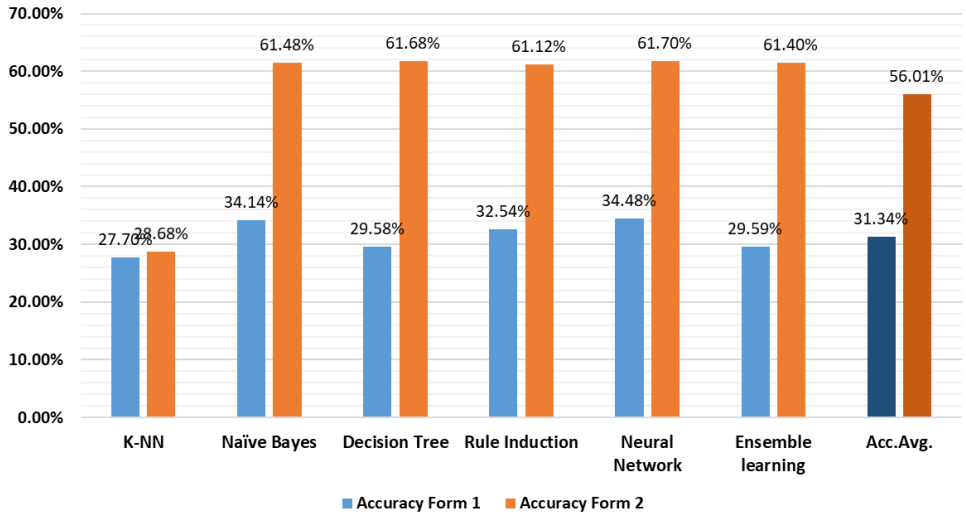
อธิบายได้ว่า ผลิตภัณฑ์กองทุนประเภท MM จากการจัดการข้อมูลรูปแบบที่ 2 มีประสิทธิภาพเฉลี่ย 68.95% มากกว่ารูปแบบที่ 1 ที่มีค่าเฉลี่ยอยู่ที่ 45.23% ในทุกอัลกอริทึม โดยมีค่าความต่างเฉลี่ยอยู่ที่ 23.72% ตามภาพประกอบ 61



ภาพประกอบ 61 ความแม่นยำในการพยากรณ์ผลิตภัณฑ์กองทุนประเภท MM จากการจัดการข้อมูล

ผลิตภัณฑ์กองทุนประเภท TF

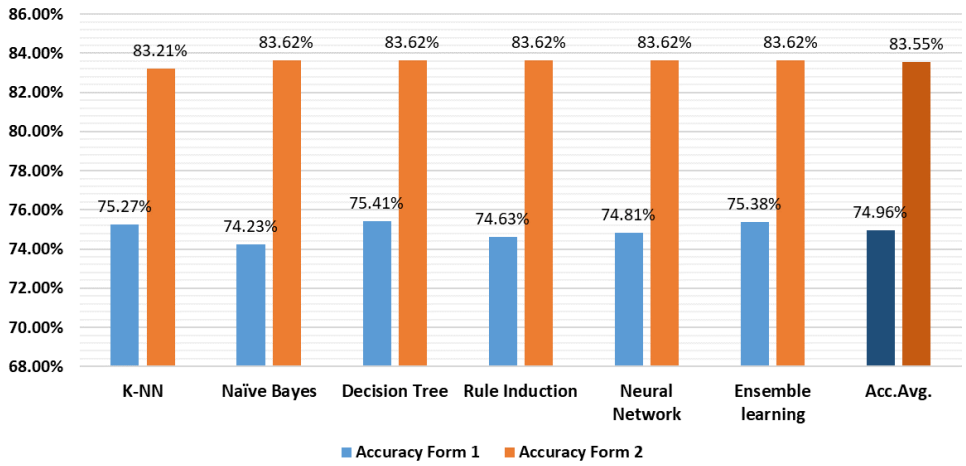
อธิบายได้ว่า ผลิตภัณฑ์กองทุนประเภท TF จากการจัดการข้อมูลรูปแบบที่ 2 มีประสิทธิภาพเฉลี่ย 56.01% มากกว่ารูปแบบที่ 1 ที่มีค่าเฉลี่ยอยู่ที่ 31.34% ในทุกอัลกอริทึม โดยมีค่าความต่างเฉลี่ยอยู่ที่ 24.67% ตามภาพประกอบ 62



ภาพประกอบ 62 ความแม่นยำในการพยากรณ์ผลิตภัณฑ์กองทุนประเภท TF จากการจัดการข้อมูล

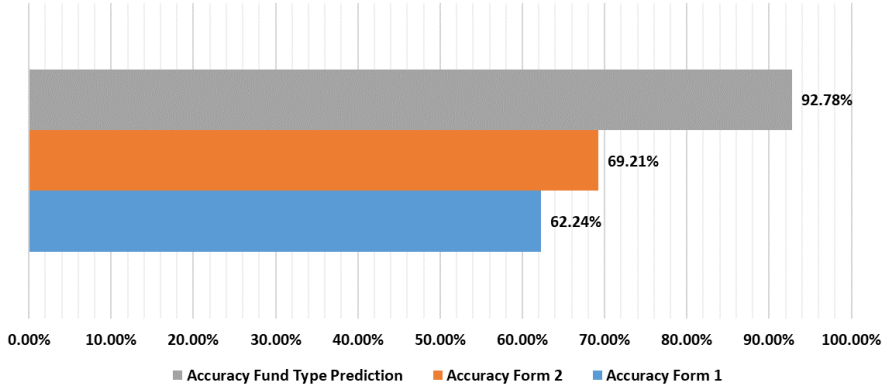
ผลิตภัณฑ์กองทุนประเภท MF

อธิบายได้ว่า ผลิตภัณฑ์กองทุนประเภท MF จากการจัดการข้อมูลรูปแบบที่ 2 มีประสิทธิภาพเฉลี่ย 83.55% มากกว่ารูปแบบที่ 1 ที่มีค่าเฉลี่ยอยู่ที่ 74.96% ในทุกอัลกอริทึม โดยมีค่าความต่างเฉลี่ยอยู่ที่ 8.60%% ตามภาพประกอบ 63



ภาพประกอบ 63 ความแม่นยำในการพยากรณ์ผลิตภัณฑ์กองทุนประเภท MF จากการจัดการข้อมูล

หากนำผลประสิทธิภาพของทุกผลิตภัณฑ์กองทุนมาเฉลี่ยแล้วพบว่า การจัดการข้อมูลรูปแบบที่ 1 (62.24%) ก่อนการพัฒนาแบบจำลอง มีประสิทธิภาพในการพยากรณ์น้อยว่ารูปแบบการจัดการข้อมูลรูปแบบที่ 2 (69.21%) เท่ากับ 6.98% และแบบจำลองพยากรณ์ประเภทกองทุนมีค่าแม่นยำมากที่สุด เท่ากับ 92.78% ตามภาพประกอบ 64



ภาพประกอบ 64 เปรียบเทียบความแม่นยำเฉลี่ยในการพยากรณ์ผลิตภัณฑ์กองทุนทุกประเภท

ผลการเปรียบเทียบประสิทธิภาพของวิธีการกับงานวิจัยที่ผ่านมา

ภายในงานวิจัย ผู้วิจัยนำผลลัพธ์ประสิทธิภาพของแบบจำลอง ทำการเปรียบเทียบประสิทธิภาพของวิธีการที่นำเสนอภายในส่วนงานวิจัยที่เกี่ยวของ เพื่อยืนยันผลลัพธ์ในการนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน จากตาราง 27 พบว่าประสิทธิภาพความแม่นยำของแบบจำลองที่นำเสนอในงานวิจัยนี้มีประสิทธิภาพดีกว่าแบบจำลองจากงานวิจัยที่ผ่านมาซึ่งผลการเปรียบเทียบประสิทธิภาพของวิธีการกับงานวิจัยที่ผ่านมาจะสอดคล้องกับวัตถุประสงค์การวิจัยข้อที่ 3 โดยมีรายละเอียดดังต่อไปนี้ โดยสามารถอธิบายและแสดงผลตามตาราง 28 ดังนี้

ตาราง 28 เปรียบเทียบประสิทธิภาพของวิธีการกับงานวิจัยที่ผ่านมา

งานวิจัย	วัตถุประสงค์	อัลกอริทึม	ค่าความแม่นยำของแบบจำลอง	เปรียบเทียบค่าความแม่นยำ
Atthaporn Sapaphan (2016)	ออกแบบระบบช่วยตัดสินใจเลือกซื้อผลิตภัณฑ์กองทุน	Scale, Decision Tree	Accuracy = 86.91%	มากกว่า 5.87%

ตาราง 28 เปรียบเทียบประสิทธิภาพของวิธีการกับงานวิจัยที่ผ่านมา (ต่อ)

งานวิจัย	วัตถุประสงค์	อัลกอริทึม	ค่าความแม่นยำของแบบจำลอง	เปรียบเทียบค่าความแม่นยำ
Thongchai Kaewkiriya (2017)	กรอบการพยากรณ์ลูกค้าประกันชีวิตตาม Multi-Algorithm	Decision Tree, Neural Network	Accuracy = 92.73%	มากกว่า 0.05%
Jaiswal, Devendra Prakash, et al. (2020)	ระบบการพยากรณ์ธุรกรรมของลูกค้า	Deep Learning	Accuracy = 91.13%	มากกว่า 1.65%
Tao, Tao, et al. (2019, May)	จำแนกประเภทการลงทุนในกองทุนรวม	XGBoost	Accuracy = 90.13%	มากกว่า 2.65%
Mashayekhi, Morteza, et al. (2018)	พยากรณ์การสมัครและถอนในกองทุนรวม	Logistic Regression	Accuracy = 88.40%	มากกว่า 4.38%
Saloni Kumari (2021)	จำแนกและการพยากรณ์โรคเบาหวาน	Majority Vote	Accuracy = 79.08%	มากกว่า 13.70%

บทที่ 5
สรุปผล อภิปราย และข้อเสนอแนะ

การวิจัยเรื่อง การพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม (Development of Customer Predictive Model for Investment using Ensemble Learning Technic) โดยมีรายละเอียดที่สามารถสรุป อภิปราย และข้อเสนอแนะ ดังนี้

1. ความมุ่งหมายของการวิจัย
2. สรุปผลการวิจัย
3. อภิปรายผลการวิจัย
4. ข้อเสนอแนะ

ความมุ่งหมายของการวิจัย

1. เพื่อพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม
2. เพื่อเปรียบเทียบประสิทธิภาพอัลกอริทึมที่ใช้ในแบบจำลองการพยากรณ์ลูกค้า
3. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการใหม่ที่นำเสนอกับงานวิจัยที่ผ่านมาเพื่อยืนยันผลลัพธ์ในการนำไปพัฒนาแบบจำลองการพยากรณ์ลูกค้าในการลงทุน

สรุปผลการวิจัย

ในการพัฒนารูปแบบการพยากรณ์ลูกค้าในการลงทุนโดยใช้เทคนิคการเรียนรู้แบบกลุ่ม โดยเลือกรูปแบบการโหวตจากการพยากรณ์ของ 5 อัลกอริทึม ได้แก่ เคเนียร์เรสเนเบอร์ นาอีฟเบย์ ต้นไม้ตัดสินใจ กฎการอุปนัยและ โครงข่ายประสาทเทียม โดยการเปรียบเทียบการจัดการข้อมูลที่ส่งผลต่อประสิทธิภาพของรูปแบบพยากรณ์ ได้แก่ การจัดการข้อมูลด้วยรูปแบบเฉลี่ยช่วงข้อมูล (รูปแบบที่ 1) มีค่าความแม่นยำเฉลี่ยเท่ากับ 62.24% และการแบ่งกลุ่มข้อมูลด้วยเคมีน (รูปแบบที่ 2) มีค่าความแม่นยำเฉลี่ยเท่ากับ 69.21%

พิจารณารายผลิตภัณฑ์กองทุน

กองทุนประเภท LR มีค่าความแม่นยำมากที่สุดเฉลี่ยเท่ากับ 97.42% ในรูปแบบการจัดการข้อมูลแบบเฉลี่ยช่วงข้อมูล และกองทุนประเภท MM, TF และ MF มีค่าความแม่นยำมากที่สุดเฉลี่ยเท่ากับ 68.95%, 56.91% และ 83.55% ตามลำดับ ในรูปแบบการจัดการข้อมูลโดยการแบ่งกลุ่มข้อมูล เคมีน

พิจารณารายอัลกอริทึม

เพื่อเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมที่ใช้ในการพัฒนารูปแบบพยากรณ์ พบว่า รูปแบบพยากรณ์ผลิตภัณฑ์กองทุนประเภท LR อัลกอริทึม เคเนียร์เรสเนเบอร์ ต้นไม้ตัดสินใจ กฎการอุปนัย โครงข่ายประสาทเทียมและเมื่อใช้เทคนิคการเรียนรู้แบบกลุ่ม มีค่าความแม่นยำมากที่สุดเท่ากันที่ 97.46% รูปแบบพยากรณ์ผลิตภัณฑ์กองทุนประเภท MM อัลกอริทึมที่มีค่าความแม่นยำตามลำดับ ได้แก่ โครงข่ายประสาทเทียม (69.40%), ต้นไม้ตัดสินใจ (69.39%), นาอีฟเบย์(69.33%), กฎการอุปนัย (69.01%) และเคเนียร์เรสเนเบอร์ (67.20%) โดยเมื่อใช้เทคนิคการเรียนรู้แบบกลุ่ม มีค่าความแม่นยำ 68.95% รูปแบบพยากรณ์ผลิตภัณฑ์กองทุนประเภท TF อัลกอริทึมที่มีค่าความแม่นยำตามลำดับ ได้แก่ Neural Network (61.70%), ต้นไม้ตัดสินใจ (61.68%), นาอีฟเบย์ (61.48%), กฎการอุปนัย (61.12%) และเคเนียร์เรสเนเบอร์ (28.68%) โดยเมื่อใช้เทคนิคการเรียนรู้แบบกลุ่ม มีค่าความแม่นยำ 61.40% และรูปแบบพยากรณ์ผลิตภัณฑ์กองทุนประเภท MF อัลกอริทึมที่มีค่าความแม่นยำมากที่สุด ได้แก่ นาอีฟเบย์ ต้นไม้ตัดสินใจ กฎการอุปนัย และโครงข่ายประสาทเทียม และเมื่อใช้เทคนิคการเรียนรู้แบบกลุ่ม มีค่าความแม่นยำมากที่สุดเท่ากันที่ 83.62% ในขณะที่อัลกอริทึมเคเนียร์เรสเนเบอร์ ค่าความแม่นยำเท่ากับ 83.21%

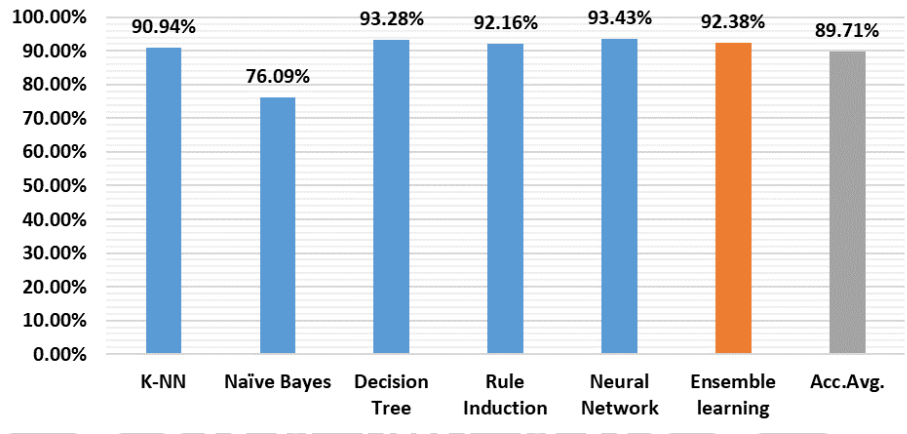
พิจารณาการกำหนดค่าพารามิเตอร์

สำหรับการตั้งค่าพารามิเตอร์ก่อนการนำไปพัฒนารูปแบบพยากรณ์ อธิบายได้ว่า อัลกอริทึมเคเนียร์เรสเนเบอร์ จากการตั้งค่าพารามิเตอร์ค่า K หรือจำนวนข้อมูลที่นำมาประมวลผลกลุ่ม โดย K= 7 เป็นค่าพารามิเตอร์ที่ดีที่สุดใน 3 ประเภทผลิตภัณฑ์กองทุนทั้งรูปแบบที่ 1 และรูปแบบที่ 2 ในขณะที่ผลิตภัณฑ์กองทุนประเภท LR ในรูปแบบที่ 1 พารามิเตอร์ที่ดีที่สุดคือ K=5 และผลิตภัณฑ์กองทุนประเภท MF ในรูปแบบที่ 2 พารามิเตอร์ที่ดีที่สุดคือ K=9 อัลกอริทึมนาอีฟเบย์ไม่มีการตั้งค่าพารามิเตอร์ อัลกอริทึมต้นไม้ตัดสินใจจากการตั้งค่าพารามิเตอร์ Criterion หรือเกณฑ์ในการจำแนกข้อมูล โดย Accuracy เป็นพารามิเตอร์ที่ดีที่สุดทั้ง 2 รูปแบบ อัลกอริทึมกฎการอุปนัยจากการตั้งค่าพารามิเตอร์ Criterion หรือเกณฑ์ในการจำแนกข้อมูล โดย Accuracy เป็นพารามิเตอร์

ที่ดีที่สุดทั้ง 2 รูปแบบ แต่ผลิตภัณฑ์กองทุนประเภท TF ในรูปแบบที่ 1 พารามิเตอร์ที่ดีที่สุดคือ Information_gain และอัลกอริทึมโครงข่ายประสาทเทียม จากการตั้งค่าพารามิเตอร์ค่า Hidden Layer Side หรือขนาดเลเยอร์ภายใน Hidden Layer และ Nodes หรือจำนวนโหนดของแต่ละ Hidden Layers โดยแต่ละผลิตภัณฑ์กองทุนมีค่าพารามิเตอร์ที่แตกต่างกันออกไป ผลิตภัณฑ์กองทุนประเภท LR MM TF และ MK ในรูปแบบที่ 1 มี Hidden Layer Size และ Node ได้แก 1 (9), 1 (9), 2 (9/8) และ 4 (9/8/7/6) ตามลำดับ และในรูปแบบที่ 2 มี Hidden Layer Size และ Node ได้แก 4 (6/6/5/5), 3 (6/5/4), 2 (6/5) และ 1 (6) ตามลำดับ

พิจารณาแบบจำลองพยากรณ์ประเภทกองทุน

โดยแบบจำลองพยากรณ์ประเภทกองทุน เป็นแบบจำลองที่ช่วยในการเลือกแบบจำลองระดับกองทุนที่มีสามารถบอกระดับหรือประกอบการพิจารณา แนะนำผลิตภัณฑ์กองทุน ซึ่งเมื่อดำเนินการพัฒนาแบบจำลองดังกล่าวด้วยเทคนิคการเรียนรู้แบบกลุ่ม พบว่ามีค่าความแม่นยำสูงถึง 92.38% ซึ่งวิเคราะห์ได้ว่า อัลกอริทึมโครงข่ายประสาทเทียมมีค่าความแม่นยำมากที่สุด (93.43%) มากกว่าการเรียนรู้แบบกลุ่ม ตามภาพประกอบ 65



ภาพประกอบ 65 ค่าความแม่นยำแบบจำลองพยากรณ์ประเภทกองทุน

อภิปรายผลการวิจัย

ผู้วิจัยทำการวิเคราะห์ผลลัพธ์ที่น่าสนใจ สำหรับการอภิปรายผลการวิจัยได้ ดังนี้

การจัดการข้อมูลและค่าความแม่นยำ

ในการแบ่งกลุ่มข้อมูลในรูปแบบเฉลี่ยช่วงข้อมูล และการแบ่งกลุ่มเคมีน แสดงให้เห็นว่า

1. การแบ่งกลุ่มข้อมูลในรูปแบบเฉลี่ยช่วงข้อมูล จากผลิตภัณฑ์กองทุนประเภท LR และ MF ข้อมูลส่วนใหญ่กว่า 97.46% และ 75.41% ถูกจัดกลุ่มอยู่ในกลุ่มที่ 1 (1-200,000) สอดคล้องกับค่าความแม่นยำที่ได้จากทุกอัลกอริทึมเฉลี่ยอยู่ที่ 97.42% และ 74.96% ตามลำดับ ซึ่งมีค่าความแม่นยำมากกว่า 70.00% แสดงให้เห็นว่า การพยากรณ์ผลิตภัณฑ์กองทุนประเภท LR และ MF มีค่าความแม่นยำตามน้ำหนักของตัวแปรพยากรณ์ ซึ่งสามารถนำไปใช้ในกรณีพยากรณ์การซื้อผลิตภัณฑ์กองทุน แต่ยังไม่เหมาะในการนำไปใช้พยากรณ์ระดับในกองทุนประเภท LR และ MF ในขณะที่การแบ่งกลุ่มข้อมูลในรูปแบบเฉลี่ยช่วงข้อมูล จากผลิตภัณฑ์กองทุนประเภท MM และ TF ที่มีข้อมูลเฉลี่ยในทุกกลุ่ม จะมีค่าความแม่นยำที่ได้จากทุกอัลกอริทึมเฉลี่ยอยู่ที่ 45.23% และ 31.34% ตามลำดับ ซึ่งมีค่าความแม่นยำน้อยกว่า 50.00% แสดงให้เห็นว่า การแบ่งกลุ่มรูปแบบเฉลี่ยช่วงข้อมูลในการพยากรณ์หรือแนะนำผลิตภัณฑ์กองทุนในงานวิจัยครั้งนี้ ยังไม่เหมาะสมในการนำไปใช้ เนื่องจากปริมาณของข้อมูลและน้ำหนักของข้อมูลพยากรณ์เมื่อมีการกระจายน้ำหนักจะทำให้ค่าประสิทธิภาพน้อยลง

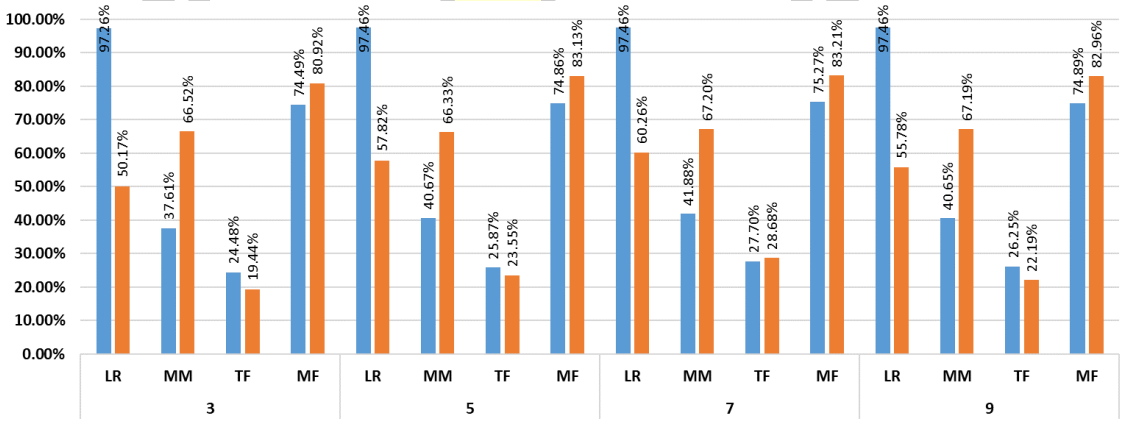
2. การแบ่งกลุ่มข้อมูลเคมีน หากเปรียบเทียบกับการแบ่งกลุ่มข้อมูลในรูปแบบเฉลี่ยช่วงข้อมูล พบว่า เป็นการแบ่งกลุ่มข้อมูลที่อธิบายตัวแปรต่างๆ ให้อยู่ในตัวแปรเดียว ซึ่งอาจต้องใช้การให้ความหมายกลุ่มจากผู้เชี่ยวชาญด้านผลิตภัณฑ์กองทุน แต่ให้ผลดีในด้านการลดจำนวนตัวแปรในการประมวลผลข้อมูล และมีการกระจายจุดเด่นของแต่ละกลุ่มได้อย่างชัดเจน ยกตัวอย่างเช่น ผลิตภัณฑ์กองทุนประเภท LR ในกลุ่ม Cluster_1 เป็นกลุ่มลูกค้าที่มีจำนวนมากที่สุดแต่มีปริมาณการซื้อผลิตภัณฑ์และการออมในรูปแบบต่างๆ น้อย เมื่อให้รูปแบบพยากรณ์ทำการพยากรณ์ผลแนะนำได้ว่า Cluster_1 ผู้แนะนำผลิตภัณฑ์กองทุนอาจทำการแนะนำผลิตภัณฑ์อื่น หรือผลิตภัณฑ์กองทุนประเภท MM ในกลุ่ม Cluster_3 เป็นกลุ่มลูกค้าที่มีปริมาณการซื้อผลิตภัณฑ์กองทุนสูง และมีปริมาณการออมปานกลาง ผู้แนะนำผลิตภัณฑ์กองทุน อาจแนะนำกองทุนประเภท MM หรือเป็นการให้เกณฑ์ตัวเลือกที่เหมาะสมในการแนะนำมากที่สุด เป็นต้น

3. การแบ่งกลุ่มข้อมูเคมีน เมื่อพิจารณาถึงค่าความแม่นยำ อธิบายได้ว่า เมื่อใช้เทคนิคแบบรวมกลุ่มมีค่าเฉลี่ยความแม่นยำของทุกผลิตภัณฑ์กองทุน มากกว่าการแบ่งกลุ่มข้อมูลแบบเฉลี่ยช่วงข้อมูลอยู่ที่ 71.09% ต่อ 62.08% และหากเฉลี่ยค่าความแม่นยำในทุกอัลกอริทึมในการวิจัยแล้ว การแบ่งกลุ่มข้อมูลเคมีน มีค่าความแม่นยำมากกว่าการแบ่งกลุ่มข้อมูลแบบเฉลี่ยช่วงข้อมูลอยู่ที่ 69.21% ต่อ 62.24% แสดงได้ว่า การแบ่งกลุ่มข้อมูลโดยวิธีการเคมีน ให้ผลลัพธ์โดยรวมที่ดีกว่าและมีการอธิบายกลุ่มข้อมูลที่ครอบคลุมมากกว่าจากชุดข้อมูลที่นำมาพัฒนารูปแบบพยากรณ์

การเปรียบเทียบค่าพารามิเตอร์และอัลกอริทึม

จากการเปรียบเทียบค่าพารามิเตอร์ก่อนการพัฒนารูปแบบพยากรณ์ มีประเด็นที่น่าสนใจ ดังนี้

1. อัลกอริทึมเคเนียร์เรสเนเบอร์จากการตั้งค่าพารามิเตอร์ K หรือจำนวนจุดข้อมูลที่นำมาประมวลผล พบว่าค่า K ที่แตกต่างกัน ไม่ได้ทำให้ประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์แตกต่างกันมากจากชุดข้อมูลที่ใช้ในงานวิจัยครั้งนี้ และเมื่อเปรียบเทียบการจัดการข้อมูลโดยอิงจากค่าพารามิเตอร์ที่แตกต่างกัน พบว่ามีผลต่อประสิทธิภาพความแม่นยำที่แตกต่างกันมาก ยกตัวอย่างเช่น ผลิตภัณฑ์กองทุนประเภท LR โดยพารามิเตอร์ K=3 ในรูปแบบเฉลี่ยช่วงข้อมูลกับเคมีน มีค่าความแม่นยำอยู่ที่ 97.26% ต่อ 50.17% เป็นต้น แต่ในขณะที่ผลิตภัณฑ์กองทุนประเภท TF โดยพารามิเตอร์ K=3 มีค่าความแม่นยำอยู่ที่ 24.48% ต่อ 19.44% จากการกระจายตัวของข้อมูลทั้ง 2 รูปแบบ ส่งผลให้มีความเปลี่ยนแปลงของรูปแบบพยากรณ์ที่ไม่แตกต่างกันมาก ตามภาพประกอบ 66

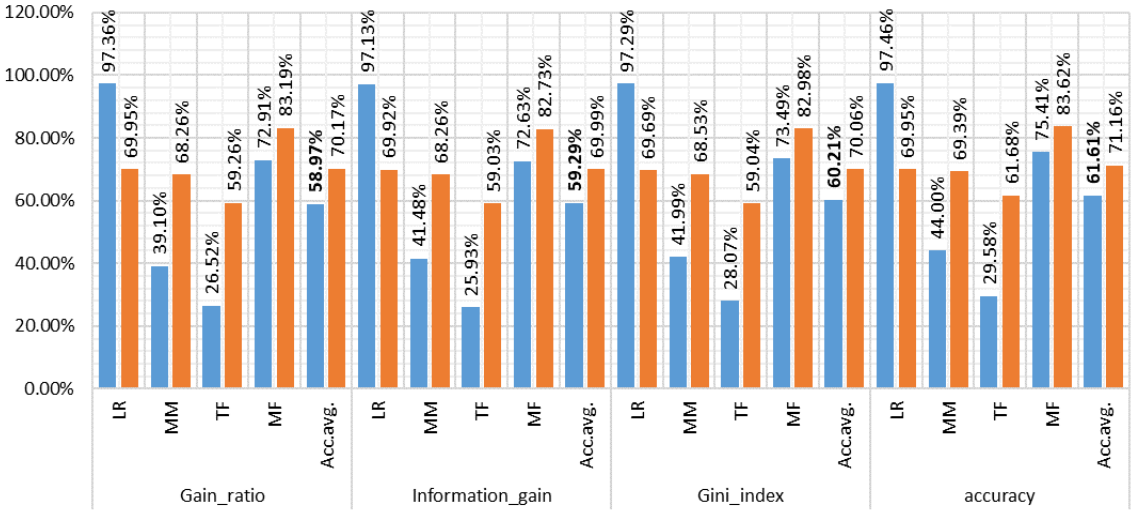


ภาพประกอบ 66 ค่าความแม่นยำต่อการจัดการข้อมูลและค่าพารามิเตอร์ของเคเนียร์เรสเนเบอร์

2. อัลกอริทึมนาอีฟเบย์ไม่มีการตั้งค่าพารามิเตอร์เพิ่มเติม ซึ่งในตัวอัลกอริทึมมีประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์ที่เทียบเท่าอัลกอริทึมอื่นที่ใช้ในงานวิจัย แต่เมื่อปรียบเทียบค่าความแม่นยำจากรูปแบบการจัดการข้อมูล พบว่าผลิตภัณฑ์กองทุนประเภท LR มีค่าความแม่นยำลดลง ในขณะที่ผลิตภัณฑ์กองทุนอื่นๆ มีประสิทธิภาพเพิ่มขึ้น

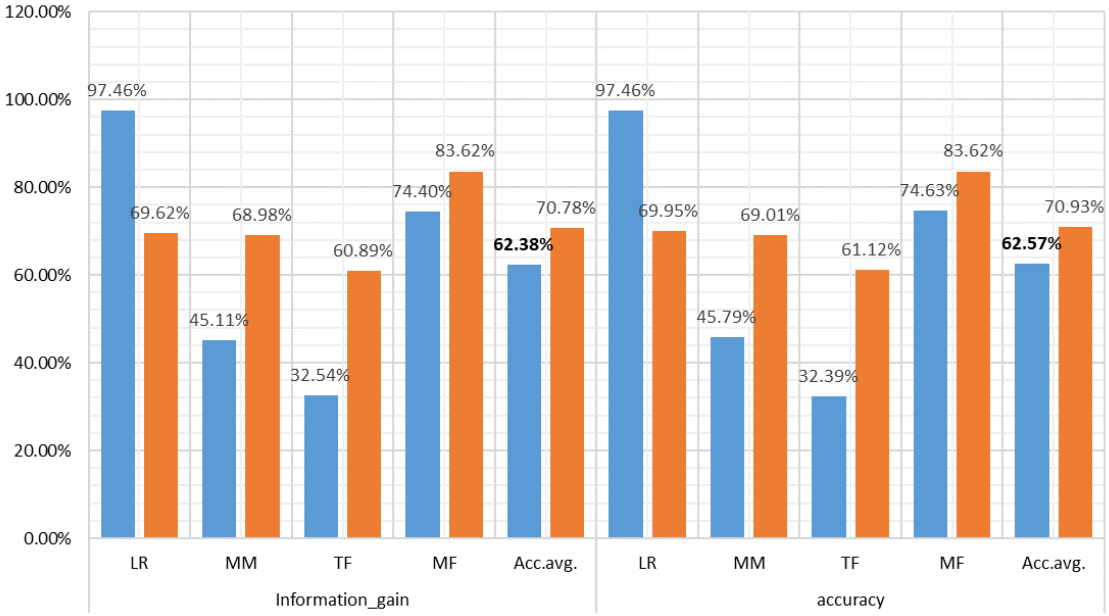
3. อัลกอริทึมต้นไม้ตัดสินใจจากการตั้งค่าพารามิเตอร์ Criterion พบว่าค่า Criterion ที่แตกต่างกัน ไม่ได้ทำให้ประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์แตกต่างกันมากจากชุดข้อมูลที่

ใช้ในงานวิจัยครั้งนี้ และเมื่อเปรียบเทียบการจัดการข้อมูลโดยอิงจากค่าพารามิเตอร์ที่แตกต่างกัน พบว่ามีผลต่อประสิทธิภาพความแม่นยำที่แตกต่างกันมาก ยกตัวอย่างเช่น ผลิตภัณฑ์กองทุนประเภท LR โดยพารามิเตอร์ Gain_Ratio ในรูปแบบเฉลี่ยช่วงข้อมูลกับเคมีน มีค่าความแม่นยำอยู่ที่ 97.36% ต่อ 69.95% เป็นต้น แต่ในขณะที่ผลิตภัณฑ์กองทุนประเภทอื่นในรูปแบบที่ 2 มีความแม่นยำเพิ่มขึ้นและมากกว่ารูปแบบที่ 1 ตามภาพประกอบ 67



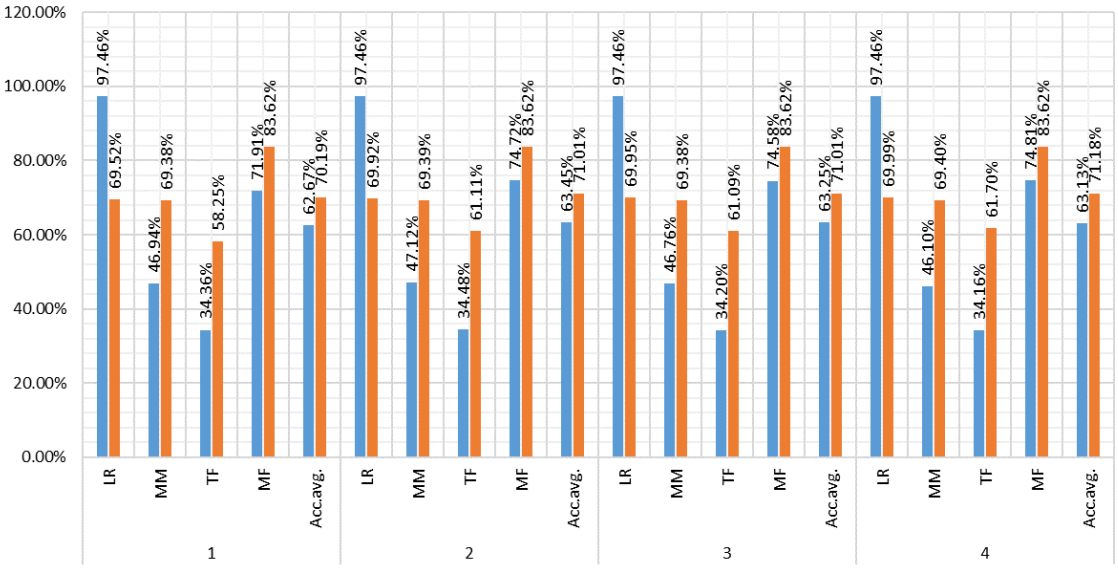
ภาพประกอบ 67 ค่าความแม่นยำต่อการจัดการข้อมูลและค่าพารามิเตอร์ของต้นไม้ตัดสินใจ

4. อัลกอริทึมกฎการอุปนัย จากการตั้งค่าพารามิเตอร์ Criterion พบว่าค่า Criterion ที่แตกต่างกัน ไม่ได้ทำให้ประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์แตกต่างกันมากจากชุดข้อมูลที่ใช้ในงานวิจัยครั้งนี้ และเมื่อเปรียบเทียบการจัดการข้อมูลโดยอิงจากค่าพารามิเตอร์ที่แตกต่างกัน พบว่ามีผลต่อประสิทธิภาพความแม่นยำที่แตกต่างกันมาก ยกตัวอย่างเช่น ผลิตภัณฑ์กองทุนประเภท LR โดยพารามิเตอร์ Information_Gain ในรูปแบบเฉลี่ยช่วงข้อมูลกับเคมีน มีค่าความแม่นยำอยู่ที่ 97.46% ต่อ 69.62% เป็นต้น แต่ในขณะที่ผลิตภัณฑ์กองทุนประเภทอื่นในรูปแบบที่ 2 มีความแม่นยำเพิ่มขึ้นและมากกว่ารูปแบบที่ 1 ตามภาพประกอบ 68



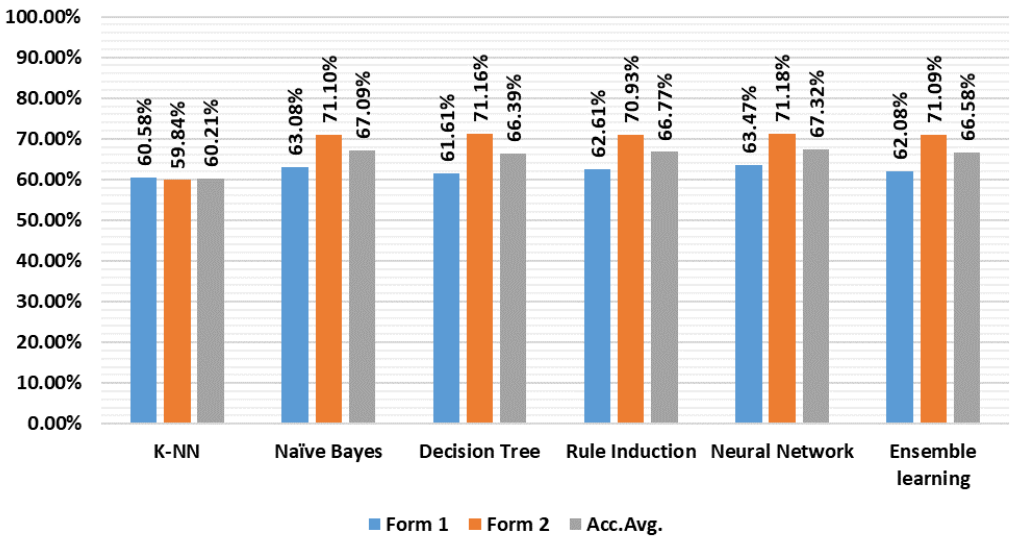
ภาพประกอบ 68 ค่าความแม่นยำต่อการจัดการข้อมูลและค่าพารามิเตอร์ของกฎการอุปนัย

5. อัลกอริทึมโครงข่ายประสาทเทียมจากการตั้งค่าพารามิเตอร์ Criterion พบว่าค่า Criterion ที่แตกต่างกัน ไม่ได้ทำให้ประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์แตกต่างกันมากจากชุดข้อมูลที่ใช้ในงานวิจัยครั้งนี้ และเมื่อเปรียบเทียบการจัดการข้อมูลโดยอิงจากค่าพารามิเตอร์ที่แตกต่างกัน พบว่ามีผลต่อประสิทธิภาพความแม่นยำที่แตกต่างกันมาก ตามภาพประกอบ 69



ภาพประกอบ 69 ค่าความแม่นยำต่อการจัดการข้อมูลและค่าพารามิเตอร์ของโครงข่ายประสาทเทียม

6. เมื่อพิจารณาค่าความแม่นยำเฉลี่ยของทุกผลิตภัณฑ์กองทุนและทุกรูปแบบการจัดการข้อมูล อัลกอริทึมโครงข่ายประสาทเทียมเป็นอัลกอริทึมที่มีค่าความแม่นยำมากที่สุด (67.32%) และถัดมาตามลำดับได้แก่ นาอีฟเบย์ (67.09%) กฎการอุปนัย (66.77%) ต้นไม้ตัดสินใจ (66.39%) และเคเนียร์เรสเนเบอร์ (60.21%) และเมื่อพิจารณาตามรูปแบบจัดการขอมูลแบบเฉลี่ยช่วงข้อมูลตามลำดับได้แก่ โครงข่ายประสาทเทียม (63.47%) นาอีฟเบย์ (63.08%) กฎการอุปนัย (62.61%) ต้นไม้ตัดสินใจ (61.61%) และเคเนียร์เรสเนเบอร์ (60.58%) สุดท้ายในรูปแบบการแบ่งกลุ่มข้อมูลเคมีน ตามลำดับได้แก่ โครงข่ายประสาทเทียม (71.18%) นาอีฟเบย์ (63.08%) ต้นไม้ตัดสินใจ (71.16%) กฎการอุปนัย (70.93%) และเคเนียร์เรสเนเบอร์ (59.84%) ตามภาพประกอบ 5.5 แสดงให้เห็นว่าอัลกอริทึมเคเนียร์เรสเนเบอร์ยังไม่เหมาะกับการใช้พัฒนารูปแบบพยากรณ์จากชุดข้อมูลในงานวิจัยนี้ และการใช้เทคนิคการรวมกลุ่ม แม้เป็นเทคนิควิธีการที่มีความแม่นยำจากการโหวตผลการพยากรณ์ แต่ก็มีความระเอียดและเวลาในการประมวลผลของแต่ละอัลกอริทึมที่นาน เมื่อเทียบกับปริมาณข้อมูลที่ใช้ในงานวิจัย ซึ่งจะเห็นได้ว่า การใช้อัลกอริทึมโครงข่ายประสาทเทียมหรืออัลกอริทึมอื่นๆ นอกจากเคเนียร์เรสเนเบอร์แล้ว มีประสิทธิภาพความแม่นยำของรูปแบบพยากรณ์ใกล้เคียงกันอยู่แล้ว



ภาพประกอบ 70 เปรียบเทียบค่าความแม่นยำเฉลี่ยของการจัดการข้อมูลในแต่ละอัลกอริทึม

7. จากผลลัพธ์ค่าความแม่นยำเฉลี่ยของแบบจำลองพยากรณ์ภายในงานวิจัย แบ่งออกเป็น แบบจำลองพยากรณ์ประเภทกองทุน 92.78% และแบบจำลองพยากรณ์ระดับกองทุนใน

รูปแบบที่ 1 62.24% และในรูปแบบที่ 2 69.21% อธิบายได้ว่า เมื่อเปรียบเทียบค่าความแม่นยำของแบบจำลองพยากรณ์ประเภทกองทุน มีค่าความแม่นยำมากกว่าในหลายๆ งานวิจัย แต่น้อยกว่างานวิจัยเกี่ยวกับ การคาดคะเนการลาออกของลูกค้าของธนาคารพาณิชย์ อยู่ที่ 6.17% ที่ใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมทชีน แต่หากเปรียบเทียบค่าเฉลี่ยความเที่ยงและค่าเฉลี่ยความระลึกพบว่าให้ค่าที่มากกว่า หากเปรียบเทียบด้านอัลกอริทึมพบว่า การเรียนกลุ่มแบบโหวตของงานวิจัยจำแนกและการพยากรณ์โรคเบาหวานให้ค่าความแม่นยำที่น้อยที่สุด แต่ไม่ได้มีเรื่องความสนใจใกล้เคียงกัน และการเรียนรู้กลุ่มแบบ XGBoost ในการจำแนกประเภทการลงทุนในกองทุนรวมเมื่อเปรียบเทียบแล้ว งานวิจัยนี้ให้ผลลัพธ์ที่ดีกว่า และเมื่อเทียบด้านการจัดการข้อมูลของงานวิจัยการออกแบบระบบช่วยตัดสินใจเลือกซื้อผลิตภัณฑ์กองทุนที่ใช้วิธีเฉลี่ยช่วงข้อมูลให้ค่าความแม่นยำที่มากกว่าผลแบบจำลองในรูปแบบที่ 2 ที่ใช้วิธีการนำเสนอผ่านการแบ่งกลุ่มด้วยเคมีน เท่ากับ 17.70% แต่เมื่อพิจารณาผลลัพธ์การทำนายแล้วไม่ได้ลงลึกถึงระดับของกอนทุน เป็นเพียงประเภทกองทุนเช่นเดียวกันกับแบบจำลองพยากรณ์ประเภทกองทุนของงานวิจัยนี้ ซึ่งให้ค่าความแม่นยำมากกว่า 5.87%

ข้อเสนอแนะ

ผู้วิจัยมีข้อเสนอแนะภายในการทำวิจัย ดังนี้

1. ข้อมูลประชากรที่นำมาใช้ในงานวิจัย ควรเลือกประชากรที่ทำการสมัครซื้อผลิตภัณฑ์กองทุนอย่างรอบครอบ หากใช้ข้อมูลที่เก่าเกินไปอาจส่งผลกระทบต่อความเป็นจริงของกลุ่มประชากรนั้นๆ ได้ อีกทั้งในด้านความหลากหลายของระดับในแต่ละผลิตภัณฑ์กองทุน ควรมีการกระจายตัวแปรพยากรณ์ที่ใกล้เคียงกัน หรือมีปริมาณที่มากกว่าที่ใช้ในงานวิจัยครั้งนี้ เพื่อเพิ่มความแม่นยำในการพยากรณ์ระดับของผลิตภัณฑ์กองทุนในระดับอื่นๆ มากกว่าพยากรณ์ผลในระดับที่มีตัวแปรพยากรณ์มากที่สุด

2. ข้อเสนอแนะด้านประเภทของผลิตภัณฑ์กองทุน จากที่ผู้วิจัยได้ทำการวิจัยไปแล้วนั้นเป็นการรวมกลุ่มของกองทุนทั้ง 4 ประเภท โดยหากได้ทำการวิเคราะห์ในแต่ละผลิตภัณฑ์ที่มีอยู่ในตลาดการเงิน อาจเป็นอีกหนึ่งรูปแบบที่น่าสนใจและมีความหลากหลายเพิ่มขึ้น

3. ในการใช้เทคนิคแบบรวมกลุ่ม อาจเหมาะในการเพิ่มประสิทธิภาพของรูปแบบพยากรณ์ได้ในระดับหนึ่ง แต่หากทดลองกับอัลกอริทึมแล้วพบว่ามีประสิทธิภาพเทียบเท่า ใกล้เคียงหรือดีกว่าอยู่แล้ว อาจไม่จำเป็นต้องเพิ่มขั้นตอนในการสร้างรูปแบบพยากรณ์ แต่หากเมื่อปริมาณ

ข้อมูลที่มากขึ้นหรือการทำข้อมูลขนาดใหญ่ (Big Data) อาจมีความจำเป็นต่อการพัฒนาอัลกอริทึมหรือการนำเทคนิคแบบรวมกลุ่มมาใช้ได้ในอนาคต

4. ในงานวิจัย ได้คัดเลือกอัลกอริทึมโดยให้ความสนใจในการเปรียบเทียบประเภทของอัลกอริทึม มีข้อเสนอแนะในการเปรียบเทียบอัลกอริทึมประเภทเดียวกันแต่คนละเทคนิควิธีการ อย่างเช่น อัลกอริทึมในกลุ่ม Tree อย่าง ต้นไม้ตัดสินใจ, ป่าแบบสุ่ม และ Gradient Boosted Trees (GBT) เป็นต้น ซึ่งมีความหลากหลายในการตั้งค่าพารามิเตอร์ที่หลากหลาย

พหุน ปณ ฑิโต ชีเว

บรรณานุกรม

บรรณานุกรม

ชิตพงษ์ กิตตินราดร. (2563). *Neural Network Algorithm*. [ออนไลน์]. ได้จาก: https://guopai.Github.Io/Ml-Blog14.Html. [สืบค้นเมื่อวันที่ 3 สิงหาคม 2564].

ณรงค์ จารขจรกุล. (2541). *ปัจจัยกำหนดผลตอบแทนจากการลงทุนในหลักทรัพย์ของประเทศไทย (วิทยานิพนธ์มหาบัณฑิต)*. มหาวิทยาลัยรามคำแหง.

ตลาดหลักทรัพย์แห่งประเทศไทย. (2558). *ประเภทของกองทุนรวม*. [ออนไลน์]. ได้จาก: https://www.set.or.th/Education/Th/Begin/Mutualfund_content02.pdf. [สืบค้นเมื่อวันที่ 4 สิงหาคม 2564].

ตลาดหลักทรัพย์แห่งประเทศไทย. (2563). *กองทุนรวม*.[ออนไลน์]. ได้จาก https://www.setinvestnow.com/Th/Mutualfund/Types-of-Mutual-Funds. [สืบค้นเมื่อวันที่ 6 สิงหาคม 2564].

ตลาดหลักทรัพย์แห่งประเทศไทย. (2564). *DCA กับ กองทุนรวมตลาดเงิน*. ได้จาก: https://www.setinvestnow.com/Th/Knowledge/Article/50-Investing-in-Money-Market-Funds-with-Dca-Strategy. [สืบค้นเมื่อวันที่ 6 สิงหาคม 2564].

นพมาศ ปักเข็ม. (2015). การเรียนรู้แบบเบย์. *เอกสารประกอบการสอนวิชาการทำเหมืองข้อมูล สาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ [ออนไลน์]. ได้จาก: https://mis.csit.Sci.Tsu.Ac.Th/Noppamas/Download/DataMining/DataMining Ch7V1.Pdf. [สืบค้นเมื่อวันที่ 3 สิงหาคม 2564].*

ฝนทิพย์ คูณแก้ว. (2555). *การสังเคราะห์โมเดลเพื่อการจำแนกตามข้อกำหนดของผู้ใช้* [วิทยานิพนธ์มหาบัณฑิต]. มหาวิทยาลัยสุรนารี.

วิภาวรรณ บัวทอง. (2557). *Classification: K-Nearest Neighbors (K-NN).* [ออนไลน์]. ได้จาก: https://wipawanblog.Files.Wordpress.Com/2014/06/Chapter-6-k-Nearest-Neighbors.pdf. [สืบค้นเมื่อวันที่ 3 สิงหาคม 2564].

ศูนย์ส่งเสริมการพัฒนาความรู้ตลาดทุน. (2553). *การวางแผนการลงทุน.* [ออนไลน์]. *กรุงเทพฯ: ศูนย์ส่งเสริมการพัฒนาความรู้ตลาดทุน สถาบันกองทุนเพื่อพัฒนาตลาดทุน ตลาดหลักทรัพย์แห่งประเทศไทย*. ได้จาก : https://www.Ookbee.com. [สืบค้นเมื่อวันที่ 6 สิงหาคม 2564].

สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์. (2564). รายชื่อบริษัทที่ได้รับใบอนุญาต การจัดการกองทุนรวม ประเภทบริษัทหลักทรัพย์. *[ออนไลน์]. ได้จาก: https://market.Sec.or.Th/Public/Orap/COMPANYPROFILE03.Aspx?Lang=th&licno=5&grptype=lcstype= [สืบค้นเมื่อวันที่ 4 สิงหาคม 2564].*

อรรถพร สะพะพันธุ์. (2559). *การออกแบบระบบช่วยตัดสินใจเลือกซื้ออผลิตภัณฑ์กองทุน กรณีศึกษาบริษัทหลักทรัพย์จัดการกองทุนแห่งหนึ่ง* [วิทยานิพนธ์มหาบัณฑิต]. สถาบันเทคโนโลยีไทย-ญี่ปุ่น.

อำภา สาระศิริ. (2559). *เทคนิคการเรียนรู้พื้นฐานโครงข่ายประสาทเทียม.* [ออนไลน์]. ได้จาก: https://www.mut.Ac.Th/Research-Detail-92. [สืบค้นเมื่อวันที่ 3 สิงหาคม 2564].

Agarwal, S. (2013). Data mining: Data mining concepts and techniques. *In 2013 International Conference on Machine Intelligence and Research Advancement*, 203–207.

Ahmad, A. , & D. L. (2007). A k-mean clustering algorithm for mixed numeric and categorical data. *Data & Knowledge Engineering*, *63*(2), 503–527.

Anthony, M., & Bartlett, P. L. (2009). Neural network learning. In *Theoretical foundations*. cambridge university press.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Chatterjee, S. (2019). Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. *Decision Support Systems*, *119*, 14–22.

Cohen, W. (1995). *Fast effective rule induction. In Machine learning proceedings 1995.*

Dietterich, T. G. (2000). Ensemble methods in machine learning. *In International Workshop on Multiple Classifier Systems* , 1–15.

Freitas, A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (10th ed., Vol. 1). Springer series in statistics.

Haowcharoen, K. (2016). *Relationship between earning quality and dividend payment of the listed companies in property and construction industry group in the stock exchange of Thailand* [Master's thesis]. Rajamangala University of Technology Thanyaburi.

He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. *Procedia Computer Science*, *31*, 423–430.

He, C., Zhu, B., Zhang, M., Zhuang, Y., He, X., & Du, D. (2015). Customers' risk type prediction based on analog complexing. *Procedia Computer Science*, *55*, 939–943.

Hernández-Nieves, E., del Canto, Á. B., Chamoso-Santos, P., de la Prieta-Pintado, F., & Corchado-Rodríguez, J. M. (2020). A Machine Learning Platform for Stock Investment Recommendation Systems. In *In International Symposium on Distributed Computing and Artificial Intelligence* (pp. 303–313). Cham.

Hirunratsamee, T. (2015). *Financial accounting* (4th ed). Vittayapat.

Hulett, C., Hall, A., & Qu, G. (2012).). Dynamic selection of k nearest neighbors in instance-based learning. *In 2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*, 85–92.

Jaiswal, D. P., Kumar, S., & Mukherjee, P. (2020). Customer transaction prediction system. *Procedia Computer Science*, *168*, 49–56.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90–95.

Kong, K., Liu, R., Zhang, Y., & Chen, Y. (2020). Predicting Liquidity Ratio of Mutual Funds via Ensemble Learning. *In 2020 IEEE International Conference on Big Data (Big Data)*, 5441–5450.

Kotsiantis, S., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3–24.

Li, J., Wang, X., & Xu, S. (2010). Prediction of customer classification based on rough set theory. *Procedia Engineering*, *7*, 366–370.

Manish K. (2012). *Ensemble Techniques for Weak Learners*. Retrieved from Http://Manish2020.Blogspot.Com/2012/12/Ensemble-of-Weak-Learners.Html.

Markowitz, H. (2015). An Encyclopedia of Modern Money Management. *Portfolio Theory*, 321.

Marr, B. (2020). *Big Data & Analytics Case Studies*. Retrieved 12 July 2021 from Https://Bernardmarr.Com/Default.Asp?ContentID=739.

Mashlakov, A., Kuronen, T., Lensu, L., Kaarna, A., & Honkapuro, S. (2021). Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Applied Energy*, *285*, 116–405.

Murray, P. W., Agard, B., & Barajas, M. A. (2015). Forecasting supply chain demand by clustering customers. *IFAC-PapersOnLine*, *48*(3), 1834–1839.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*.

Reichstein, C., & Härting, R. C. (2018). Potentials of changing customer needs in a digital world–a conceptual model and recommendations for action in tourism. *Procedia Computer Science*, *126*, 1484–1494.

Sachdev, S. H. (2020). *Choosing number of Hidden Layers and number of hidden neurons in Neural Networks*. Retrieved from https://www.Linkedin.com/Pulse/Choosing-Number-Hidden-Layers-Neurons-Neural-Networks-Sachdev.

SangSoi, C. (2015). *The relationship between financial ratios and stock price of companies listed in the stock market of Thailand: A case study of service sector* [Master Degree's Thesis]. Bangkok University.

Sungkaew, J. (2015). *Investment* (4th ed.). Thammasat University.

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2020). Restaurants store management based on demand forecasting. *Procedia CIRP*, *88*, 580–583.

Tao, T., Yan, K., & Yang, S. (2019). Classification of mutual fund investment types with advanced machine learning models. *In 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)* , 84–89.

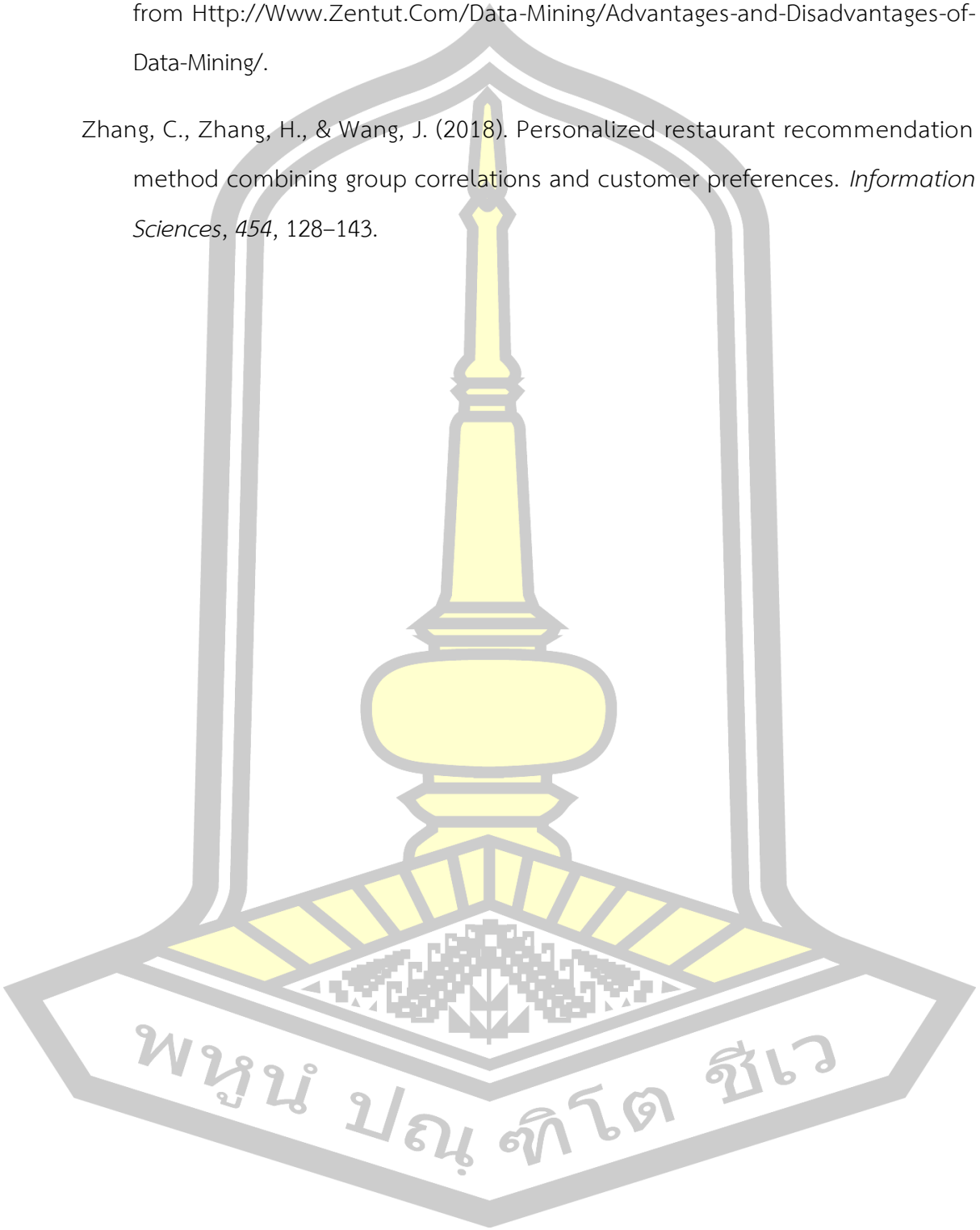
Thongchai Kaewkiriya. (2017). Design of e-learning content for student guidance based on ILS and 4MAT theory. *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, 241–245. https://doi.org/10.1109/CSCITA.2017.8066561

Xia, J., Zhang, S., Cai, G., Li, L., Pan, Q., Yan, J., & Ning, G. (2017). Adjusted weight voting algorithm for random forests in handling missing values. *Pattern Recognition*, *69*, 52–60.

Yong, X., Ying-xin, P., & Jian-bin, X. (2018). The Improvement of Personalized Recommendation System Based on Sparse Data of Financial Products. *World Journal of Research and Review*, *7*(2), 262–601.

Zentut website. (2016). *Advantages and Disadvantages of Data Mining*. Retrieved from Http://Www.Zentut.Com/Data-Mining/Advantages-and-Disadvantages-of-Data-Mining/.

Zhang, C., Zhang, H., & Wang, J. (2018). Personalized restaurant recommendation method combining group correlations and customer preferences. *Information Sciences*, *454*, 128–143.



ประวัติผู้เขียน

ชื่อ	นายธงชัย แก้วกิริยา	
วันเกิด	วันที่ 10 มกราคม 2521	
สถานที่เกิด	ตำบลอินทร์บุรี อำเภออินทร์บุรี จังหวัดสิงห์บุรี	
สถานที่อยู่ปัจจุบัน	เลขที่ 222/287 หมู่บ้านสราญสิริ ถนนปัญญาอินทรา แขวงบางชัน เขตคลองสามวา กรุงเทพมหานคร ไปรษณีย์ 10510	
ตำแหน่งหน้าที่การงาน	ที่ปรึกษาด้านเทคโนโลยีสารสนเทศ	
สถานที่ทำงานปัจจุบัน	บริษัท ซิกม่า เทค จำกัด	
ประวัติการศึกษา	พ.ศ. 2543	ปริญญาตรี ครุศาสตร์อุตสาหกรรมบัณฑิต (ค.อ.บ.) สาขาเทคโนโลยีคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
	พ.ศ. 2549	ปริญญาตรี วิศวกรรมศาสตรบัณฑิต (วศ.บ.) สาขาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม สถาบันเทคโนโลยีปทุมวัน
	พ.ศ. 2548	ปริญญาโท วิศวกรรมศาสตรมหาบัณฑิต (วศ.ม.) สาขาวิศวกรรมไฟฟ้าและสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
	พ.ศ. 2562	ปริญญาโท บริหารธุรกิจมหาบัณฑิต (บธ.ม.) สาขาการจัดการ มหาวิทยาลัยรามคำแหง
	พ.ศ. 2557	ปริญญาเอก Doctor of Engineering (D.Eng) สาขา Electrical Engineering and Information Science (Computer Science and Intelligent System) Osaka Prefecture University, Japan
	พ.ศ. 2559	ปริญญาเอก ปรัชญาดุษฎีบัณฑิต (ปร.ด.) สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
	พ.ศ. 2566	ปริญญาเอก ปรัชญาดุษฎีบัณฑิต (ปร.ด.)

สาขาวิชาบริหารธุรกิจและนวัตกรรมดิจิทัล

มหาวิทยาลัยมหาสารคาม