การเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมด้วยเทคนิคการเรียนรู้ของเครื่อง

ธันย์ชนก ขำประดิษฐ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตศึกษา สาขาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2567

COMPARING THE PERFORMANCE OF DEBT REPAYMENT ABILITY FOR EXISTING BORROWERS USING MACHINE LEARNING TECHNIQUES

Thanchanok Khampradit

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology
Graduate Studies
Thai-Nichi Institute of Technology
Academic Year 2024

หัวข้อวิทยานิพนธ์	การเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมด้วยเทคนิคการเรียนรู้ของเครื่อง
โดย	ธันย์ชนก ขำประดิษฐ์
สาขาวิชา	เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์

บัณฑิตศึกษา สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....รองอธิการบดีฝ่ายวิชาการ
(รองศาสตราจารย์ ดร.วรากร ศรีเชวงทรัพย์)
วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.วรภัทร ไพรีเกรง)

..... กรรมการ
(ดร.ภูวดล ศิริกองธรรม)

..... กรรมการ
(ดร.อภิชญา นิ้มคุ้มภัย)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์
(ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์)

ธันย์ชนก ขำประดิษฐ์ : การเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมด้วยเทคนิคการเรียนรู้ของเครื่อง. อาจารย์ที่ปรึกษา : ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์, 65 หน้า.

งานวิจัยนี้มีวัตถุประสงค์สร้างแบบจำลองข้อมูลการในการวิเคราะห์ความสามารถในการชำระหนี้คืนโดยใช้เทคนิคการเรียนรู้ของเครื่อง และเพื่อเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมาะสมในการจำแนกความสามารถในการชำระหนี้คืนจากข้อมูลลูกค้าผู้กู้รายเดิม เพื่อช่วยธนาคารประกอบการตัดสินใจในการอนุมัติสินเชื่อ โดยการนำข้อมูลในอดีตมาวิเคราะห์เพื่อสะท้อนถึงความสามารถในการชำระหนี้ในอนาคต และทำให้ธนาคารมีเครื่องมือที่ง่ายต่อการอนุมัติในการปล่อยสินเชื่อของลูกค้าที่เคยมีประวัติการขอสินเชื่อจากธนาคาร จากการศึกษานี้ใช้ข้อมูลประเมินศักยภาพลูกหนี้ของธนาคารแห่งหนึ่ง จำนวน 4,334 ราย โดยใช้เทคนิคการเรียนรู้ของเครื่องเพื่อเปรียบเทียบความสามารถในการชำระหนี้และเปรียบเทียบโมเดลที่มีผลต่อความสามารถในการชำระหนี้ โดยเทคนิคที่เลือกใช้ ได้แก่ Logistic Regression Neural Network Decision Tree และ Random Forest ด้วยโปรแกรม Visual Studio Code จากผลการศึกษาพบว่า โมเดล Random Forest มีความแม่นยำสูงและแปรปรวนต่ำ วัดประสิทธิภาพค่าเฉลี่ยความถูกต้องแม่นยำ (Accuracy) คิดเป็น 95.98% และค่าเฉลี่ยความคลาดเคลื่อนในการทำนาย (RMSE) 20.99% นอกจากนี้ต้นเงินค้างในบิล (The Principal Remains on the Bill) มีผลต่อความสามารถในการชำระหนี้คืนมากที่สุด โมเดล Random Forest จึงมีความเหมาะสมที่จะนำไปใช้ในการพิจารณาอนุมัติสินเชื่อ โดยวัดจากค่าประสิทธิภาพความแม่นยำสูงสุดจากการสร้างโมเดลที่ทำนายจากพฤติกรรมการชำระหนี้ของผู้กู้รายเดิมในอดีต จะช่วยส่งผลให้การปล่อยอนุมัติสินเชื่อของธนาคารมีประสิทธิภาพและมีความแม่นยำมากยิ่งขึ้น

บัณฑิตศึกษา
สาขาวิชา เทคโนโลยีสารสนเทศ
ปีการศึกษา 2567

ลายมือชื่อนักศึกษา .....
ลายมือชื่ออาจารย์ที่ปรึกษา .....

THANCHANOK KHAMPRADIT : COMPARING THE PERFORMANCE OF DEBT REPAYMENT ABILITY FOR EXISTING BORREWERS USING MACHINE LEARNING TECHNIQUES. ADVISOR : ACTING SUB LT. DR. PICHITCHAI KAMIN, 65 PP.

This research aims to develop data models to analyze debt repayment ability using machine learning techniques and to compare the efficiency of suitable techniques in classifying repayment ability based on previous borrower data. The goal is to assist banks in their decision-making process for loan approvals by analyzing historical data to reflect future repayment capacity. It also provides financial institutions with a tool that simplifies loan approval processes for customers with a previous loan history. The study utilized data from 4,334 borrowers from a bank to evaluate debtor potential. Machine learning algorithms were applied to compare repayment abilities and assess models that influence repayment. The selected techniques include Logistic Regression Neural Network Decision Tree and Random Forest using the Visual Studio Code program.

The study found that the Random Forest model had the highest accuracy and the lowest variance, with an average accuracy of 95.98% and an average RMSE of 20.99%. Additionally, the Principal Remaining on the Bill was identified as the most significant factor influencing repayment ability. Therefore, the Random Forest model is suitable for use in loan approval processes, as it demonstrated the highest predictive accuracy when analyzing the repayment behavior of previous borrowers. This can lead to more efficient and accurate loan approvals for the bank.

Graduate Studies Student's Signature.....

Field of Study Information Technology Advisor's Signature.....

Academic Year 2024

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างยิ่งจากอาจารย์ที่ปรึกษา ว่าที่ร้อยตรี ดร.พิชิตชัย คำอินทร์ ผู้ให้ความรู้ คำปรึกษา คำแนะนำ ชี้แนะวิธีการต่าง ๆ และให้ความอนุเคราะห์ช่วยเหลือทุกอย่างในการทำวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณ ไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณประธานและคณะกรรมการสอบวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.วรภัทร ไพรีเกรง ดร.ภูวดล ศิริกองธรรม และ ดร.อภิชญา นิ้มคุ้มภัย ผู้ให้คำแนะนำข้อแก้ไขที่ควรปรับปรุงจนวิทยานิพนธ์ฉบับนี้สมบูรณ์

ขอกราบขอบพระคุณคณาจารย์ สถาบันเทคโนโลยีไทย-ญี่ปุ่น ผู้ประสิทธิ์ประสาทวิชาความรู้ และบุคลากรคณะเทคโนโลยีสารสนเทศ ผู้ให้คำแนะนำในการจัดทำเอกสารที่เกี่ยวข้อง และให้ความช่วยเหลือในการดำเนินการขั้นตอนต่าง ๆ ของทางบัณฑิตวิทยาลัย รวมทั้งขอขอบพระคุณผู้บริหารที่ให้ความกรุณาให้ผู้วิจัยสามารถนำข้อมูลไปใช้ในวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณครอบครัวอันเป็นที่รักและเพื่อน ๆ ที่คอยให้คำแนะนำ กำลังใจ และเป็นแรงผลักดันที่สำคัญ ทำให้ผู้วิจัยสามารถดำเนินงานวิจัยและจัดทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายนี้ผู้วิจัยหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์แก่ผู้อื่นต่อไป รวมถึงการนำไปประยุกต์ใช้ให้เกิดประโยชน์ในด้านการวิเคราะห์อนุมัติสินเชื่อ เพื่อพัฒนาต่อยอดให้มีความก้าวหน้าและมีประสิทธิภาพมากยิ่งขึ้น

ธันย์ชนก ขำประดิษฐ์

สารบัญ

สารบัญ (ต่อ)

สารบัญตาราง

สารบัญรูป

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเศรษฐกิจไทยมีแนวโน้มที่มีการขยายตัวเพิ่มขึ้นอย่างต่อเนื่อง แม้ในปี 2566 ที่ผ่านมามีอัตราการขยายตัวเพิ่มขึ้นแต่ยังไม่ครอบคลุมทุกภาคเศรษฐกิจ พบว่า ภาคการส่งออกสินค้าและภาคอุตสาหกรรมยังมีแนวโน้มการขยายตัวที่ค่อนข้างช้ากว่าภาคเศรษฐกิจอื่น ๆ อย่างไรก็ตามในปี 2567 และปี 2568 มีอัตราการขยายตัวเพิ่มขึ้นอย่างสมดุล เศรษฐกิจจะปรับตัวดีขึ้นตามการขยายตัวของภาคการส่งออกสินค้าและมีการฟื้นตัวอย่างต่อเนื่องของภาคการท่องเที่ยวแต่ยังต้องระมัดระวังในภาคการส่งออกสินค้าออกไปต่างประเทศที่อาจจะไม่ได้รับผลดีเท่าที่ควรเนื่องจากยังมีปัญหาเศรษฐกิจโลก การเกิดสงคราม ภาวะเงินเฟ้อ เป็นต้น ตั้งแต่สถานการณ์โควิด-19 คลี่คลายจะเห็นได้ว่าเศรษฐกิจไทยเริ่มมีการฟื้นตัวขึ้นมาบ้าง เครื่องชี้วัดเศรษฐกิจหลายด้านก็เริ่มกลับมาอยู่ในระดับเทียบเท่าหรือสูงกว่าช่วงก่อนสถานการณ์โควิด-19 แล้ว โดยเฉพาะในด้านอุปสงค์ ในส่วนการบริโภคและรายได้ประชาชน แต่ในด้านอุปทาน ส่วนของภาคการผลิตยังฟื้นตัวได้ไม่ดีนักโดยเฉพาะการผลิตสินค้าส่งออกไปยังประเทศคู่ค้า โดยเฉพาะประเทศจีนที่ยังชะลอตัวอยู่และด้านวัฏจักรอิเล็กทรอนิกส์โลก (Electronic Cycle) ที่ยังฟื้นตัวไม่เต็มที่ แม้ว่าภาคการท่องเที่ยวจะฟื้นตัวต่อเนื่องแต่ยังค่อนข้างฟื้นตัวช้าเนื่องจากนักท่องเที่ยวบางชาติ เช่น จีน มีการเข้ามาท่องเที่ยวในประเทศที่ลดลงรวมถึงรายจ่ายต่อหัวเริ่มลดลง

สินเชื่อหรือหนี้ล้วนมีความหมายที่คล้ายกัน ต่างกันเพียงการพิจารณาเท่านั้น หนี้หรือเครดิตเป็นข้อสัญญาผูกผันที่จะชำระสิ่งมีค่าในอนาคต โดยปกติจะระบุเป็นจำนวนเงิน เพราะเงินทำหน้าที่เป็นหน่วยวัดมูลค่า (Unit of Account) และเป็นมาตรฐานเลื่อนการชำระหนี้ในอนาคต (Stand of Deferred Payment) โดยที่ฝ่ายเสียสละเงินหรือของมีค่าให้กับผู้อื่นและได้รับคำมั่นสัญญาที่จะได้รับชำระคืนในอนาคต เรียกว่า เครดิต แต่ถ้าผู้ที่เป็นฝ่ายได้รับเงินหรือสิ่งของและมีคำมั่นสัญญาจะชำระเงินในอนาคต เรียกว่า หนี้ [1]

หนี้ครัวเรือนของคนไทยอ้างอิงจากข้อมูลสินเชื่อในระบบที่อยู่ในเครดิตบูโร ณ เดือนมีนาคม 2565 พบว่าคนไทยที่มีหนี้สูงถึง 37% หรือราว 25 ล้านคน คิดเป็น 1 ใน 3 ของประชากรไทยทั้งหมด โดยมีสัดส่วนเพิ่มขึ้น 30% จากปี 2560 นอกจากนี้คนไทยยังมีหนี้ในปริมาณที่เพิ่มสูงขึ้น โดยประมาณ 57% ของคนไทยมีหนี้มีหนี้ตั้งแต่ 100,000 บาทจนถึง 1 ล้านบาท ซึ่งกลุ่มคนที่มีหนี้มากกว่า 100,000 บาท มีอยู่ถึง 43% และกลุ่มคนที่มีหนี้เกิน 1 ล้านบาท มีอยู่ 14% โดยมูลค่าหนี้เฉลี่ยต่อคนอยู่ที่ 520,000 บาท ในภาพรวมมูลค่าหนี้ของคนไทยมีการเพิ่มขึ้นเกือบ 2 เท่าในช่วง 10 ปีที่ผ่านมา

หากดูสัดส่วนของคนที่มีหนี้แยกตามจำนวนบัญชีหนี้ที่มี พบว่า 32% ของคนไทยที่มีหนี้ 4 บัญชีขึ้นไป หากดูสัดส่วนบัญชีหนี้ แยกตามผลิตภัณฑ์ พบว่าคนไทยมีหนี้ที่ไม่ก่อให้เกิดรายได้ในประเภทของสินเชื่อส่วนบุคคล 39% และสินเชื่อบัตรเครดิต 29% ซึ่งมากกว่าหนี้ที่ก่อให้เกิดรายได้ โดยสัดส่วนหนี้ที่ก่อให้เกิดรายได้ประเภทบ้านมีสัดส่วน 4% และประเภทธุรกิจเพียง 4%

การให้สินเชื่อของธนาคารเป็นการใช้เงินทุนเพื่อก่อให้เกิดรายได้หลักของธนาคาร โดยธนาคารจะได้รับรายได้ในรูปแบบของดอกเบี้ย ยิ่งธนาคารอนุมัติสินเชื่อมากเท่าไหร่ ธนาคารก็จะได้รับรายได้เพิ่มมากขึ้นเท่านั้น นอกจากนี้การอนุมัติการให้สินเชื่อยังเป็นวิธีการที่ธนาคารนำเงินที่ได้จากการะดมเงินทุนแล้วนำไปกระจายเงินทุนในรูปแบบของสินเชื่อ ถือเป็นการช่วยแก้ปัญหาของประชาชนคนอื่น ๆ ซึ่งเป้าหมายหลักของการให้สินเชื่อคือ การได้รับเงินต้นและดอกเบี้ยคืนจากลูกหนี้ภายในระยะเวลาตามที่สัญญากำหนด หากธนาคารดำเนินการให้สินเชื่อที่ดี ได้รับการชำระหนี้คืนภายในระยะเวลาที่กำหนด ธนาคารก็จะได้รับกำไรเพิ่มขึ้น แต่ในอีกแง่หนึ่งธนาคารก็ย่อมแบกรับความเสี่ยงจากการไม่ได้รับชำระหนี้คืนภายในระยะเวลาที่กำหนดหรือไม่ได้รับเงินคืนทั้งต้นเงินและดอกเบี้ยเช่นกัน อาจส่งผลให้เกิดหนี้สูญ ผลประกอบการธนาคารลดลง กำไรลดลงและส่งผลเสียต่อความน่าเชื่อถือของธนาคารได้ ดังนั้น ธนาคารจึงมีความจำเป็นอย่างมากที่ต้องจัดการในด้านการอนุมัติสินเชื่อ การปล่อยสินเชื่ออย่างระมัดระวังและรอบคอบมากที่สุด ซึ่งปัญหาหลักของธนาคารที่ต้องการจะปล่อยสินเชื่อคือ ธนาคารต้องการทราบว่าภายหลังจากที่อนุมัติสินเชื่อไปแล้ว ผู้กู้จะสามารถมาชำระหนี้ได้หรือไม่ จะไม่ผิดนัดชำระหนี้ ไม่ผิดเงื่อนไขตามสัญญากู้ [2] และผู้กู้ลักษณะใดที่มีความเสี่ยงสูงอาจก่อให้เกิดหนี้เสีย หรือประวัติการชำระหนี้แบบใดส่งผลให้เห็นสมควรอนุมัติสินเชื่อโดยอาศัยจากประวัติการชำระหนี้ของผู้กู้เดิม นอกจากนี้ธนาคารยังสามารถบวกค่าความเสี่ยงที่จะเกิดกับผู้กู้กลุ่มนี้ได้เพื่อป้องกันการเกิดหนี้สูญที่จะเกิดขึ้น

ผู้วิจัยได้ตระหนักถึงปัญหาดังกล่าวที่ธนาคารพบเจอ ปัจจัยต่าง ๆ ที่เกิดขึ้นบ่อย และการที่ธนาคารต้องแบกรับความเสี่ยงที่อาจจะเกิดหนี้เสีย แม้จะมีเครื่องมือที่เข้ามาช่วยพิจารณาประวัติการชำระหนี้ในอดีตของผู้กู้รายเดิม แต่ยังไม่สามารถเข้ามาช่วยวิเคราะห์ได้เต็มประสิทธิภาพและมีความแม่นยำชัดเจนเพียงพอ ผู้วิจัยจึงอยากจะสร้างโมเดลเพื่อเป็นเครื่องมือทางเลือกในการช่วยวิเคราะห์การอนุมัติสินเชื่อให้ผู้กู้รายเดิม โดยนำข้อมูลในอดีตของผู้กู้มาทำนายแนวโน้มความเสี่ยงที่จะผิดนัดชำระหนี้ที่อาจเกิดขึ้นในอนาคต ทำให้เกิดความถูกต้องแม่นยำและนำโมเดลที่ได้จาก Machine Learning เข้าไปปรับใช้ควบคู่กับเครื่องมือ Credit Scoring Model ของธนาคารที่มีอยู่เพื่อช่วยให้การพิจารณาการขอสินเชื่อมีความแม่นยำและมีประสิทธิภาพมากยิ่งขึ้น นอกจากนี้จะช่วยให้ธนาคารสามารถกลั่นกรองลูกค้าชั้นดี มีคุณลักษณะที่เหมาะสมสำหรับการปล่อยอนุมัติสินเชื่อได้อีกทางเลือกหนึ่ง

1.2 วัตถุประสงค์ของการศึกษา

1.2.1 เพื่อสร้างแบบจำลองข้อมูลการในการวิเคราะห์ความสามารถในการชำระหนี้คืนโดยใช้เทคนิคการเรียนรู้ของเครื่อง

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมาะสมในการจำแนกความสามารถในการชำระหนี้คืนจากข้อมูลลูกค้าผู้กู้รายเดิม

1.3 ขอบเขตการวิจัย

1.3.1 ศึกษาเทคนิคและทดสอบการวิเคราะห์ความสามารถในการชำระหนี้คืนโดยใช้ข้อมูลลูกค้าผู้กู้รายเดิมโดยใช้ประวัติข้อมูลลูกค้าผู้กู้รายเดิมย้อนหลัง ระยะเวลา 12 เดือน

1.3.2 เปรียบเทียบอัลกอริทึมที่มีผลต่อความสามารถในการชำระหนี้คืน โดยใช้เทคนิคการเรียนรู้ของเครื่อง ดังต่อไปนี้ Logistic Regression Neural Network Decision Tree และ Random Forest

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้เทคนิคการเรียนรู้ของเครื่องที่เหมาะสมในการนำมาใช้วิเคราะห์การชำระหนี้คืนสำหรับข้อมูลขนาดใหญ่

1.4.2 ได้แนวทางประยุกต์ใช้งานจริงในธนาคาร โดยการนำโมเดลที่ได้ไปช่วยสนับสนุนการตัดสินใจการปล่อยสินเชื่อให้กับลูกค้า

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็นการเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมด้วยเทคนิคการเรียนรู้ของเครื่อง โดยมีวัตถุประสงค์เพื่อสร้างแบบจำลองข้อมูลในการวิเคราะห์ความสามารถในการชำระหนี้คืนของผู้กู้รายเดิมโดยใช้เทคนิคการเรียนรู้ของเครื่องและเพื่อเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมาะสมในการจำแนกความสามารถในการชำระหนี้คืนจากข้อมูลลูกค้าผู้กู้รายเดิม โดยใช้ข้อมูลจากสถาบันการเงิน โดยแต่ละหัวข้อมีรายละเอียดดังนี้

2.1 แนวคิดและทฤษฎีเกี่ยวกับคุณภาพชีวิต

2.2 แนวคิดเกี่ยวกับหลักการบริหารสินเชื่อ

2.3 ระบบประเมินความเสี่ยงสินเชื่อรายย่อย

2.4 กระบวนการวิเคราะห์ข้อมูล

2.5 ทฤษฎีการเรียนรู้ของเครื่อง

2.6 แนวคิดการวัดค่าความถูกต้องของเครื่องมือ

2.7 แนวคิดการทดสอบประสิทธิภาพแบบจำลอง

2.8 แนวคิดทฤษฎีงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎีเกี่ยวกับคุณภาพชีวิต

UNESCO ได้นิยามคุณภาพชีวิต หมายถึง ระดับความเป็นอยู่ที่ดีของสังคมและระดับความพึงพอใจในความต้องการส่วนหนึ่งของมนุษย์ ดังนั้นคุณภาพชีวิตจึงเป็นการใช้ชีวิตที่ดีมีความสุขพึงพอใจในชีวิตและสภาพแวดล้อมที่เกี่ยวข้องกับความเป็นอยู่ในการดำเนินชีวิตของปัจเจกบุคคลในสังคม

WHO [3] ให้ความหมายว่า คุณภาพชีวิต หมายถึง การรับรู้หรือเข้าใจเกี่ยวกับปัจเจกบุคคลที่มีต่อสถานภาพชีวิตของตนเองตามบริบทของวัฒนธรรมที่เกิดขึ้นและค่านิยมที่ใช้ชีวิตอยู่ นอกจากนี้ยังมีความเกี่ยวข้องกันกับความคาดหวัง มาตรฐานและความกังวลสนใจที่เกิดขึ้นรอบตัว คุณภาพชีวิตเป็นมโนมติที่มีขอบเขตกว้างขวางครอบคลุมเรื่องต่าง ๆ ที่สลับซับซ้อน ได้แก่ สุขภาพทางกาย สภาวะทางจิต ความสัมพันธ์ต่าง ๆ ทางสังคม ความเชื่อส่วนบุคคลและสัมพันธภาพที่มีต่อสิ่งแวดล้อม

Padilla and Grant [4] ได้กล่าวถึง คุณภาพชีวิตเป็นการรับรู้ความพึงพอใจในตนเองที่มีโอกาสเปลี่ยนแปลงได้ตามภาวะสุขภาพ โดยวัดจากความสามารถในการทำงานตามส่วนต่าง ๆ ของร่างกาย

คุณภาพชีวิตเรียกกว้าง ๆ ว่า การเป็นอยู่ที่ดี (Well-being) ซึ่งหมายถึงการเป็นอยู่ที่ดีของตนเองและสิ่งแวดล้อมตามสภาพทั่วไป คุณภาพชีวิตจะแสดงออกมาในรูปแบบของความต้องการ (Want) เมื่อได้รับการตอบสนองความต้องการแล้ว ส่งผลให้บุคคลนั้น ๆ มีความสุขหรือความพึงพอใจเกิดขึ้น แต่ในทางสังคมและสภาพแวดล้อม การมีชีวิตที่ดีอย่างมีคุณภาพต้องไม่เป็นภาระหรือก่อให้เกิดปัญหาทางสังคมด้วยเช่นกัน

กรมการพัฒนาชุมชนให้นิยามไว้ว่า คุณภาพชีวิต หมายถึง การดำรงชีวิตของมนุษย์ให้เหมาะสมตามความจำเป็นพื้นฐานที่สังคมได้กำหนดไว้ในช่วงเวลาหนึ่ง การที่จะกล่าวว่า มนุษย์จะมีคุณภาพชีวิตที่ดีได้ก็ต่อเมื่อสมาชิกในครอบครัวหรือชุมชนนั้นมีชีวิต มีความเป็นอยู่ตามความจำเป็นพื้นฐานที่มีครบถ้วน ซึ่งเกณฑ์ความจำเป็นพื้นฐานที่กำหนดไว้แล้วนั้นสามารถเปลี่ยนแปลงได้ตามสภาพเศรษฐกิจและสังคมที่เปลี่ยนแปลงไปในช่วงเวลานั้นๆ [5]

พัณณิน กิติพราภรณ์ [6] ได้ให้คำจำกัดความของคุณภาพชีวิต คือ ชีวิตที่มีความสุข ซึ่งความสุขนั้นเกิดจากความสุข 2 ทาง ได้แก่ ความสุขสุขทางกาย หมายถึง มีความเป็นอยู่ที่ดี อาทิเช่น มีที่อยู่อาศัย มีสุขภาพ และHealth Care ที่ดี มีสาธารณูปโภคที่ดี เช่น การคมนาคมที่ดี มีสภาพแวดล้อมที่ดี เช่น น้ำ อากาศที่บริสุทธ์ และยังรวมไปถึงการพักผ่อนและสันทนาการที่ดีตามสมควรอีกด้วย และความสุขทางใจที่ได้จากการรู้จักพอดี ความพอใจในสถานภาพที่เป็นอยู่ การมีทัศนคติที่ดีต่อตนเองและผู้อื่น มีความรัก ความอบอุ่นผูกพันในครอบครัวและเพื่อนมนุษย์มีความอดทนเสียสละทำประโยชน์แก่สังคม

ศิริ ฮามสุโพธิ์ [7] กล่าวถึงคุณภาพชีวิต คือ บุคคลที่สามารถดำรงชีวิตอยู่ร่วมกับสังคมได้อย่างเหมาะสม โดยไม่เป็นภาระและบุคคลนั้นไม่ก่อให้เกิดปัญหากับสังคม เป็นบุคคล ที่มีความสมบูรณ์ทั้งร่างกายและจิตใจ สามารถใช้ชีวิตที่ชอบธรรมและสอดคล้องกับสภาพแวดล้อมของสังคม บุคคลที่สามารถแก้ไขปัญหาได้ รวมไปถึงการแสวงหาสิ่งที่ปรารถนาให้ได้มาอย่างถูกต้องภายใต้ทรัพยากรที่มีอยู่

Power Bullinger and WHOQOL Group [8] กล่าวถึง เครื่องมือชี้วัดคุณภาพชีวิตขององค์การอนามัยโลกฉบับภาษาไทย (WHOQOL-BREF-THAI) โดยได้มีการพัฒนาเครื่องมือชี้วัดคุณภาพชีวิตด้วยวิธีการทบทวนแนวคิดของคุณภาพชีวิตและการศึกษาเชิงคุณภาพ เพื่อกำหนดองค์ประกอบของคุณภาพชีวิต WHOQOL– BREF–THAI ประกอบด้วยแบบภาวะวิสัย (Perceived Objective) และแบบอัตวิสัย (Self-report Subjective) โดยองค์ประกอบของคุณภาพชีวิต 4 ด้าน มีดังนี้

(1) ด้านสุขภาพกาย (Physical Domain) คือ การรับรู้สภาพร่างกายของตนเองที่มีผลต่อชีวิตประจำวัน เช่น การรับรู้สภาพร่างกายที่มีความสมบูรณ์แข็งแรง การรับรู้ถึงความสุขสบายกาย การรับรู้ถึงความสามารถที่สามารถจัดการกับความเจ็บปวดทางร่างกายได้ การรับรู้ถึงพละกำลัง

ในการดำเนินชีวิต การรับรู้ถึงความเป็นอิสระที่ไม่พึ่งพาผู้อื่น การรับรู้ถึงความสามารถในการเคลื่อนไหว การรับรู้ถึงความสามารถในการทำงาน เป็นต้น

(2) ด้านจิตใจ (Psychological Domain) คือ การรับรู้สภาพจิตใจของตนเองที่มีผลต่อการใช้ชีวิต เช่น การรับรู้ถึงความรู้สึกดีที่มีต่อตนเอง การรับรู้ถึงภาพลักษณ์ของตนเอง การรับรู้ถึงความรู้สึกภาคภูมิใจในตนเอง การรับรู้ถึงความมั่นใจในตนเอง การรับรู้ถึงความสามารถของตนเองในการจัดการกับความกังวลหรือเศร้าที่มี การรับรู้เกี่ยวกับความเชื่อต่าง ๆ เป็นต้น

(3) ด้านความสัมพันธ์ทางสังคม (Social Relationships) คือ การรับรู้ระหว่างความสัมพันธ์ของตนเองกับบุคคลอื่น การรับรู้ถึงการได้รับความช่วยเหลือจากบุคคลอื่นในสังคม การรับรู้ว่าตนเองก็เป็นผู้ให้ความช่วยเหลือบุคคลอื่นในสังคมด้วยเช่นกัน เป็นต้น

(4) ด้านสิ่งแวดล้อม (Environment) คือ การรับรู้เกี่ยวกับสิ่งแวดล้อมที่มีผลต่อการดำเนินชีวิต เช่น การรับรู้ว่าตนเองมีชีวิตอย่างอิสระ มีความปลอดภัยและมั่นคงในชีวิต การรับรู้ได้ว่าตนเองอยู่ในสิ่งแวดล้อมที่ดี ปราศจากมลพิษ การคมนาคมสะดวก เป็นต้น

นิศารัตน์ ศิลปเดช [9] มีการจำแนกองค์ประกอบของคุณภาพชีวิตไว้ ดังนี้

(1) ความสมบูรณ์ด้านร่างกายและสติปัญญา หมายถึง การมีคุณภาพชีวิตที่ดีนั้นต้องมีความปกติในด้านร่างกายและสติปัญญา ได้แก่ การมีสุขภาพอนามัยที่แข็งแรง มีพละกำลังที่สามารถทำงานได้ดีเช่นเดียวกับคนอื่น ๆ รวมทั้งสามารถแก้ไขปัญหาที่เกิดขึ้นได้

(2) ความสมบูรณ์ด้านจิตใจและอารมณ์ หมายถึง การเป็นมนุษย์ที่มีจิตใจดี อารมณ์แจ่มใสมั่นคง ไม่หงุดหงิด ไม่โมโหง่าย มองโลกในแง่ดี โอบอ้อมอารีช่วยเหลือผู้อื่น โดยการมีพื้นฐานเป็นบุคคลที่มีจิตใจดีจะช่วยให้เกิดความสุขและสงบในการดำรงชีวิต

(3) ความสมบูรณ์ด้านสังคมและสิ่งแวดล้อม หมายถึง การเป็นคนที่ได้รับการยอมรับจากบุคคลอื่น เนื่องจากเป็นบุคคลที่มีมนุษยสัมพันธ์ดี มีความสามารถในการปรับตัวและยอมรับความสามารถของบุคคลอื่น การมีความสมบูรณ์ด้านสังคมและสิ่งแวดล้อมจะช่วยให้บุคคลนั้นมีชีวิตที่เหมาะสมเข้ากันได้ดีกับสังคมและสิ่งแวดล้อมที่เป็นอยู่

(4) ความสมบูรณ์ด้านปัจจัยการดำรงชีพ หมายถึง ความสามารถที่จะจัดหาสิ่งจำเป็นต่าง ๆ ที่จะช่วยให้ชีวิตดำรงอยู่ได้อย่างดีตามฐานะของตนเองตลอดจนตามสภาพเศรษฐกิจและสังคม ปัจจัยการดำรงชีพที่จำเป็น คือ ปัจจัย 4 ประกอบด้วย อาหาร เครื่องนุ่งห่ม ที่อยู่อาศัย และยารักษาโรค การมีปัจจัยการดำรงชีพที่จำเป็นอย่างพอเพียงจะช่วยให้บุคคลนั้นมีความสะดวกสบายนำไปสู่ความสุขและความพึงพอใจในชีวิต

คุณภาพชีวิตเป็นสิ่งที่บ่งบอกถึงระดับการพัฒนาคุณภาพชีวิตของคนในประเทศ หากมีคนในประเทศมีคุณภาพชีวิตที่ดี สามารถใช้ชีวิตบนความจำเป็นพื้นฐานต่อการดำรงชีวิตของมนุษย์ได้ หากคุณภาพชีวิตของคนในประเทศดีแล้วและมีคุณภาพชีวิตที่ดีในด้านร่างกาย จิตใจ สังคมและ

สิ่งแวดล้อม ย่อมส่งผลไปยังประเทศชาติด้วย เมื่อคุณภาพชีวิตดีย่อมส่งผลให้สังคมและประเทศชาติดียิ่งขึ้นไปด้วย

2.2 แนวคิดเกี่ยวกับหลักการบริหารสินเชื่อ

สินเชื่อ (Credit) เป็นเครื่องมือชนิดหนึ่งในการเป็นสื่อกลางของการแลกเปลี่ยนเงินตรา (Money) แม้ว่าสินเชื่อจะไม่ใช่เงินตรา (No Money) แต่สินเชื่อก็เป็นที่ยอมรับกันอย่างกว้างขวางว่ามีความใกล้เคียงกับเงินตรามากที่สุด แต่การจะใช้สินเชื่อเป็นสื่อกลางในการแลกเปลี่ยนนั้นมีความแตกต่างจากการใช้เงินตรา เนื่องจากสินเชื่อมีลักษณะเป็นสัญญาผูกพัน (Promise) อันมีผลต่อเนื่องไปสู่อนาคต ซึ่งจะต้องมีการไถ่ถอนหนี้โดยต้องมีการชำระหนี้ตามสัญญาผูกพันเกิดขึ้น เมื่อฝ่ายหนึ่งหรือเจ้าหนี้ส่งมอบสินค้าหรือบริการให้อีกฝ่ายหนึ่งซึ่งคือ ลูกหนี้ ย่อมคาดหวังว่าลูกหนี้จะปฏิบัติตามข้อตกลงต่อกันว่าจะชำระหนี้ให้ตามเวลาที่กำหนดและตามจำนวนที่มีการตกลงกันไว้ ซึ่งเจ้าหนี้จะต้องพิจารณาถึงความสามารถของลูกหนี้จากการประกอบอาชีพและมีความซื่อสัตย์ต่อกัน อีกทั้งหากมีเหตุสุดวิสัยที่อาจเกิดขึ้นลูกหนี้ต้องร่วมกันในการแก้ปัญหาภาระหนี้สินด้วยความเต็มใจ โดยเจ้าหนี้ย่อมมีความเสี่ยงในการรับภาระหนี้ จึงต้องมีการเรียกร้องผลตอบแทนในรูปแบบของดอกเบี้ย ค่าธรรมเนียม ส่วนลดรับ รวมถึงเรียกร้องสิ่งที่มีค่า สินทรัพย์ เพื่อเป็นหลักประกัน ดังนั้นสินเชื่อถือเป็นปัจจัยที่มีความสำคัญต่อการขยายตัวของของการผลิตสินค้าและบริการ ในทางสังคมสินเชื่อยังทำให้พฤติกรรมและคุณภาพในชีวิตของคนเปลี่ยนแปลงไป โดยเฉพาะการเปลี่ยนแปลงในพฤติกรรมการผลิต การบริโภคและการใช้จ่ายในภาคครัวเรือนและภาคธุรกิจ

2.2.1 ความหมายของสินเชื่อ

(1) ความหมายของสินเชื่อโดยทั่วไป

คำว่าสินเชื่อ (Credit) มาจากศัพท์ในภาษาลาตินของคำว่า Credere แปลว่าความเชื่อถือหรือความไว้วางใจ (To Trust or to Believe) ว่าฝ่ายที่รับสินเชื่อจะต้องมีการชำระหนี้คืนภายในระยะเวลาและเงื่อนไขที่กำหนด ความหมายของสินเชื่อ คือ ความสามารถในการกู้ยืมเงินหรือความสามารถในการได้รับสินค้าบริการเป็นเงินเชื่อ โดยมีคำมั่นสัญญาว่าจะชำระคืน (Repayment) ในอนาคต

(2) ความหมายของสินเชื่อที่มีลักษณะเฉพาะ

1. ความหมายของสินเชื่อในแง่ของผู้บริโภค หมายถึง ความสามารถที่จะได้รับสินค้าหรือบริการไปใช้ก่อนล่วงหน้า แต่มีข้อตกลงกันว่าจะมีการชำระสินค้าหรือบริการภายในระยะเวลาที่กำหนด

2. ความหมายของสินเชื่อในแง่ของการค้า หมายถึง ความเชื่อถือที่ผู้ขายมีต่อผู้ซื้อ โดยผู้ขายมอบสินค้าหรือบริการให้แก่ผู้ซื้อไปก่อน โดยมีข้อตกลงร่วมกันว่าจะชำระค่าสินค้าหรือบริการในภายหลังตามที่ตกลงกัน ก่อให้เกิดภาวะความเป็นลูกหนี้ต่อกัน

3. ความหมายของสินเชื่อในแง่ของสถาบันการเงิน หมายถึง บริการชนิดหนึ่งของสถาบันการเงินที่ก่อให้เกิดรายได้หลัก

<u>กระบวนการให้สินเชื่อ</u> มี 3 ขั้นตอน ดังนี้

1. การเกิดรายการสินเชื่อ (Credit Transaction) โดยเริ่มจากบุคลทั้ง 2 ฝ่าย มีการตกลงกันที่จะทำการแลกเปลี่ยนสินค้าหรือบริการหรือการให้กู้ยืม โดยมีสัญญาและเงื่อนเวลาที่จะชำระเงินคืนในอนาคต

2. สถานะทางสินเชื่อ (Credit Standing) เมื่อได้มีการตกลงกันที่จะให้สินเชื่อแล้ว สิ่งที่จะต้องพิจารณาและตัดสินใจในขั้นต่อมา คือ สถานะทางสินเชื่อของผู้ขอกู้หรือขอใช้บริการสินเชื่อว่ามีความน่าเชื่อถือมากน้อยเพียงใดทั้งในด้านลักษณะส่วนตัว ผลประกอบการชื่อเสียงของการดำเนินธุรกิจตลอดจนความสามารถที่จะชำระหนี้ สถานะทางสินเชื่อจะเป็นตัวบ่งชี้ของการยอมรับในการกำหนดวงเงินสินเชื่อ เงื่อนไขเวลาของการให้สินเชื่อว่ามีมากน้อยเพียงใดและอย่างไรซึ่งสิ่งเหล่านี้จะมีความแตกต่างกันในผู้ขอสินเชื่อแต่ละราย

3. การใช้เครื่องมือประกอบด้านสินเชื่อหรือตราสารสินเชื่อ (Credit Instrument) เป็นขั้นตอนสุดท้ายของกระบวนการสินเชื่อที่จะต้องมีการทำหลักฐานเพื่อแสดงการตกลงเกี่ยวกับวงเงินสินเชื่อ เงื่อนไขและระยะเวลาที่ทั้ง 2 ฝ่ายตกลงยอมรับ ได้แก่ ตราสาร หรือสัญญาที่เป็นหลักประกันการชำระหนี้ในอนาคต เช่น หนังสือสัญญาเงินกู้ ตั๋วสัญญาใช้เงิน ตั๋วแลกเงิน หรือเช็ค เป็นต้น ดังนั้นเครื่องมือประกอบด้านสินเชื่อหมายถึงหลักฐานแสดงถึงสภาพหนี้ เงื่อนไขและระยะเวลาการชำระหนี้ที่จัดทำขึ้นเพื่อป้องกันความเสี่ยงในการประกอบการด้านสินเชื่อ [10]

<u>บทบาทของสินเชื่อ</u>

1. บทบาทของสินเชื่อต่อผู้บริโภค สินเชื่อสามารถทำให้คุณภาพชีวิตของผู้บริโภคดีขึ้นโดยที่ผู้บริโภคสามารถหาสินค้าและบริการมาใช้เพื่อตอบสนองความต้องการของตนเองได้อย่างสะดวกรวดเร็ว โดยเฉพาะสินค้าที่มีความจำเป็นและสินค้าที่มีมูลค่าสูง เช่น บ้าน ที่ดิน นอกจากนี้สินเชื่อยังช่วยบรรเทาความทุกข์ในคราวจำเป็นได้ เช่น การเจ็บป่วยกะทันหัน

2. บทบาทของสินเชื่อต่อผู้ผลิตและผู้ให้บริการ สินเชื่อทำให้ผู้ผลิตมีทุนในการขยายการผลิต หากการผลิตเพิ่มจำนวนได้มากขึ้น ต้นทุนการผลิตก็จะลดลง นอกจากนี้ยังทำให้ผู้ผลิตสามารถระบายสินค้าออกไปได้อย่างรวดเร็วด้วยวิธีการขายสินค้าเป็นเงินเชื่อ ซึ่งจะช่วยลดต้นทุนของ

การจัดเก็บสินค้าและการดูแลรักษาสินค้าอีกด้วย อีกทั้งยังเป็นการขยายช่องทางการจำหน่ายสินค้าให้กว้างขวางมากยิ่งขึ้น

3. บทบาทของสินเชื่อต่อสถาบันการเงิน สถาบันการเงินเกือบทุกแห่งมีรายได้จากธุรกิจสินเชื่อเป็นรายได้หลัก โดยสถาบันการเงินมีหน้าที่ในการระดมเงินออมจากผู้มีเงินเหลือเก็บออมและนำเงินดังกล่าวมาหมุนเวียนให้แก่ผู้ที่ต้องการใช้เงิน เพื่อให้เกิดสภาพคล่องทางเศรษฐกิจ กระตุ้นให้เกิดการหมุนเวียนทางเศรษฐกิจมากขึ้น

4. บทบาทของสินเชื่อด้านเศรษฐศาสตร์ สินเชื่อถือเป็นการเพิ่มความต้องการของ Supply ของเงินในระบบเศรษฐกิจ หากมีเพิ่มขึ้นอย่างเหมาะสมพอดีกับความต้องการของ Demand ของตลาด ย่อมส่งผลดีต่อเศรษฐกิจในภาพรวม หากมี Supply มากเกินไปจะส่งผลให้เกิดเงินเฟ้อขึ้นได้ หากไม่มีการควบคุมการให้สินเชื่อย่อมก่อให้เกิดเงินเฟ้อเพิ่มสูงขึ้นอีกในระบบเศรษฐกิจ [11]

<u>การจำแนกข้อมูล</u>

ข้อมูลสินเชื่อประกอบด้วยข้อมูลที่เป็นตัวเลขและข้อมูลที่ไม่เป็นตัวเลข ข้อมูลแบ่งได้เป็น 2 ประเภท ดังนี้

1. ข้อมูลเชิงคุณภาพ เช่น ประวัติส่วนตัวของผู้ขอสินเชื่อ ความซื่อสัตย์ ความมีชื่อเสียง ความสามารถในการประกอบการ ประวัติการชำระหนี้ การหาข้อมูลดังกล่าวทำได้ โดยการสอบถามจากตัวผู้ขอให้สินเชื่อหรือบุคคลที่ใกล้ชิด ตรวจสอบที่บ้านหรือสถานประกอบการของผู้ขอใช้สินเชื่อ

2. ข้อมูลเชิงปริมาณ ข้อมูลส่วนนี้เป็นข้อมูลที่เป็นตัวเลข ซึ่งรวบรวมมาสำหรับใช้วิเคราะห์ฐานะของผู้ขอสินเชื่อสถานะของโครงการที่ขอสินเชื่อ

<u>หลักเกณฑ์เบื้องต้นในการพิจารณาสินเชื่อ</u>

ในการพิจารณาอนุมัติสินเชื่อจะขึ้นอยู่กับนโยบายและหลักเกณฑ์ของผู้ให้สินเชื่อแต่ละราย โดยทั่วไปแล้วมีปัจจัยหลัก ๆ ที่ใช้ประกอบการพิจารณา ดังนี้

1. นโยบายสินเชื่อของผู้ให้สินเชื่อ เช่น ผู้ให้สินเชื่อบางรายอาจกำหนดว่าผู้ยื่นขอสินเชื่อต้องไม่มีประวัติการค้างชำระในช่วง 12 เดือนย้อนหลัง หรืองดให้สินเชื่อแก่ลูกค้าใหม่ในกลุ่มอาชีพหรือกลุ่มอุตสาหกรรมที่มีความเสี่ยงสูง

2. วัตถุประสงค์ในการขอสินเชื่อ เช่น ใช้เป็นเงินทุนหมุนเวียนในการทำธุรกิจหรือลงทุนขยายโรงงาน ซึ่งจะเป็นข้อมูลประกอบการพิจารณาความสามารถในการชำระหนี้ในอนาคต

3. คุณลักษณะและความสามารถในการชำระหนี้ของผู้ขอสินเชื่อ ซึ่งจำเป็นต้องมีการวิเคราะห์ความสามารถของผู้กู้เพื่อประเมินความเสี่ยง การพิจารณาคุณลักษณะจะใช้หลัก 7C’s และ 5P’s ดังนี้

เกณฑ์การให้สินเชื่อ 7C (The 5C's of Credit)

เป็นเกณฑ์การให้สินเชื่อที่ธนาคารนำมาพิจารณาและประเมินความน่าเชื่อถือในการอนุมัติสินเชื่อ ประกอบด้วย

1. ตัวผู้กู้ (Character)

พิจารณาจากคุณสมบัติของผู้ขอสินเชื่อเกี่ยวกับ สถานะ อายุ อาชีพ ความมั่นคงของรายได้และประวัติการชำระหนี้ เพื่อดูข้อมูลในอดีตเกี่ยวกับการขอสินเชื่อและการชำระหนี้ที่ผ่านมา

2. ความสามารถในการชำระหนี้ (Capacity)

ความสามารถในการจ่ายชำระหนี้คือพิจารณาดูจากระยะเวลาที่กำหนดไว้ในอดีต รวมถึงความมั่นคงของรายได้ที่จะนำมาชำระหนี้ในอนาคต โดยดูจากคุณสมบัติของผู้ขอสินเชื่อเกี่ยวกับรายได้ ความสามารถในการหากำไร รายรับ รายจ่าย ภาระค่าใช้จ่าย ภาระหนี้สินที่มีอยู่และข้อมูลเครดิตในปัจจุบัน

3. เงินทุน (Capital)

พิจารณาจากคุณสมบัติของผู้ขอสินเชื่อเกี่ยวกับเงินทุน สินทรัพย์ หรือบัญชีเงินฝากของผู้ขอสินเชื่อเพื่อเป็นหลักประกันการให้กู้ยืม โดยพิจารณาจากสัดส่วนที่มีอยู่เหมาะสมหรือไม่ที่ใช้ในการชำระหนี้ที่อาจเปลี่ยนสภาพเป็นความสามารถในการชำระหนี้ในอนาคตได้

4. หลักประกัน (Collateral)

พิจารณาจากคุณสมบัติของผู้กู้เกี่ยวกับหลักประกัน ดูจากสินทรัพย์ที่นำมาค้ำประกันและผู้ค้ำประกัน ซึ่งหลักประกันเป็นปัจจัยในการลดความเสี่ยงของการให้สินเชื่อ ในกรณีที่ผู้กู้ไม่สามารถชำระหนี้ได้ โดยหลักประกันสามารถแบ่งได้เป็น 2 รูปแบบ คือ หลักประกันแบบบุคคลค้ำประกัน เป็นบุคคลธรรมดาหรือนิติบุคคล หากเป็นบุคคลธรรมดาพิจารณาจากรายได้ ฐานะทางการเงิน ความมั่นคงทางอาชีพ สินทรัพย์ หนี้สิน และความสามารถในการชำระหนี้ของตัวผู้ค้ำประกัน หากเป็นนิติบุคคลจะพิจารณาจากผลประกอบกิจการย้อนหลัง กำไรของกิจการ เงินทุน ความสามารถในการชำระหนี้ และหลักประกันแบบใช้หลักทรัพย์ค้ำประกัน พิจารณาจากที่ตั้ง มูลค่าหลักทรัพย์

5. สภาวการณ์เศรษฐกิจทั่วไป (Condition)

พิจารณาจากสภาวะทางเศรษฐกิจ ซึ่งเป็นปัจจัยภายนอกที่ไม่สามารถควบคุมได้ ส่งผลกระทบต่อรายได้ของผู้ขอสินเชื่อ เช่น สภาวะตลาด เงินเฟ้อ แนวโน้มธุรกิจ การแข่งขันทางธุรกิจ นโยบายของรัฐ ความเคลื่อนไหวของราคาสินค้าที่มีผลกระทบต่อความเป็นไปได้ของรายได้ของผู้ขอสินเชื่อ ซึ่งจะมีผลต่อความสามารถในการชำระหนี้ในอนาคต

6. Country & Currency

เป็นการพิจารณาถึงความเสี่ยงที่เกิดจากความไม่แน่นอนของภาวะทางเศรษฐกิจ สังคม การเมือง หรือปัจจัยภายนอกอื่น ๆ เช่น ภัยธรรมชาติ ภัยจากการเกิดโรคระบาด ความไม่สงบทางสังคม

การเมือง เป็นต้น โดยธนาคารปล่อยสินเชื่อให้ผู้ขอสินเชื่อที่อยู่ในประเทศ แต่แหล่งที่มาของเงินที่จะนำมาชำระหนี้ของผู้ขอสินเชื่อมาจากต่างประเทศ ซึ่งแหล่งที่มาของเงินอาจมาจากลูกหนี้การค้า เจ้าหนี้การค้า แหล่งนำเข้าสินค้าและวัตถุดิบ แหล่งตลาดการส่งออกเพื่อจำหน่ายสินค้า คู่สัญญาทางการค้าหรือคู่ค้าในการก่อภาระผูกพันจากธุรกรรมด้าน Trade Finance นอกจากนี้อาจเกิดความผันผวนด้านอัตราแลกเปลี่ยนของสกุลเงินที่ใช้เป็นตัวกลางในการดำเนินธุรกรรมทางการค้ากับประเทศดังกล่าวจนอาจ ส่งผลกระทบต่อระดับความเสี่ยงหรือความน่าเชื่อถือในการดำเนินธุรกิจ ซึ่งอาจส่งผลกระทบต่อฐานะทางการเงินและการดำเนินงานของผู้ขอสินเชื่อ ดังนั้นการติดตามภาวะเศรษฐกิจ สังคม การเมือง และการตรวจสอบรายชื่อแต่ละประเทศอยู่ใน Country Risk หรือไม่ รวมถึงการพิจารณาแนวโน้ม โอกาส และสาเหตุที่จะเกิดการเปลี่ยนแปลงการผันผวนของอัตราแลกเปลี่ยนจึงเป็นสิ่งสำคัญและจำเป็นในการวิเคราะห์สินเชื่อ เพื่อหาแนวทางป้องกันความเสี่ยงต่าง ๆ ที่อาจเกิดขึ้นได้

7. Control

การควบคุมการใช้สินเชื่อ (Credit Control) เป็นการให้ความสำคัญด้านวัตถุประสงค์ของแหล่งใช้ไปของสินเชื่อและการควบคุมการติดตามการใช้สินเชื่อ เพื่อให้เกิดการนำไปใช้อย่างมีประสิทธิภาพกับธุรกิจมากที่สุด โดยต้องมีความสอดคล้องและเป็นไปตามวัตถุประสงค์ที่อนุมัติสินเชื่อ รวมถึงสามารถตรวจสอบย้อนหลังได้ ซึ่งการควบคุมและติดตามที่นอกเหนือไปจากระเบียบและข้อกำหนดที่เจ้าหน้าที่ต้องควบคุมปฏิบัติตามเงื่อนไขต่าง ๆ ที่กำหนดขึ้น เพื่อควบคุมติดตามการใช้สินเชื่อ อาจประกอบด้วยเงื่อนไขที่ต้องปฏิบัติและควบคุมในช่วงก่อนเบิกจ่ายสินเชื่อ ระหว่างใช้สินเชื่อและเงื่อนไขหลังเบิกใช้สินเชื่อ เมื่อสินเชื่อได้รับการอนุมัติพร้อมเงื่อนไขดังกล่าวแล้วปรากฏว่าผู้ขอสินเชื่อมีความประสงค์จะขอผ่อนปรน ผ่อนผัน ปรับเปลี่ยน หรือยกเลิกเงื่อนไขต่าง ๆ ต้องมีการจัดทำสัญญาเพื่อนำเสนอคำขอพร้อมเหตุผลสนับสนุน เพื่อทำการปรับเปลี่ยนเงื่อนไขดังกล่าวไปยังผู้มีอำนาจอนุมัติในการปรับเปลี่ยนและจัดเก็บเอกสารเพื่อการตรวจสอบย้อนหลังจากผู้มีส่วนเกี่ยวข้องได้ โดยเงื่อนไขและข้อกำหนดของเงินกู้จะเป็นไปตามความเสี่ยงของเงินกู้ ยิ่งสินเชื่อที่มีความเสี่ยงสูงย่อมต้องมีเงื่อนไขและข้อกำหนดต่าง ๆ ที่เข้มงวดรัดกุมกว่า เช่น มีอัตราดอกเบี้ยเงินกู้ที่สูงกว่ากำหนดระยะเวลาผ่อนชำระชัดเจน มีค่างวดผ่อนชำระต่อเดือนที่สูงกว่าคนที่มีความเสี่ยงต่ำ เป็นต้น

<u>นโยบาย 5P (The 5P’s Policy)</u>

การวิเคราะห์สินเชื่อโดยใช้นโยบาย 5P [12] เป็นอีกวิธีหนึ่งที่สามารถใช้เป็นเครื่องมือที่ช่วยในการพิจารณาการให้สินเชื่อ หลักการพิจารณาประกอบด้วย

1. ตัวบุคคล (People)

ตัวบุคคลเป็นการพิจารณาบุคคลที่จะกู้เงิน พิจารณาว่ามีความรับผิดชอบเพียงใด มีความสำเร็จทางธุรกิจเพียงใด

2. วัตถุประสงค์ (Purpose)

วัตถุประสงค์ของการกู้เงินพิจารณาถึงความต้องการในการกู้เงินว่าจะทำเงินไปทำอะไร และดูแนวทางการใช้เงินตามความเหมาะสม

3. การชำระหนี้ (Payment)

การชำระหนี้เป็นการพิจารณาถึงความสามารถในการชำระหนี้ของธุรกิจนั้น ๆ ว่ามีสามารถชำระหนี้ได้ตรงตามระยะเวลาที่กำหนดหรือไม่ โดยพิจารณาความสามารถในการทำกำไรของธุรกิจ เพราะเงินที่จะต้องมาชำระหนี้เป็นเงินที่มาจากกำไรของธุรกิจหลังหักภาษี ไม่ใช่จากการขายทรัพย์สินหรือกู้ยืมและคำนวณแนวโน้มสภาพคล่องของธุรกิจที่อาจเกิดขึ้นในอนาคต

4. การป้องกัน (Protection)

การป้องกันเป็นการพิจารณาและหาแนวทางลดความเสี่ยงของการให้สินเชื่อ หากแผนการดำเนินธุรกิจของผู้กู้ไม่เป็นไปตามที่วางไว้ เช่น หากไม่สามารถดำเนินธุรกิจต่อไปได้ในอนาคตหรือขาดความสามารถในการบริหารจัดการธุรกิจจะมีบุคคลภายนอกหรือสินทรัพย์เข้ามารับผิดชอบหนี้หรือไม่

5. ความเสี่ยงภัย (Prospective)

ความเสี่ยงภัยเป็นการพิจารณาผลได้ผลเสียที่อาจจะเกิดขึ้นกับการให้สินเชื่อว่ามีความเหมาะสมหรือไม่ เช่น เปรียบเทียบอัตราดอกเบี้ยกับการเสี่ยงภัยในธุรกิจ ความยุ่งยากในการเรียกเก็บเงิน

<u>ขั้นตอนของงานสินเชื่อ</u>

ในการดำเนินงานสินเชื่อสามารถกำหนดขอบเขตของการดำเนินงานกว้าง ๆ ได้เป็น 2 ส่วน คือ ส่วนของการดำเนินการให้สินเชื่อและส่วนของการเรียกเก็บหนี้ การดำเนินการให้สินเชื่อมีขั้นตอน ดังนี้

1. แสวงหาและคัดเลือกลูกค้า
2. รับใบคำขอใช้บริการสินเชื่อ
3. สัมภาษณ์ลูกค้าผู้ต้องการสินเชื่อ
4 การวิเคราะห์สินเชื่อ เป็นการวิเคราะห์ทั้งในเชิงปริมาณและเชิงคุณภาพ
5. การกำหนดวงเงินสินเชื่อและการอนุมัติสินเชื่อ
6. การแจ้งให้ลูกค้าทราบการอนุมัติสินเชื่อ

7. การจ่ายสินเชื่อตามที่ได้รับอนุมัติ

8. การตรวจสอบการใช้เงินกู้

<u>การดำเนินการเก็บหนี้</u>

เป็นการปฏิบัติเพื่อให้ได้รับชำระหนี้คืน มีขั้นตอน ดังนี้

1. การแจ้งหนี้หรือการเตือน

2. การติดตามทวงหนี้

3. การใช้มาตรการที่รุนแรง

วิธีการเรียกเก็บหนี้ควรดำเนินการตามมาตรการตั้งแต่เบาไปหาหนักและควรแยกแยะให้ออกว่าลูกหนี้รายใดควรใช้มาตรการใด นอกจากนี้พนักงานที่ดำเนินการเรียกเก็บหนี้จะต้องมีข้อมูลที่ถูกต้องชัดเจนและมีกลยุทธ์ที่แยบยล โดยจะต้องมีการวางแผนและเตรียมการแก้ไขปัญหาที่เกิดขึ้นระหว่างการเรียกเก็บหนี้เป็นการล่วงหน้า เพื่อให้การเรียกเก็บหนี้บรรลุเป้าหมาย

<u>แนวคิดเกี่ยวกับความเสี่ยงและการบริหารความเสี่ยง</u>

ความเสี่ยง หมายถึง ความไม่แน่นอนที่จะเกิดเหตุการณ์ขึ้นจริง โดยมักมีผลลัพธ์การคาดการณ์ไว้ให้เป็นไปในทิศทางใดทางหนึ่ง หากมีการเบี่ยงเบนผลลัพธ์ไปในทางที่แตกต่างจากการคาดการณ์ไว้ ย่อมส่งผลกระทบกับกิจการที่จะก่อให้เกิดความล้มเหลวในด้านการบริหารงานโดยเฉพาะด้านการเงิน

การบริหารความเสี่ยง คือ กระบวนการที่ปฏิบัติโดยผู้บริหารและบุคลากรในองค์กรช่วยกันกำหนดกลยุทธ์และวิธีการดำเนินงาน โดยการบริหารความเสี่ยงออกแบบขึ้นมา เพื่อให้สามารถบ่งบอกถึงเหตุการณ์ที่อาจจะส่งผลกระทบต่อองค์กรได้ ทำให้องค์กรสามารถวางแผนจัดการความเสี่ยงให้อยู่ในระดับที่ยอมรับได้ เพื่อเสริมสร้างความมั่นใจในการช่วยกันบรรลุเป้าหมายขององค์กร [13] กระบวนการบริหารความเสี่ยงมีการกำหนดขั้นตอน และวิธีการในการบริหารความเสี่ยงเพื่อให้เป็นไปตามระบบ ซึ่งตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) ได้กำหนดขั้นตอนการบริหารความเสี่ยงไว้ 8 ขั้นตอน ประกอบด้วย

1. สภาพแวดล้อมภายในองค์กร (Internal Environment)

2. การกำหนดวัตถุประสงค์ (Objective Setting)

3. การบ่งชี้เหตุการณ์ (Event Identification)

4. การประเมินความเสี่ยง (Risk Assessment)

5. การตอบสนองความเสี่ยง (Risk Response)

6. กิจกรรมการควบคุม (Control Activities)

7. ข้อมูลและการติดต่อสื่อสาร (Information and Communication)

8. การติดตาม (Monitoring)

สาเหตุของการค้างชำระหนี้หรือไม่สามารถชำระหนี้ได้

การค้างชำระหนี้เป็นสิ่งที่หลีกเลี่ยงไม่ได้ที่ธนาคารต้องเผชิญในการดำเนินงาน เนื่องจากการให้สินเชื่อแก่ลูกค้าที่มาขอสินเชื่อ ถึงแม้ว่าจะมีการตรวจสอบคุณสมบัติของผู้กู้แล้วก็ตาม แต่ยังมีอีกหลายปัจจัยที่ส่งผลให้ลูกค้าไม่สามารถชำระหนี้คืนได้ตามระยะเวลาที่ระบุไว้ในสัญญา ธนาคารจึงต้องมีระบบการตรวจสอบติดตามหนี้หลังจากที่อนุมัติสินเชื่อไปแล้ว [14]

1. ปัจจัยที่เกิดจากภายใน เป็นปัจจัยที่เกิดขึ้นภายในธนาคาร สามารถควบคุมการเปลี่ยนแปลงที่เกิดขึ้นได้ ได้แก่

1.1 การเปลี่ยนแปลงของอัตราดอกเบี้ย

1.2 การประเมินราคาหลักทรัพย์ที่ไม่เหมาะสม

1.3 ระบบการติดตามและควบคุมหนี้ของธนาคาร

1.4 การอำนวยสินเชื่อของธนาคารที่ไม่มีการกลั่นกรองที่ดี

2. ปัจจัยที่เกิดขึ้นภายนอก เป็นปัจจัยที่ธนาคารไม่สามารถควบคุมได้ หากปัจจัยเปลี่ยนแปลงไปจะส่งผลกระทบต่อตัวลูกหนี้ ได้แก่

2.1 ภาวะเศรษฐกิจ เป็นปัญหาสำคัญของผู้ประกอบธุรกิจ หากเศรษฐกิจดีย่อมส่งผลดีต่อผลประกอบการของธุรกิจ หากเศรษฐกิจย่ำแย่จะส่งผลกระทบต่อผลประกอบการของธุรกิจ ทำให้รายได้น้อยลง เกิดการจ้างงานน้อยลง

2.2 นโยบายของรัฐบาล อาจมีการกำหนดราคาสินค้าต่าง ๆ ให้เป็นไปตามกลไกตลาด ส่งผลให้ค่าครองชีพมีการเปลี่ยนแปลง

2.3 ค่านิยมและเทคโนโลยี ปัจจุบันมีการเปลี่ยนแปลงที่รวดเร็วอยู่ตลอดเวลาอาจส่งผลให้ธุรกิจบางประเภทเปลี่ยนแปลงไปอย่างรวดเร็วด้วยเช่นกัน

2.4 ภัยธรรมชาติหรือเหตุการณ์ไม่คาดคิด เช่น ไฟไหม้ น้ำท่วม ส่งผลให้สินค้าหรือผลผลิตได้รับความเสียหาย

3. ปัจจัยที่เกิดจากตัวลูกหนี้เอง ได้แก่

3.1 ลูกหนี้นำเงินไปใช้ผิดวัตถุประสงค์

3.2 การย้ายถิ่นอาศัย การเปลี่ยนงาน การถูกเลิกจ้าง

3.3 ลูกหนี้ถึงแก่กรรมหรือเจ็บป่วยเรื้อรัง ทุพพลภาพ สภาพครอบครัวหย่าร้าง

3.4 ลูกหนี้ใช้จ่ายเงินฟุ่มเฟือย มีหนี้สินภายนอกมาก

3.5 การทุจริตของผู้บริหารในกิจการ

3.6 ลูกหนี้เจตนาไม่ยอมชำระหนี้หรือนำเงินไปชำระหนี้ภายนอกก่อนนำเงินไปชำระคืนแก่ธนาคาร

<u>แนวทางการแก้ปัญหาของสถาบันการเงิน</u>

1. เข้าใจปัญหาของลูกค้า ธนาคารต้องยกระดับความพึงพอใจรวมทั้งรักษาและดึงดูดให้ลูกค้าให้ได้มากที่สุด ซึ่งธนาคารจะต้องใส่ใจ กล่าวคือ การเข้าใจปัญหาที่ลูกค้ากำลังเผชิญและหาทางแก้ปัญหาที่ตอบโจทย์และสร้างคุณค่าได้มากที่สุด เช่น เข้าใจปัญหาเรื่องการรอคิวที่สาขานานของลูกค้า จึงแก้ปัญหาด้วยการให้บริการจองคิวล่วงหน้าผ่านแอปพลิเคชัน ซึ่งจะแจ้งเตือนเมื่อใกล้ถึงคิวรับบริการของลูกค้า ทำให้ลูกค้าสามารถวางแผนตารางประจำวันและทำธุรกรรมได้สะดวกยิ่งขึ้น [15]

2. ผนวก ESG เข้ากับการดำเนินธุรกิจ สถาบันการเงินและธนาคารพาณิชย์ต้องคำนึงถึงสิ่งแวดล้อม สังคม และธรรมาภิบาล (Environmental, Social and Governance: ESG) มากขึ้น โดยได้นำปัจจัยด้าน ESG มาผนวกเข้ากับกระบวนการทางธุรกิจ เช่น สินเชื่อพลังงานสะอาด ซึ่งเป็นการขยายโอกาสทางธุรกิจไปยังกลุ่มลูกค้าใหม่และผลิตภัณฑ์ใหม่ ๆ

3. เสริมขีดความสามารถด้านคลาวด์และการวิเคราะห์ข้อมูล ธนาคารต้องคำนึงถึงการเพิ่มขีดความสามารถด้วยระบบคลาวด์และการวิเคราะห์ข้อมูล เช่น การย้ายฐานข้อมูลเข้าสู่ระบบคลาวด์ ซึ่งจะช่วยลดความเสี่ยงด้านการบริหารจัดการข้อมูลและจัดการกับปริมาณของธุรกรรมที่เพิ่มขึ้นได้อย่างรวดเร็ว ส่งผลให้ลูกค้าเกิดความไว้วางใจและสร้างความน่าเชื่อถือของธนาคาร

4. จับมือพันธมิตรทางธุรกิจ การแสวงหาโอกาสจากความร่วมมือกับพันธมิตรก็เป็นเรื่องสำคัญที่จะช่วยเพิ่มศักยภาพของธุรกิจ ซึ่งปัจจุบันมีทั้งพันธมิตรที่เป็นสถาบันการเงินและที่ไม่ใช่สถาบันการเงิน เป็นโอกาสดีที่ทำให้ลูกค้าใช้บริการของธนาคารและพันธมิตรทำให้ลูกค้าสามารถเข้าถึงบริการทางการเงินที่ตอบโจทย์ได้ง่ายและสะดวกมากยิ่งขึ้น

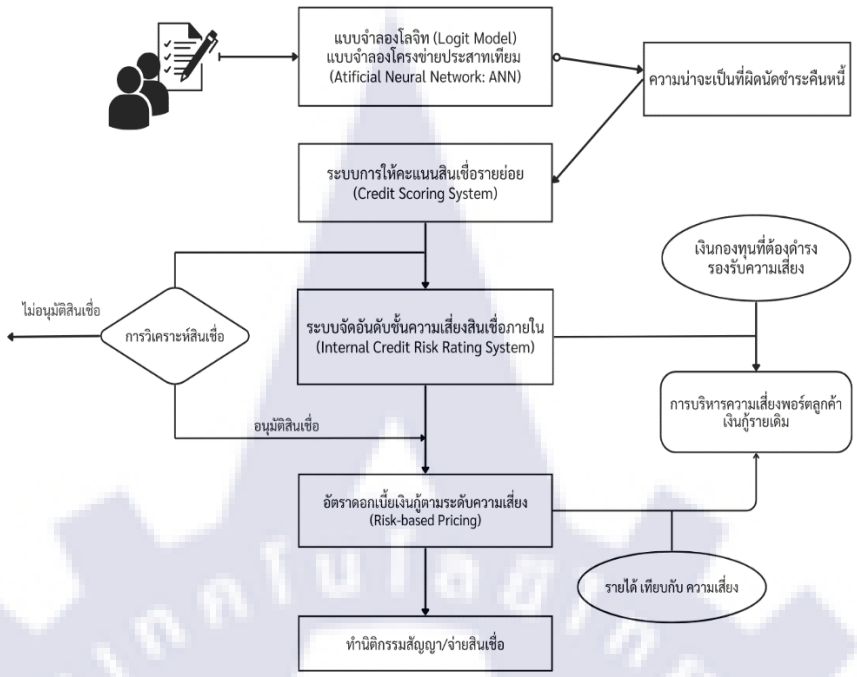
5. เพิ่มศักยภาพการเติบโตผ่านการควบรวมกิจการ ผู้บริหารอาจพิจารณาการควบรวมกิจการ (Mergers and Acquisitions: M&A) เพื่อให้ได้มาซึ่งสินค้า บริการและนวัตกรรมใหม่ ๆ ที่จะยิ่งช่วยทำให้ธุรกิจเติบโตได้เป็นทวีคูณมากกว่าที่บริษัทจะสามารถทำได้เองเพียงลำพัง

2.3 ระบบประเมินความเสี่ยงสินเชื่อรายย่อย

เนื่องด้วยธนาคารแห่งประเทศไทยมีการกำหนดให้สถาบันการเงินมีการพัฒนาระบบประเมินความเสี่ยงสินเชื่อรายย่อยด้วยการให้คะแนนและจัดอันดับชั้นความเสี่ยงตามมาตรฐานสากล (Statistical Model Based with Basel II Criteria) เพื่อใช้เป็นเครื่องมือบริหารความเสี่ยงของแต่ละสถาบันการเงินและสำหรับใช้พิจารณาอนุมัติ/ไม่อนุมัติสินเชื่อในกิจกรรมสินเชื่อของแต่ละสถาบัน

Credit Scoring Model (CSM) [16] เป็นเครื่องมือที่ช่วยประเมินความน่าเชื่อถือและความสามารถในการชำระหนี้ของผู้ขอสินเชื่อ สร้างขึ้นมาเพื่อช่วยลดปัญหาในการปล่อยกู้ให้แก่สถาบันการเงิน โดยเครื่องมือ Credit Scoring ถูกพัฒนาขึ้นจากข้อมูลธุรกรรมทางการเงินทั้งหมดของผู้ขอสินเชื่อในสถาบันการเงินนั้น ๆ เพื่อใช้ในการพิจารณาอนุมัติสินเชื่อ อนุมัติวงเงินสินเชื่อ โดยพิจารณาจากประวัติการชำระหนี้ของลูกค้าในอดีต เพื่อดูว่ามีการใช้จ่ายเป็นอย่างไร มีการกู้ยืมอะไรบ้าง จำนวนเท่าไร มีการชำระเงินตรงเวลาไหม มีผิดนัดชำระหรือไม่ เป็นต้น จากนั้นจึงนำข้อมูลที่ได้มาสร้างแบบจำลอง เพื่อทำการวิเคราะห์และคำนวณออกมาเป็นคะแนน ในส่วนของคะแนนที่ได้จะมีการแปลงให้อยู่ในลำดับชั้นของลูกหนี้ที่แตกต่างกัน ซึ่งสถาบันการเงินแต่ละแห่งมีหลักเกณฑ์และความเข้มงวดที่แตกต่างกัน หากมีเกณฑ์การพิจารณาที่เข้มงวดจนเกินไปอาจทำให้สูญเสียโอกาสทางธุรกิจได้ ในทางกลับกันหากสถาบันการเงินกำหนดเกณฑ์ที่ประณีประนอมมากเกินไปก็อาจทำให้ได้รับลูกค้าที่มีลักษณะกลุ่มเสี่ยงสูงเข้ามาได้เช่นกัน

แบบจำลองที่ใช้ในการวิเคราะห์ Credit Scoring ได้มีการพัฒนาอย่างต่อเนื่อง โดยในช่วงแรกของการพัฒนานั้นจะเป็นแบบจำลองทางคณิตศาสตร์และสถิติ (Statistical and mathematical programming) เช่น แบบจำลอง Linear Probability Logistic Regression Linear Discriminant Analysis (LDA) และ Linear and quadratic programming ในยุค Big Data ที่มีข้อมูลขนาดใหญ่และมีความหลากหลายของข้อมูลมากขึ้น ประกอบกับมีการพัฒนาเทคนิค Machine Learning ที่ใช้ในการวิเคราะห์ข้อมูล เช่น Random forest Gradient Boosting Neural Networks และ Support Vector Machines (SVM) ซึ่งทำให้ผลลัพธ์การพยากรณ์มีความแม่นยำเพิ่มขึ้นกว่าเดิมในอดีต อย่างไรก็ตามในปัจจุบันสถาบันการเงินส่วนใหญ่ยังคงนิยมใช้แบบจำลองทางคณิตศาสตร์และสถิติเป็นหลัก โดยเฉพาะ Logistic Regression เนื่องจาก Logistic Regression สามารถอธิบายความสัมพันธ์เชิงสาเหตุได้ง่ายกว่าและชัดเจนกว่า ผู้วิจัยเล็งเห็นว่าเป็นโอกาสที่ดีที่จะทบทวนและพิจารณาข้อมูลหรือตัวแปรเพิ่มเติม เพื่อช่วยเพิ่มประสิทธิภาพของเครื่องมือ Credit Scoring มากขึ้น อีกทั้งยังสามารถนำไปปรับใช้ในกระบวนการพิจารณาการอนุมัติสินเชื่อเพื่อเป็นประโยชน์ต่อลูกค้าและสถาบันการเงินอีกด้วย



รูปที่ 2.1 กรอบแนวคิดการพัฒนาระบบบริหารความเสี่ยงพอร์ตลูกค้าเงินกู้รายเดิมของ ธ.ก.ส.

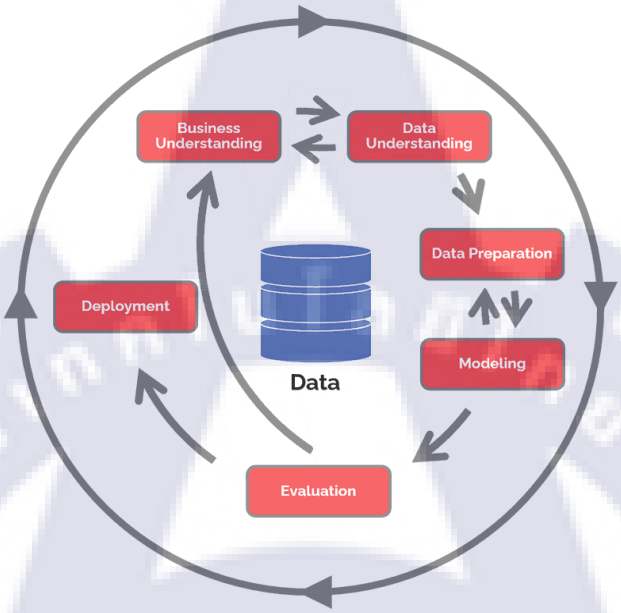
จากรูปที่ 2.1 แสดงกรอบแนวคิดการพัฒนาระบบบริหารความเสี่ยงพอร์ตลูกค้าเงินกู้รายเดิมของธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร (ธ.ก.ส.) โดยขั้นตอนการขอสินเชื่อเริ่มต้นจากลูกค้าผู้กู้รายเดิมเข้ามาแสดงความประสงค์ที่จะขอสินเชื่อเพิ่ม หลังจากนั้นเจ้าหน้าที่จะทำการรวบรวมข้อมูลและตรวจสอบข้อมูลของลูกค้า ต่อไปจะเป็นการวิเคราะห์หาความน่าจะเป็นที่จะผิดนัดชำระหนี้คืน เมื่อได้คะแนนของลูกค้าแล้วจะสามารถระบุความเสี่ยงของลูกค้าผ่านระบบการประเมินความเสี่ยงสินเชื่อรายย่อยและนำคะแนนไปจัดอันดับชั้นความเสี่ยงสินเชื่อภายในพอร์ตลูกค้าผู้กู้รายเดิม เพื่อดูความเสี่ยงของลูกค้า ระบบที่ใช้งานประกอบด้วย

1. การพยากรณ์ความน่าจะเป็นที่จะผิดนัดชำระหนี้คืนลูกค้าผู้กู้รายเดิมแต่ละราย
2. ระบบการจัดอันดับชั้นความเสี่ยงสินเชื่อพอร์ตลูกค้าเงินกู้รายเดิม
3. ระบบการกำหนดอัตราดอกเบี้ยเงินกู้ตามระดับความเสี่ยงลูกค้าผู้กู้รายเดิมแต่ละราย
4. ระบบการกำหนดโครงสร้างอัตราดอกเบี้ยเงินกู้ตามอันดับชั้นความเสี่ยงพอร์ตลูกค้าผู้กู้รายเดิม

หากลูกค้าอยู่ในเกณฑ์ที่ธนาคารกำหนดไว้จะถือว่าผ่านการประเมิน โดยลูกค้าจะได้รับอัตราดอกเบี้ยสินเชื่อตามลำดับชั้นความเสี่ยงของลูกค้าแต่ละราย ทั้งนี้การพิจารณาอนุมัติสินเชื่อขึ้นอยู่กับนโยบายและหลักเกณฑ์ของแต่ละสถาบันการเงิน

2.4 กระบวนการวิเคราะห์ข้อมูล

Cross-Industry Standard Process for Data Mining: CRISP-DM เป็นกระบวนการวิเคราะห์ข้อมูลมาตรฐานที่ได้รับความนิยมใช้งาน ซึ่งเป็นการวิเคราะห์ข้อมูลด้าน Data Mining ประกอบด้วย 6 ขั้นตอน ดังนี้



รูปที่ 2.2 กระบวนการทำงาน CRISP-DM

1. การทำความเข้าใจธุรกิจ (Business Understanding)

เป็นขั้นตอนแรกในกระบวนการ คือ การเข้าใจจุดประสงค์ในการทำธุรกิจทำความเข้าใจธุรกิจ เข้าใจปัญหาและวัตถุประสงค์ของโครงการ จากนั้นทำการแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล ดูว่าจะประยุกต์ใช้การวิเคราะห์ข้อมูลมาใช้งานจริงได้อย่างไร เช่น การคาดการณ์ว่าลูกค้าคนใดควรได้รับสินเชื่อที่ขอกู้มาพร้อมทั้งระบุผลลัพธ์ที่ต้องการและวางแผนในการดำเนินงาน

2. การทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูลที่เกี่ยวข้องให้ปริมาณข้อมูลเพียงพอเหมาะสม หลังจากนั้นจะนำข้อมูลมาตรวจสอบคุณภาพ ดูความถูกต้องของข้อมูล พิจารณาว่ามีข้อมูลอะไรบ้างที่จะนำมาวิเคราะห์เพื่อตอบโจทย์ เลือกข้อมูลที่เก็บรวบรวมมาที่จะนำไปใช้วิเคราะห์และทำความเข้าใจข้อมูลที่ได้มาเพื่อใช้ไปวิเคราะห์ในขั้นตอนต่อไป

วิธีการเลือกกลุ่มตัวอย่าง มี 2 แบบใหญ่ [17]

2.1 การเลือกกลุ่มตัวอย่างที่ไม่เป็นไปตามโอกาสทางสถิติ (Non-Probability Sampling) ประกอบด้วย 4 วิธี

(1) การสุ่มแบบบังเอิญ (Accidental Sampling) เป็นการเก็บข้อมูลให้ครบตามต้องการโดยไม่มีกฎเกณฑ์ที่แน่นอน

(2) การสุ่มแบบกำหนดโควต้า (Quota Sampling) จะเป็นการกำหนดกลุ่มย่อยตามที่ต้องการโดยอาศัยสัดส่วนขององค์ประกอบกลุ่มประชากรตาม เพศ การศึกษา

(3) การสุ่มแบบเจาะจง (Purposive Sampling) เป็นการเลือกกลุ่มที่ผู้วิจัยใช้เหตุผลในการเลือกเพื่อความเหมาะสมในการวิจัย

(4) การสุ่มตามความสะดวก (Convenience Sampling) เป็นการเลือกกลุ่มตัวอย่างตามความสะดวกในเรื่องที่ศึกษา

2.2 การเลือกกลุ่มตัวอย่างที่เป็นไปตามโอกาสทางสถิติ (Probability Sampling) มี 4 วิธี ประกอบด้วย

(1) การสุ่มอย่างง่าย (Simple Random Sampling) เช่น วิธีการจับฉลาก (Lottery) และวิธีการใช้ตารางเลขสุ่ม (Random Table)

(2) การสุ่มอย่างมีระบบ (Systematic Random Sampling) เช่น การเรียงลำดับบัญชีรายชื่อหาช่วงของการเลือกตัวอย่าง โดยใช้ประชากรทั้งหมดหารด้วยขนาดของกลุ่มตัวอย่าง

(3) การสุ่มแบบเป็นชั้นภูมิ (Stratified Random Sampling) มีการจัดแบ่งประชากรเป็นกลุ่มหรือชั้นย่อย ๆ ก่อน แล้วเลือกสุ่มตัวอย่างตามสัดส่วน (Proportional) ในแต่ละชั้น จากนั้นจึงใช้การสุ่มอย่างง่าย เช่น แบ่งนักศึกษาตามคณะต่าง ๆ หาขนาดกลุ่มตัวอย่าง จากนั้นเทียบสัดส่วนตามขนาดแล้วจับฉลาก เป็นต้น

(4) การสุ่มแบบหลายขั้นตอน (Multi-stage Sampling) ใช้การสุ่มหลายแบบ เช่น แบ่งเป็นกลุ่ม แล้วแบ่งเป็นชั้นภูมิ แล้วสุ่มอย่างง่าย

3. การเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมข้อมูล เป็นกระบวนการที่ใช้เวลานานที่สุด เพราะข้อมูลต้องมีความถูกต้อง ครบถ้วนเพื่อสามารถนำไปใช้งานได้ โดยเป็นการเตรียมข้อมูลที่ยังเป็นข้อมูลดิบที่ได้มาจากการรวบรวมมาทำให้กลายเป็นข้อมูลที่สมบูรณ์ ซึ่งขั้นตอนนี้เป็นการทำความเข้าใจข้อมูล มีตัวแปรอะไรบ้าง โดยแบ่งเป็น 3 ขั้นตอนย่อย ดังนี้

3.1 Data Selection การคัดเลือกข้อมูลเป็นการเลือกข้อมูลหรือตัวแปรที่จะนำมาใช้วิเคราะห์ข้อมูล

3.2 Data Cleaning การแก้ไขข้อมูลที่ผิดพลาด เป็นการกรองข้อมูลเอาส่วนที่ไม่ได้ใช้งานออก เช่น แก้ไขข้อมูลที่เป็นค่าว่าง (Missing Value)

3.3 Data Integration จะใช้ในกรณีที่มีการนำข้อมูลจากหลายๆ แหล่งมาใช้ร่วมกัน

3.4 Data Reduction เป็นการลดขนาดของข้อมูลลงให้เหลือเพียงข้อมูลที่จำเป็น

3.5 Data Transformation การแปลงข้อมูล เป็นการจัดรูปของข้อมูลให้สามารถนำไปวิเคราะห์ เช่น การแปลงข้อมูลประเภท Text ให้อยู่ในรูปแบบของตัวเลข

4. Modeling การสร้างโมเดล

ในขั้นตอนนี้จะเป็นการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining โดยจะเลือกและทดลองสร้างหลาย ๆ โมเดลที่คาดว่าสามารถแก้ไขปัญหาที่ต้องการได้ จากนั้นค่อย ๆ ปรับค่าพารามิเตอร์ในแต่ละโมเดล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้ เพื่อให้ได้โมเดลที่เหมาะสมที่สุดมาใช้ในการแก้ไขปัญหา เป็นการประยุกต์ใช้เทคนิควิเคราะห์ข้อมูล โดยวิธีการวิเคราะห์ข้อมูลแบ่งได้ 2 แบบ ดังนี้

4.1 Unsupervised Learning เป็นการวิเคราะห์ข้อมูลแบบไม่มีคำตอบมาให้ เช่น การจัดกลุ่มแบบ K-mean

4.2 Supervised Learning เป็นการวิเคราะห์ข้อมูลที่มีคำตอบให้มาแล้ว เช่น คาดการณ์ว่าลูกค้าคนใดบ้างจะชำระหนี้ตามกำหนดระยะเวลา

5. Evaluation การวัดผล

เป็นการวัดผลประสิทธิภาพของโมเดล วัดผลว่าโมเดลมีประสิทธิภาพเพียงพอหรือไม่ และเพื่อเปรียบเทียบแต่ละเทคนิคว่าเทคนิคใดดีกว่ากัน

6. Deployment นำไปใช้งาน

เป็นการนำโมเดลที่เหมาะสมที่สุดไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหาที่ต้องการและเป็นตัวช่วยในการตัดสินใจทางธุรกิจ

2.5 ทฤษฎีการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine Learning) [18] เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) เป็นกระบวนการที่ทำให้คอมพิวเตอร์มีความสามารถในการเรียนรู้ด้วยตนเอง หลักการทำงานเกี่ยวข้องกับการศึกษาและการพัฒนาอัลกอริทึมที่สามารถเรียนรู้และปรับตัวตามข้อมูลที่ได้รับโดยอาศัยแบบจำลอง (Model) ที่สร้างจากชุดข้อมูลที่มีความสัมพันธ์กันและสามารถนำไปทำนายหรือใช้ในการตัดสินใจภายหลัง ประเภทการเรียนรู้ของเครื่องได้ผ่านการสังเคราะห์รูปแบบการเรียนรู้ออกเป็น 3 รูปแบบ ดังนี้

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการเรียนรู้โดยอาศัยข้อมูลที่ป้อนเข้าไปเก็บไว้เป็นตัวอย่าง เพื่อให้คอมพิวเตอร์ใช้ในการเปรียบเทียบกับข้อมูลที่เข้ามาใหม่แล้ว

ทำนายหรือจัดหมวดหมู่ที่มีความเหมือนกันมากที่สุดให้อยู่ด้วยกัน ผลลัพธ์ที่ได้ คือ การจัดหมวดหมู่ (Classification) และการวิเคราะห์การถดถอย (Regression) ตัวอย่างที่เห็นได้ชัดของ Machine Learning ในกลุ่ม Supervised Learning ที่ถูกนำมาประยุกต์ใช้งานในเชิงธุรกิจ คือ การคำนวณราคาบ้านหรือการวิเคราะห์ผลฟุตบอล

การเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่เป็นที่รู้จัก เช่น Classification และ Regression ซึ่งวิธี Classification เป็นการทำนายผลลัพธ์ที่มีความต่อเนื่อง เช่น ผลลัพธ์ที่ได้จากการทำนายออกมาเป็นตัวเลข ส่วนวิธี Regression เป็นการทำนายผลลัพธ์ที่ข้อมูลไม่ต่อเนื่อง ไม่ใช่ตัวเลข เช่น ผลลัพธ์ที่ได้จากการทำนายเป็นสีเขียว สีแดง สีส้ม เป็นต้น

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้ที่ไม่มีข้อมูลตัวอย่างที่ใช้บอกว่าข้อมูลนั้นคืออะไร แต่จะเรียนรู้จากการหาความสัมพันธ์จากข้อมูลนำเข้า (Input) โดยพิจารณาจากรูปแบบ (Patterns) หรือโครงสร้างของข้อมูล (Data structure) แล้วนำมาจัดเป็นกลุ่ม (Cluster) บนพื้นฐานของความเหมือน (Similarities) และความแตกต่าง (Differences) ระหว่างรูปแบบของข้อมูลนำเข้า (Input Patterns) ตัวอย่างที่เห็นได้ชัดของ Machine Learning ในกลุ่ม Unsupervised Learning ที่ถูกนำมาประยุกต์ใช้งานในเชิงธุรกิจ คือ ระบบแนะนำผลิตภัณฑ์ เช่น การแนะนำคลิปวิดีโอใน YouTube ที่ทำการแบ่งหมวดหมู่ของคลิปวีดีโอต่าง ๆ

การเรียนรู้แบบไม่มีผู้สอนสามารถแบ่งเป็น 2 ประเภทหลัก ๆ คือ Clustering และ Dimension Reduction โดย Clustering คือ การจัดกลุ่มข้อมูลซึ่งในบางครั้งมีข้อมูลจำนวนมากที่ไม่สามารถใส่ Label ได้หมด แต่รู้ว่าข้อมูลนั้นประกอบด้วยจำนวน Class อะไรบ้าง อาจจะระบุให้มีการแบ่งข้อมูลออกเป็นจำนวน Class ตามที่ต้องการ จากนั้นค่อยมาระบุว่าแต่ละกลุ่มที่พบคืออะไร ส่วน Dimension reduction คือ การลดจำนวนมิติเพื่อบีบอัดข้อมูลเป็นกลไกที่ทำให้ไม่จำเป็นต้องเก็บข้อมูลไว้ครบ แต่ก็ยังสามารถจำแนกข้อมูลได้ เช่น มีข้อมูลสามมิติของข้อมูลหลาย ๆ คลาส สิ่งที่ต้องการ คือ การลดข้อมูลให้เหลือสองมิติและยังสามารถนำไปแยกประเภทคลาสได้ดีเท่าเดิม

3. การเรียนรู้แบบเสริมกำลัง (Reinforcement learning) เป็นการเรียนรู้แบบไม่มีผู้สอน แต่จะมีปฏิสัมพันธ์กับสิ่งแวดล้อม ทำให้เกิดการกระตุ้นในการสร้างแบบจำลอง เพื่อตอบสนองว่าจะต้องทำอย่างไรต่อไป ตัวอย่างที่เห็นได้ชัดของMachine Learning ในกลุ่ม Reinforcement Learning ที่ถูกนำมาประยุกต์ใช้งานในเชิงธุรกิจ คือ AlphaGo ที่สามารถเล่นเกมโกะให้ชนะผู้เล่นระดับโลก ระบบการจัดการ Portfolio ให้ตัดสินใจเลือกอัตราส่วนของสินทรัพย์

การเรียนรู้ของเครื่อง (Machine Learning) เป็นกระบวนการทำงานที่ต้องการให้คอมพิวเตอร์มีความสามารถเรียนรู้สิ่งต่างๆ ด้วยตัวเองหรือมีการทำงานเป็นไปอย่างอัตโนมัติ โดยมีรูปแบบการเรียนรู้แตกต่างกันไป การเลือกประเภทการเรียนรู้ของเครื่องจะขึ้นอยู่กับวัตถุประสงค์ของการนำไปใช้งาน

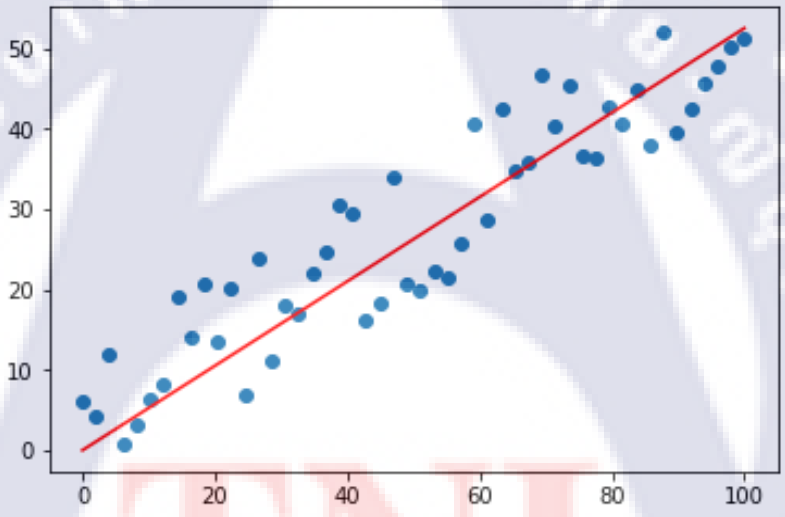
2.6 การวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ

1. การถดถอยโลจิสติก (Logistic Regression: LR) คือ การวิเคราะห์ถดถอยเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ซึ่งได้แก่ ตัวประมาณการ (Predictor, X) และตัวตอบสนอง (Response, Y) โดยเป็นความสัมพันธ์แบบเชิงเส้น (Linear) โดยที่ขั้นตอนการทำ Regression ต้องมีการเก็บจำนวนตัวอย่างที่จะใช้ศึกษาเป็นจำนวนมากพอเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร และในหลายๆ ครั้ง เพื่อนำมาหาสมการความสัมพันธ์

$$y = ax + b$$

เมื่อ $a = \text{slope}$

$$b = y - \text{int ercept} \qquad (2.1)$$



รูปที่ 2.3 ตัวอย่างการวิเคราะห์การถดถอยโลจิสติก

จากรูปที่ 2.3 จะเห็นได้ว่าในแผนภาพมีหลายจุด นั้นก็คือจุดที่บ่งบอกว่า เมื่อมีค่าเปลี่ยนแปลงไปจะทำให้ค่า y เปลี่ยนแปลงตามไป ดังนั้น จำนวนของจุดจึงมีส่วนสำคัญในการทำการวิเคราะห์การถดถอย การวิเคราะห์โลจิสติกมีเป้าหมายก็เพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ซึ่งก็คือตัวแปรเกณฑ์ โดยอาศัยสมการโลจิสติกที่สร้างขึ้นจากชุดตัวแปรทำนาย (x's) ที่มีข้อมูลเป็นตัวแปรที่อยู่ในระดับช่วง Interval Scale หากเป็นข้อมูลเชิงกลุ่มจะต้องมีการแปลงเป็นตัวแปรทวิที่มีค่า 0 กับ 1 ก่อน โดยที่ระหว่างตัวแปรทำนายจะต้องมีความสัมพันธ์กันต่ำ โดยใช้เกณฑ์ค่า r หรือความสัมพันธ์ ไม่เกิน .65 ถ้าใช้เกณฑ์ของ Burns and Grove [19] หรือถ้าใช้เกณฑ์ของ Stevens [20]

ค่า r ไม่เกิน .80 ซึ่งถ้าหากเกิดความสัมพันธ์กันสูงจะทำให้เกิดปัญหา Multicollinearity และในการวิเคราะห์จะต้องใช้ขนาดตัวอย่างหรือ n มากกว่าหรือเท่ากับ 30 เท่า ของจำนวนตัวแปรทำนาย

2. โครงข่ายประสาทเทียม (Neural Network) เป็นการจำลองการทำงานของสมองมนุษย์ด้วยโปรแกรมคอมพิวเตอร์ เป็นแนวความคิดที่ต้องการให้คอมพิวเตอร์มีความชาญฉลาดในการเรียนรู้และฝึกฝนได้ สามารถนำความรู้และทักษะไปแก้ปัญหาต่าง ๆ

2.1 ข้อมูลป้อนเข้า (Input) เป็นข้อมูลที่เป็นตัวเลข หากเป็นข้อมูลเชิงคุณภาพต้องแปลงให้อยู่ในรูปเชิงปริมาณที่โครงข่ายประสาทเทียมยอมรับได้

2.2 ข้อมูลส่งออก (Output) คือ ผลลัพธ์ที่เกิดขึ้นจริง (Actual Output) จากกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม

2.3 ค่าน้ำหนัก (Weights) คือ สิ่งที่ได้มาจากการเรียนรู้หรือค่าความรู้ (Knowledge) โดยค่านี้จะเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่น ๆ ที่อยู่ในรูปแบบเดียวกันต่อไป

2.4 ฟังก์ชันผลรวม (Summation Function: S) เป็นผลรวมของข้อมูลป้อนเข้า (Ai) และค่าน้ำหนัก (Wi)

2.5 ฟังก์ชันการแปลง (Transfer Function) เป็นการคำนวณการจาลองการทำงาของโครงข่ายประสาทเทียม เช่น ซิกมอยด์ฟังก์ชัน (Sigmoid Function) ฟังก์ชันไฮเปอร์โบลิกแทนเจนต์ (Hyperbolic Tangent Function) เป็นต้น

3. ต้นไม้ตัดสินใจ (Decision Tree) เป็นโมเดลทางคณิตศาสตร์ที่ใช้ในการทำนายประเภทของวัตถุ โดยจะพิจารณาจากลักษณะของวัตถุ เป็นการตัดสินใจเป็นแบบแผนผังต้นไม้ประกอบด้วยส่วนที่เป็นโหนดภายใน (Inner Node) หรือที่เรียกว่าโหนดของต้นไม้ ทำหน้าที่แทนตัวแปรต่าง ๆ ที่ใช้ในการตัดสินใจ ส่วนที่เป็นกิ่ง (Branch) ใช้แทนค่าที่เป็นได้ของตัวแปรและส่วนที่เป็นโหนดใบ (Leaf Node) ใช้แทนค่าคำตอบของตัดสินใจหรือการจำแนก

โหนดภายใน (Internal Node) คือ คุณลักษณะต่าง ๆ ของข้อมูล ซึ่งเมื่อข้อมูลใด ๆ ตกลงมาที่โหนดจะใช้คุณลักษณะนี้เป็นตัวตัดสินใจว่าข้อมูลจะไปในทิศทางใด โดยโหนดภายในที่เป็นจุดเริ่มต้นของต้นไม้ เรียกว่า โหนดราก

กิ่ง (Branch, Link) เป็นค่าของคุณลักษณะในโหนดภายในที่แตกกิ่งนี้ออกมา ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าของคุณลักษณะในโหนดภายในนั้น

โหนดใบ (Leaf Node) คือ กลุ่มต่างๆ ซึ่งเป็นผลลัพธ์ในการจำแนกประเภทข้อมูล

ต้นไม้ตัดสินใจเป็นเทคนิคที่ใช้ได้กับข้อมูลที่มีค่าต่อเนื่อง (Continuous Values) และข้อมูลที่มีค่าไม่ต่อเนื่อง (Discrete Values) โดยต้นไม้ตัดสินใจที่เป็นโหนดใบจะแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง เรียกว่าต้นไม้ตัดสินใจแบบจำแนก (Classification Trees) และต้นไม้ตัดสินใจที่โหนดใบเป็นข้อมูลต่อเนื่อง (Continuous Values) เรียกว่าต้นไม้ตัดสินใจแบบถดถอย (Regression Trees)

ต้นไม้ตัดสินใจจะมีโครงสร้างการตัดสินใจเป็นแบบต้นไม้ที่ประกอบด้วย ส่วนที่โหนด กิ่ง และใบ ซึ่งวิธีการสร้างต้นไม้นั้นมีขั้นตอนวิธีให้เลือกหลากหลาย เช่น ขั้นตอนวิธี CART ID3 C4.5 SLIQ และ SPRINT เป็นต้น โดยแต่ละวิธีจะมีความแตกต่างของเกณฑ์ที่ใช้พิจารณาเลือกโหนดและลักษณะการแตกกิ่งที่เหมาะสม

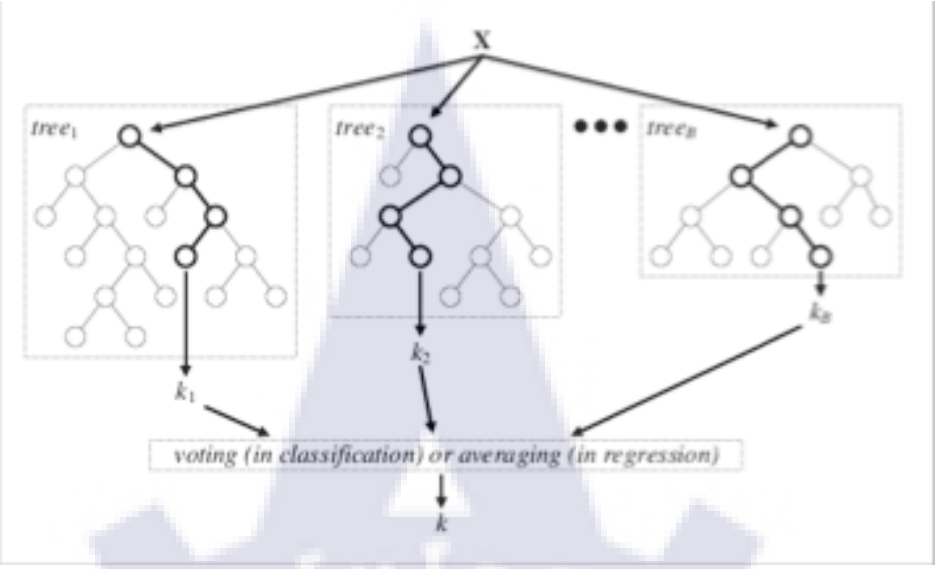
ขั้นตอนของการสร้างต้นไม้แบ่งเป็น 2 ประเด็นหลัก ดังนี้

3.1 การตัดสินใจว่าจะแตกกิ่งหรือแตกข้อมูลเป็นอย่างไร ซึ่งจะเป็นเรื่องเกี่ยวกับการกำหนดเงื่อนไขการแตกกิ่งข้อมูล เช่น แตกเป็น 2 กิ่ง 3 กิ่ง หรือมากกว่านั้น เป็นต้น ซึ่งสามารถทำได้โดยใช้ชนิดของข้อมูลเป็นเงื่อนไขว่าจะให้ข้อมูลเป็นแบบไหน เช่น ข้อมูลเป็นนามบัญญัติหรือกลุ่ม (Category Data) หรือข้อมูลเป็นตัวเลขต่อเนื่อง เป็นต้น และเป็นเรื่องเกี่ยวกับการเลือกจุดที่เหมาะสมสำหรับการแตกกิ่ง นั่นคือการทดสอบว่าแอททริบิวต์หรือตัวแปรใดเหมาะสมที่จะเป็นโหนดเพื่อแตกกิ่งของปัญหามากที่สุด

3.2 การตัดสินใจว่าจะหยุดแตกกิ่งเมื่อไหร่ เช่น หยุดเมื่อกิ่งนั้นมีจำนวนน้อยกว่าค่าที่กำหนดหรือหยุดเมื่อค่าเกณฑ์การทดสอบของการแตกกิ่งไม่ดีขึ้น เป็นต้น โดยขั้นตอนวิธีสร้างต้นไม้แต่ละวิธีจะมีกลไกการทำงานที่เป็นแบบวนซ้ำเลือกแอทริบิวต์ที่ให้ค่าเกณฑ์การทดสอบที่เหมาะสมที่สุดเพื่อแตกกิ่งลงไปเรื่อย ๆ จนพบเงื่อนไขการหยุดแตกกิ่ง

4. การวิเคราะห์ป่าแบบสุ่ม (Random Forest) เป็นวิธีที่มีการพัฒนามากขึ้นในการทำเหมืองข้อมูลและค้นหาความรู้ (Knowledge Discovery) มีคุณสมบัติในการจำแนกข้อมูลที่ประกอบไปด้วยชุดข้อมูลจำแนกโครงสร้างแบบต้นไม้ใช้สำหรับการจำแนกและการถดถอยของปัญหาสามารถใช้ในการทำนายตัวแปรต่อเนื่องและสามารถใช้ในการประมาณความน่าจะเป็นที่เกิดขึ้น

แนวคิดของ Random Forest [21] คือ การสร้างตัวแบบด้วยวิธีการ Decision Tree ซึ่งเป็นเทคนิคที่ให้ผลลัพธ์ในลักษณะเป็นโครงสร้างของต้นไม้ ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) แต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัวทดสอบกิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบและใบ (Leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (Class) โดยผลลัพธ์ที่ได้จากการทำนายแนวคิดของ Random Forest คือ การสร้างตัวแบบด้วยวิธีการ Decision Tree ขึ้นมาหลาย ๆ ตัวแบบ โดยวิธีการสุ่มตัวแปรแล้วนำผลที่ได้แต่ละตัวแบบมารวมกันโดยวิธีการจำแนกข้อมูลจะนับผลที่มีจำนวนซ้ำกันมากที่สุดและเลือกออกมาเป็นผลลัพธ์สุดท้ายและสำหรับการถดถอยจะเป็นการเฉลี่ยผลของต้นไม้ทุก ๆ ต้น ดังแสดงในภาพที่ 2.4



รูปที่ 2.4 แนวคิดของ Random Forest [22]

ข้อดีของวิธีป่าแบบสุ่ม คือ ให้ผลการทำนายที่แม่นยาและเกิดปัญหา Overfitting น้อย สามารถทำงานได้อย่างมีประสิทธิภาพบนฐานข้อมูลขนาดใหญ่แต่การวิเคราะห์ในทางทฤษฎียังเป็นเรื่องที่ยากและหากใช้สำหรับการทำนายจะให้ผลการทำนายไม่เกินช่วงของข้อมูลฝึกฝน (Training Data)

Random Forest เป็นเครื่องที่นิยมใช้มากและใช้ได้ดีในหลาย ๆ ปัญหา ทั้งแบบ Regression และ Classification โดย Random Forest มีการพัฒนาต่อยอดมาจากโมเดล Decision Tree ต่างกันที่ Random Forest เป็นการเพิ่มจำนวนต้นไม้ (Tree) เป็นหลาย ๆ ต้น ทำให้ประสิทธิภาพการทำงานและพยากรณ์สูงขึ้น โดยมีหลักการทำงาน คือ แบ่งข้อมูลออกเป็นต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ ต้น โดยแต่ละต้นจะได้รับคุณลักษณะ (Feature) และข้อมูล (Data) ที่ไม่เหมือนกันทั้งหมด เพื่อทำให้ได้ต้นไม้ที่มีความหลากหลายและมีความอิสระต่อกันมากขึ้น

การทำงานของ Random Forest มีดังนี้

4.1 ทำการสุ่มเลือก Feature และ Data จากชุดข้อมูลทั้งหมดที่มี

4.2 สร้างต้นไม้ตัดสินใจจากข้อมูลตัวอย่างแต่ละชุดและหาค่าพยากรณ์ต้นไม้แต่ละต้น

4.3 เลือกจำนวนต้นไม้ตัดสินใจที่ต้องการ จากนั้นทำซ้ำในขั้นตอน 1 และ 2 อีกรอบ

4.4 หาค่าพยากรณ์ โดยค่าพยากรณ์ที่ได้จะเป็นการให้ต้นไม้ตัดสินใจแต่ละต้นหาค่าพยากรณ์แต่ละต้น ในกรณีที่เป็นปัญหาเพื่อการจำแนก (Classification) จะใช้วิธีผลโหวตมากที่สุด (Majority Vote) โดยค่าพยากรณ์ของต้นไม้ตัดสินใจที่ได้รับค่าผลโหวตมากที่สุดจะถูกเลือกให้เป็น

ค่าพยากรณ์ แต่ถ้าเป็นปัญหาวิเคราะห์การถดถอย (Regression) จะใช้วิธีคำนวณหาค่าเฉลี่ย (Mean) โดยเอาค่าพยากรณ์ของทุกต้นไม้ตัดสินใจมาคำนวณหาค่าเฉลี่ยเพื่อแสดงเป็นค่าพยากรณ์

Random Forest ใช้ได้ทั้งกับปัญหา Classification และ Regression และใช้ได้ทั้งกับข้อมูลมีโครงสร้าง (Structured) และไม่มีโครงสร้าง ข้อมูลมีโครงสร้าง เช่น ข้อมูลจัดเก็บเป็นตาราง หรือคอลัมน์แล้ว เป็นต้น ส่วนข้อมูลไม่มีโครงสร้าง เช่น รูปภาพ หรือข้อความ เป็นต้น

2.7 แนวคิดการวัดค่าความถูกต้องของเครื่องมือ

ในการวัดค่าความถูกต้องของเครื่องมือนั้น งานวิจัยนี้จะใช้การเปรียบเทียบแต่ละวิธีด้วยค่าต่าง ๆ ได้แก่ ค่าความแม่นในการทำนาย ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย โดยงานวิจัยจะทำการประเมินประสิทธิภาพของแต่ละวิธีโดยใช้ Confusion Matrix และนำมาเปรียบเทียบประสิทธิภาพของแต่ละโมเดล

Confusion Matrix เป็นเครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนายหรือ Prediction โดยทำนายจากโมเดลที่สร้างขึ้นใน Machine Learning โดยมีไอเดียจากการวัดว่าสิ่งที่เราคิดหรือที่โมเดลที่ทำนายกับสิ่งที่เกิดขึ้นจริง มีสัดส่วนเป็นอย่างไร จำแนกมี 2 ประเภทคือ ทายถูก (Positive) และ ทายผิด (Negative)

ผลการทำนายทั้งหมดที่เป็นไปได้จะมีทั้งหมด 4 ดังนี้

True Positive (TP) คือ สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นจริง ในกรณีทำนายว่า จริง และสิ่งที่เกิดขึ้นก็คือ จริง

True Negative (TN) คือ สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณีทำนายว่า ไม่จริง และสิ่งที่เกิดขึ้นก็คือ ไม่จริง

False Positive (FP) คือ สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น ในกรณีทำนายว่า จริง แต่สิ่งที่เกิดขึ้นคือ ไม่จริง

False Negative (FN) คือ สิ่งที่ทำนายไม่ตรงกับที่เกิดขึ้นจริง ในกรณีทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้นคือ จริง

TP TN FP FN ในตารางจะแทนด้วยค่าความถี่ซึ่งสามารถนำค่าเหล่านี้

	Actually Positive	Actually Negative
Predicted Positive	True Positives (TP)	False Positives (FP)
Predicted Negative	False Negatives (FN)	True Negatives (TN)

รูปที่ 2.5 ตาราง Confusion Matrix

โดยสามารถนำมาคำนวณเป็นค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความอ่อนไหว (Recall) และค่าประสิทธิภาพ (F1-score) ได้ดังนี้

1. ค่าความถูกต้อง (Accuracy) คือ การคำนวณจำนวนคำตอบที่สามารถพยากรณ์ถูกต้องเทียบกับจำนวนคำตอบทั้งหมดที่นำไปให้พยากรณ์ ซึ่งนำมาประยุกต์ใช้กับงานวิจัยฉบับนี้โดยการวัดค่าความถูกต้องของตัวพยากรณ์

$$\%Accuracy = 100 - \%Error$$

โดยที่

$$\text{Relative Error} = \left| \frac{x_{mea} - x_1}{x_1} \right| \qquad (2.2)$$

$$\%Error = \text{Relative Error} \times 100$$

เมื่อ X_{mea} คือ ค่าที่ได้จากการวัด (Measure Value)

X_t คือ ค่าจริง (True Value)

2. ค่าความแม่นยำ (Precision) คือ การคำนวณความแม่นยำหรือจำนวนที่ทำนายถูกจากข้อมูลที่ทำนายเป็นคลาสที่สนใจอยู่ โดยสามารถนำมาประยุกต์ใช้กับงานวิจัยฉบับนี้ โดยการวัดค่าความแม่นยำของตัวพยากรณ์ที่ทำนายผลออกมาเป็นโรคไตเรื้อรังหรือไม่

$$\text{Precision} = \left| \frac{x_i - x_m}{x_m} \right| \quad (2.3)$$

$$x_m = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (2.4)$$

โดยที่ x_m คือค่าเฉลี่ยของการวัด

x_i คือค่าการวัดแต่ละครั้ง

n คือจำนวนครั้งของการวัด

$\sum_{i=1}^{n} x_i$ คือผลรวมค่าของการวัดทั้งหมด

3. ค่าความอ่อนไหว (Recall) คือ ค่าความอ่อนไหว หรือจำนวนข้อมูลที่ทำนายถูกต่อข้อมูลที่ทำนายถูกรวมกับข้อมูลที่ทำนายผิด

4. ค่าประสิทธิภาพ (F1-Score) หรือการวัดประสิทธิภาพของโมเดล หาได้โดยการใช้ค่าน้ำหนักเฉลี่ยของค่าความแม่นยำและค่าความอ่อนไหวของแต่ละกลุ่มเป้าหมาย

5. ค่าความแม่นในการทำนาย คือ การแสดงการวัดที่ได้มีความแม่นยำในรูปแบบอัตราส่วนโดยคิดเป็นร้อยละ โดยนำค่าจากตารางเมทริกซ์ความสับสนมาแทนในสมการ

Accuracy = จำนวนข้อมูลที่จำแนกถูกว่าคำตอบเป็นบวกและลบ x 100 %

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2.5)$$

6. ค่าเฉลี่ยความคลาดเคลื่อนของการทำนาย (RMSE) คือ ค่าเฉลี่ยของความคลาดเคลื่อนในการทำนาย โดยที่การยกกำลังสอง (Square) จะทำให้ค่าคลาดเคลื่อนที่มีค่ามากกว่ามีน้ำหนักมากขึ้น และการใช้รากที่สอง (Square Root) จะทำให้เราได้ค่าคลาดเคลื่อนเฉลี่ยในการทำนายที่มีขนาดเป็นเดียวกับของตัวแปรที่ถูกทำนาย ซึ่งถ้า RMSE มีค่าน้อย แสดงให้เห็นว่าโมเดลมีประสิทธิภาพในการทำนายได้ดีตามข้อมูลที่ให้มา

RMSE มีข้อดีในการให้น้ำหนักมากกว่ากับค่าความคลาดเคลื่อนของข้อมูลที่ห่างไกลมากกว่าค่าความคลาดเคลื่อนของข้อมูลที่ใกล้เคียงกับค่าที่ทำนาย

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (2.6)$$

โดย n คือ จำนวนข้อมูลในชุดข้อมูล

y_i คือ ค่าจริง (Actual Value) ของข้อมูลที่ตัวแปรต้นทาง (Target Variable)

$\hat{y}_i$ คือ ค่าที่ทำนายได้ (Predicted Value) จากโมเดลสำหรับข้อมูลตัวแปรต้นทาง y_i

7. ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ ค่าวัดความแม่นของการทำนายที่วัดจากค่าความคลาดเคลื่อนโดยไม่คำนึงถึงทิศทางของความคลาดเคลื่อน MAE มีหน่วยวัดหน่วยเดียวกับค่าสังเกต

$$\text{Mean Absolute Error} = \frac{\sum_{i-1}^{n}(y_i - \widehat{y_i})}{N} \quad (2.7)$$

โดย n คือ จำนวนข้อมูลในชุดข้อมูล

y_i คือ ค่าจริง (Actual Value) ของข้อมูลที่ตัวแปรต้นทาง (Target Variable)

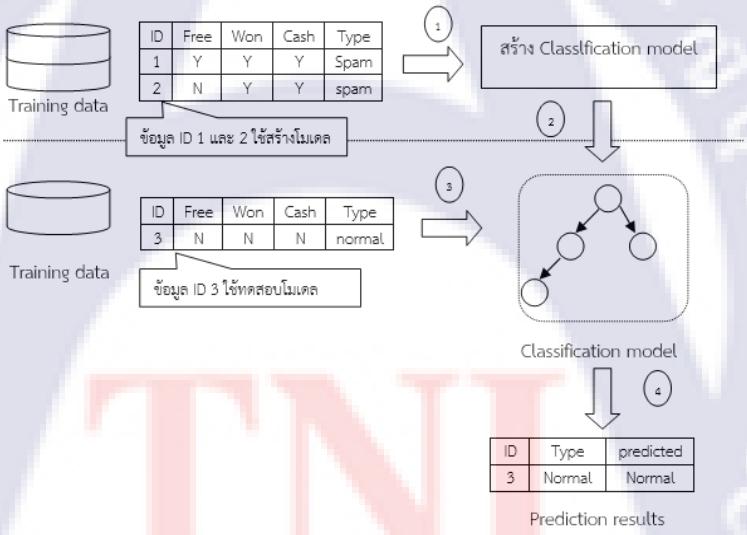
$\hat{y}_i$ คือ ค่าที่ทำนายได้ (Predicted Value) จากโมเดลสำหรับข้อมูลตัวแปรต้นทาง y_i

ทั้ง MAE และ RMSE เป็นวิธีการวัดความคลาดเคลื่อนของการทำนาย โดยที่ RMSE มีความสามารถในการแสดงความคลาดเคลื่อนที่มีน้ำหนักมากกว่ากับค่าที่ทำนายที่ห่างออกไปมากจากค่าจริง ในขณะที่ MAE มีความสามารถในการวัดความคลาดเคลื่อน โดยไม่ให้ความสำคัญกับความห่างหากของค่าที่ทำนายในทุก ๆ กรณี ซึ่งขึ้นอยู่กับลักษณะของงานและวัตถุประสงค์ของการทำนายข้อมูล

2.8 แนวคิดการทดสอบประสิทธิภาพแบบจำลอง

ในการทดสอบประสิทธิภาพของแบบจำลองนั้นโดยทั่วไปแล้วมีวิธีการทดสอบอยู่ 3 วิธี ได้แก่ วิธีการ Hold out-validation Cross validation และ Stratified Sampling

1. วิธีการ Hold out – validation คือ การแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% สมมติในกรณีที่ข้อมูลทั้งหมดคือร้อยละ 100 ระบบจะทำการคัดเลือกข้อมูลร้อยละ 70 หรือร้อยละ 80 เพื่อนำมาใช้ในการสร้างแบบจำลองและข้อมูลส่วนที่สอง คือ ข้อมูลส่วนที่เหลือซึ่งคิดเป็นร้อยละ 30 หรือร้อยละ 20 จะนำมาใช้ในการทดสอบประสิทธิภาพของโมเดล ซึ่งในการเลือกข้อมูลแต่ละครั้งอาจจะไม่เป็นข้อมูลชุดเดียวกัน ทำให้ผลของค่าความถูกต้องในแต่ละครั้งจะมีค่าแตกต่างกันเล็กน้อย ตัวอย่างเช่นในรูป แบ่งข้อมูล Training Data ในตารางออกเป็น 2 ตัวอย่างสำหรับการสร้างโมเดลและข้อมูล 1 ตัวอย่างใช้ในการทดสอบประสิทธิภาพของโมเดล คือ แบ่งข้อมูลออกเป็น 2 ชุด และ Training Data สำหรับสร้างโมเดลและ Testing Data สำหรับทดสอบ



รูปที่ 2.6 การทำงานของ Hold out – Validation

วิธีการ Hold out สำหรับ Training Model นั้น โดยกระบวนนี้จะทำการแบ่งข้อมูลออกเป็นการแบ่งส่วนต่าง ๆ และใช้การแบ่งส่วนเดียวสำหรับ Training Model และการแบ่งส่วนอื่น ๆ สำหรับการตรวจสอบและทดสอบแบบจำลอง วิธีการ Hold out เหมาะสำหรับการประเมินแบบจำลองและการเลือกแบบจำลอง เมื่อใช้ข้อมูลทั้งหมดสำหรับ Training Model โดยใช้โมเดลที่แตกต่างกัน

ปัญหาในการประเมินแบบจำลองและการเลือกแบบจำลองที่เหมาะสมที่สุดยังคงอยู่ งานหลักคือการค้นหาว่าโมเดลใดจากทุกโมเดลที่มีข้อผิดพลาดการวางนัยทั่วไปต่ำที่สุด

Hold out – Validation มักใช้เมื่อชุดข้อมูลมีขนาดเล็กและมีข้อมูลไม่เพียงพอที่จะแบ่งออกเป็นสามชุด (การเทรนนิ่ง การตรวจสอบความถูกต้อง และการทดสอบ) แนวทางนี้มีข้อดี คือ ใช้งานได้ง่าย แต่อาจจะมีความละเอียดว่าข้อมูลแบ่งออกเป็นสองชุดอย่างไร หากมีการแบ่งแบบไม่สุ่มผลลัพธ์ก็อาจมีอคติ โดยรวมแล้ววิธี Hold out สำหรับการประเมินแบบจำลอง เป็นจุดเริ่มต้นที่ดีสำหรับ Training Model แต่ควรใช้ด้วยความระมัดระวัง

2. วิธีการตรวจสอบไขว้กัน (Cross-Validation) [23] คือ วิธีการตรวจสอบค่าความผิดพลาดในการคาดการณ์ของแบบจำลอง โดยพื้นฐานของวิธีการตรวจสอบไขว้กัน คือ การสุ่มตัวอย่าง โดยเริ่มจากการแบ่งชุดข้อมูลออกเป็นส่วน ๆ และนำชุดข้อมูลบางส่วนไปสอนแบบจำลองพยากรณ์ ในอีกส่วนของข้อมูลจะถูกนำไปใช้ทดสอบแบบจำลองที่ถูกสอน เป็นการตรวจสอบไขว้กัน วิธีการนี้ได้รับความนิยมใช้เป็นอย่างมาก มักเรียกว่า K-fold Cross-Validation โดยจะแบ่งข้อมูลออกเป็น k ชุดเท่า ๆ กัน เช่น K = 5 หมายถึง จะมีชุดข้อมูลจำนวน 5 ชุด ซึ่งจะคำนวณค่าความผิดพลาด 5 รอบ โดยแต่ละรอบของการคำนวณจะเลือกข้อมูลออกมา 1 ชุดเพื่อเป็นข้อมูลทดสอบและข้อมูลอีก 4 ชุดจะถูกใช้เป็นข้อมูลสำหรับการเรียนรู้ ดังรูปที่ 2.7

รอบที่ 1 : ชุดสอน	2	3	4	5	ชุดทดสอบ	1
รอบที่ 2 : ชุดสอน	3	4	5	1	ชุดทดสอบ	2
รอบที่ 3 : ชุดสอน	4	5	1	2	ชุดทดสอบ	3
รอบที่ 4 : ชุดสอน	5	1	2	3	ชุดทดสอบ	4
รอบที่ 5 : ชุดสอน	1	2	3	4	ชุดทดสอบ	5

รูปที่ 2.7 การทำงานของ Cross – validation

Cross-Validation เป็นหนึ่งในหัวข้อสำคัญเกี่ยวกับการทดสอบโมเดลการเรียนรู้ ซึ่งถูกคิดค้นโดย K. Dunlap และ K. R. Popper [24] การวัดประสิทธิภาพนี้ยังเคยถูกใช้ในสายงานด้านจิตวิทยาก่อนจะนำมาใช้ในสาย Machine Learning แนวคิด คือ การแบ่งข้อมูลเป็น k ส่วนเท่า ๆ กัน เพื่อสร้างและทดสอบโมเดล (Train + Validate) คำนวณค่าเฉลี่ย Accuracy หรือ Error เช่น ประสิทธิภาพของแบบจำลองก่อนที่จะนำโมเดลไปใช้ทำนายข้อมูลชุดทดสอบ

การทำงานประกอบด้วยสองขั้นตอนหลัก ได้แก่ การแยกข้อมูลออกเป็นส่วนย่อย (Folding) ทำการ Training และ Testing นิยมใช้ K-Fold CV กับโมเดลที่ต้องมีการปรับจูนค่า Parameters เช่น Decision Tree (Cp Max Depth Split Rule) Random Forest (Mtry Ntree) KNN (k Distance) เป็นต้น

3. วิธีการ Stratified Sampling เป็นการสุ่มตัวอย่างที่แบ่งข้อมูลออกเป็นกลุ่มย่อย (Stata) ที่มีลักษณะรวมกันก่อนที่จะทำการสุ่มตัวอย่างจากแต่ละกลุ่ม โดยมีจุดประสงค์ เพื่อให้การสุ่มตัวอย่างสะท้อนโครงสร้างได้อย่างถูกต้อง โดยเฉพาะในกรณีที่ข้อมูลไม่มีความสมดุลหรือขนาดของข้อมูลที่มีขนาดแตกต่างกัน ซึ่งการใช้ Stratified Sampling จะทำให้ผลการวิเคราะห์มีความแม่นยำขึ้น

การนำวิธี Stratified Sampling ไปใช้เพื่อช่วยลดความลำเอียงและเพิ่มความแม่นยำของการพยากรณ์ ทำให้ได้กลุ่มตัวอย่างที่ครอบคลุมกลุ่มย่อยที่มีลักษณะเฉพาะเจาะจง โดยการสุ่มตัวอย่างแบบ Stratified Sampling ช่วยให้มั่นใจว่าแต่ละกลุ่มย่อยมีตัวแทนอยู่ในชุดข้อมูล เพื่อไม่ให้ข้อมูลเกิดความลำเอียงไปกลุ่มใหญ่มากเกินไป นอกจากนี้ยังปรับปรุงความแม่นยำในการทำนายและเพิ่มประสิทธิภาพของการประเมินผลใน Cross-Validation ซึ่งจะช่วยให้แต่ละชุดข้อมูลย่อยที่ใช้ในการฝึกโมเดลและทดสอบมีสัดส่วนของคลาสที่มีความสม่ำเสมอกัน ทำให้การประเมินผลโมเดลมีความถูกต้องแม่นยำมากขึ้น โดยเฉพาะปัญหาความไม่สมดุลของข้อมูล ซึ่งการแบ่งแบบสุ่มโดยไม่มีการจัดสรรอย่างเป็นระบบอาจทำให้ผลลัพธ์ดีแบบไม่น่าเชื่อ

2.9 งานวิจัยที่เกี่ยวข้อง

ระพีวรรณ อุดมรัตน์ [25] ได้ศึกษาเกี่ยวกับ การศึกษาลักษณะของลูกค้าธุรกิจขนาดกลางที่มีโอกาสเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ (NPL): กรณีศึกษาของสถาบันการเงินแห่งหนึ่ง โดยมีจุดประสงค์เพื่อทำนายความสัมพันธ์ปัจจัยที่เป็นหนี้ที่ไม่ก่อให้เกิดรายได้ ใช้เทคนิค Logistic Regression มาใช้เพื่อวัดผลและวิเคราะห์เชิงสถิติ โดยนำใช้ข้อมูลจากระบบ Core Banking มาวิเคราะห์ข้อมูลลูกค้าธุรกิจขนาดกลาง ตั้งแต่ปี 2557-2561 จำนวน 45,627 ราย ทำการประมวลผลข้อมูลทั้ง 8 ตัวแปร ผลการศึกษาพบว่า ปัจจัยที่มีความสัมพันธ์กับการเป็นหนี้ที่ไม่ก่อให้เกิดรายได้คือ วงเงินสินเชื่อและจำนวนวันค้างชำระ โดยทั้งสองปัจจัยมีความสัมพันธ์ในทิศทางเดียวกับการเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ ขณะที่อีก 6 ปัจจัย คือ อัตราส่วนภาระหนี้ต่อรายได้อัตราส่วนหนี้สินต่อทุน อัตรากำไรสุทธิ อัตราผลตอบแทนของสินทรัพย์ อัตราส่วนความสามารถในการจ่ายดอกเบี้ย และอัตราทุนหมุนเวียนไม่พบว่ามีความสัมพันธ์กับการเป็นหนี้ที่ไม่ก่อให้เกิดรายได้

กันทิมา มีนิล [26] ได้ศึกษาเรื่องการศึกษาเพื่อพยากรณ์ลูกค้า NPL สินเชื่อเพื่อที่อยู่อาศัยของธนาคารอาคารสงเคราะห์ และจัดกลุ่มการเข้าสู่สภาวะความเสี่ยง มีจุดมุ่งหมายเพื่อศึกษาปัจจัยที่มีผลต่อการเกิดลูกค้า NPL ศึกษาพฤติกรรมที่มีผลก่อให้เกิดแนวโน้มลูกค้า NPL และแนวทางในการลดความเสี่ยง ใช้ฐานข้อมูลลูกค้าสินเชื่อที่อยู่อาศัยทั้งที่เป็น NPL และไม่เป็น NPL จำนวน 10,000 รายการ

ใช้โมเดลการเรียนรู้ของเครื่อง ได้แก่ Decision Tree J48 K-Nearest Neighbor และ Naïve Bayes ผลการศึกษาพบว่า จากการนำข้อมูลมาทดสอบด้วยโมเดล KNN K=2 โดยใช้การพยากรณ์ ประเภท Training Set ข้อมูลมีความแม่นยำ 94.66% ประเภท Supplied set Test ด้วยการ Re-evaluate Model ข้อมูลมีความแม่นยำ 93.80% และประเภท Cross Validation Folds 10 ข้อมูลมีความแม่นยำ อยู่ 89.74%

ปารดา ศัสตุระ [27] ศึกษาเรื่องการประยุกต์เหมืองข้อมูลในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ กรณีศึกษาธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่ จุดมุ่งหมายเพื่อศึกษาปัจจัยที่ไม่ก่อให้เกิดรายได้ NPLs และการประยุกต์เหมืองข้อมูล ในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ วิธีการใช้อัลกอริทึมการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ 3 แบบ ได้แก่ Decision Tree Naive Bayes และ Random Forest ข้อมูลที่ใช้รวบรวมข้อมูลจากลูกค้าที่ได้รับสินเชื่อ Covid - 19 ของธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่ จำนวน 7,215 ราย ผลการศึกษาพบว่า ปัจจัยที่มีผลต่อการเกิดหนี้ที่ไม่ก่อให้เกิดรายได้ ได้แก่อาชีพ สถานะภาพ ระดับการศึกษา รายได้เพศ และผลการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อจากอัลกอริทึม Decision Tree มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 96.15%

เฉลิมชาติ ชัยวิลา และศิวาพร ฟองทอง [28] ได้ศึกษาเรื่องปัจจัยที่ส่งผลต่อการชำระหนี้ของลูกค้าธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร สาขาบ้านผือ จังหวัดอุดรธานี โดยมีวัตถุประสงค์เพื่อศึกษาพฤติกรรมทางด้านหนี้สินและศึกษาปัจจัยที่มีผลต่อการชำระหนี้ของลูกค้าเงินกู้ของธนาคารเพื่อการเกษตรและสหกรณ์การเกษตรสาขาบ้านผือ อำเภอบ้านผือ จังหวัดอุดรธานี จากงานวิจัยพบว่า พฤติกรรมทางด้านหนี้สินของกลุ่มตัวอย่างเป็นลูกค้าเงินกู้เดิมของธนาคาร ซึ่งมียอดเงินกู้ครั้งแรกเฉลี่ยอยู่ที่ 97,531 บาท โดยลูกค้ามักจะใช้หลักประกันเป็นอสังหาริมทรัพย์และกู้เงินแบบกลุ่มบุคคล โดยมียอดเงินกู้คงเหลือเฉลี่ยอยู่ที่ 297,237 บาท อัตราดอกเบี้ยเงินกู้เฉลี่ยร้อยละ 7.27 และบางส่วนมีการพักชำระหนี้เงินต้นอยู่ โดยกลุ่มตัวอย่างส่วนใหญ่มีการชำระเงินตรงตามกำหนด ปัจจัยที่ส่งผลต่อการชำระหนี้ของลูกค้าในทางบวกอย่างมีนัยสำคัญทางสถิติ คือ ปัจจัยด้านธนาคาร ได้แก่ หลักประกันเงินกู้ประเภทอสังหาริมทรัพย์และความพึงพอใจต่อพนักงาน ในทางตรงกันข้ามปัจจัยที่ส่งผลต่อการชำระหนี้ในทางลบอย่างมีนัยสำคัญทางสถิติ คือ ปัจจัยด้านเศรษฐกิจ ได้แก่ ค่าใช้จ่ายฉุกเฉินระหว่างช่วงการชำระหนี้และปัจจัยด้านธนาคาร ได้แก่ อัตราดอกเบี้ยเงินกู้

สุพัตรา มนัสชล และนลินี ทินนาม [29] ได้ศึกษาเรื่องศักยภาพในการชำระหนี้ของผู้กู้โครงการสินเชื่อฟื้นฟู SMEs จากอุทกภัยและภัยพิบัติ ปี 2560 ของธนาคารพัฒนาวิสาหกิจขนาดกลางและขนาดย่อมแห่งประเทศไทย สาขานครศรีธรรมราช โดยมีวัตถุประสงค์เพื่อศึกษาศักยภาพในการชำระหนี้ของผู้กู้โครงการสินเชื่อฟื้นฟู SMEs จากอุทกภัยและภัยพิบัติปี 2560 ของธนาคาร

พัฒนาวิสาหกิจขนาดกลางและขนาดย่อมแห่งประเทศไทย สาขานครศรีธรรมราช พบว่า ผู้กู้โครงการสินเชื่อฟื้นฟู SMEs จากอุทกภัยและภัยพิบัติ ปี 2560 ของธนาคารพัฒนาวิสาหกิจขนาดกลางและขนาดย่อมแห่งประเทศไทย สาขานครศรีธรรมราช ส่วนใหญ่มีประสบการณ์ในการทำธุรกิจที่ขอกู้ในครั้งนี้ 10 ปีขึ้นไป มีประวัติการติดต่อกับสถาบันการเงิน 10 ปีขึ้นไปไม่เคยผิดนัดชำระหนี้ มีความสามารถในการชำระหนี้สูง ใช้เงินทุนในการดำเนินธุรกิจจากการกู้มากกว่าเงินทุนส่วนตัวและใช้บริการค้ำประกันวงเงินกู้จากบรรษัทประกันสินเชื่ออุตสาหกรรมขนาดย่อม (บสย.) ซึ่งแสดงให้เห็นว่าผู้กู้ในโครงการฯ มีจุดแข็งเกี่ยวกับศักยภาพในการชำระหนี้ 2 ด้าน คือ ด้านคุณสมบัติของผู้กู้ (Character) ซึ่งพิจารณาจากประสบการณ์ในการดำเนินธุรกิจ ประวัติการติดต่อสถาบันการเงินและประวัติการชำระหนี้ ส่วนในด้านความสามารถในการชำระหนี้ (Capacity) พิจารณาจากความสามารถในการทำกำไรเพื่อไปชำระหนี้เงินต้นและดอกเบี้ย แต่อย่างไรก็ตามจากผลการวิจัยพบประเด็นที่ธนาคารต้องพึงระวัง2 ด้าน คือ ด้านเงินทุน (Capital) ซึ่งพิจารณาจากแหล่งเงินทุนที่ใช้ในการดำเนินธุรกิจและด้านหลักประกัน (Collateral) ซึ่งพิจารณาจากหลักทรัพย์หรือสิ่งที่ผู้กู้นำมาค้ำประกันเงินกู้ โดยผลการวิจัยนี้ ธนาคารสามารถนำไปใช้ในการวางแผนหาแนวทางควบคุมดูแลและติดตามการชำระหนี้ของผู้กู้ในโครงการนี้ เพื่อให้ผู้กู้สามารถชำระหนี้ได้อย่างราบรื่นตามวัตถุประสงค์ของโครงการโดยไม่ก่อให้เกิดหนี้เสียต่อธนาคาร

ศิริลักษณ์ บุญชัยสุข และอุษณา แจ้งคล้อย [30] มีการศึกษาเรื่องปัจจัยที่มีผลต่อพฤติกรรมการจ่ายชำระหนี้ของลูกหนี้กองทุนส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการในเขตจังหวัดนครราชสีมา มีจุดประสงค์เพื่อศึกษาเปรียบเทียบปัจจัยที่มีผลต่อพฤติกรรมการจ่ายชำระหนี้ของลูกหนี้กองทุนส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการในเขตจังหวัดนครราชสีมา จากการศึกษา พบว่า ปัจจัยที่มีผลต่อพฤติกรรมการชำระหนี้ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 95 คือ ปัจจัยด้านอาชีพและปัจจัยด้านอายุของสัญญากู้ยืม ส่วนปัจจัยด้านเพศ ปัจจัยด้านอายุ ปัจจัยด้านสถานะผู้กู้และปัจจัยด้านประเภทความพิการ ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

ชลธิชา สุวรรณพิทักษ์ [31] ได้ศึกษาเรื่องปัจจัยที่มีผลต่อการชำระหนี้ตามสัญญาปรับปรุงโครงสร้างหนี้ของลูกหนี้สินเชื่อบุคคล กรณีศึกษา ธนาคารกสิกรไทย จำกัด (มหาชน) หน่วยบริหารคุณภาพสินทรัพย์ พัทยากลาง โดยมีวัตถุประสงค์เพื่อศึกษาลักษณะส่วนบุคคล ปัจจัยด้านภาระหนี้ของลูกหนี้ที่มีกับธนาคารและปัจจัยด้านความสามารถในการชำระหนี้ของลูกหนี้สินเชื่อบุคคลที่ทำสัญญาปรับปรุงโครงสร้างหนี้กับบริษัท ธนาคารกสิกรไทย จำกัด (มหาชน) หน่วยบริหารคุณภาพสินทรัพย์พัทยากลางและเพื่อศึกษาปัจจัยที่มีผลต่อการชำระหนี้ตามสัญญาปรับปรุงโครงสร้างหนี้สินเชื่อบุคคลของบริษัท ธนาคารกสิกรไทย จำกัด (มหาชน) หน่วยบริหารคุณภาพสินทรัพย์พัทยากลาง จากการศึกษาพบว่า ตัวแปรด้านสถานภาพสมรส ตัวแปรด้านรายได้เฉลี่ยต่อเดือน ตัวแปรจำนวนเงินผ่อนชำระต่อเดือน ตัวแปรด้านประเภทหลักประกัน ปัจจัยด้านความสามารถ

ในการชำระหนี้ และปัจจัยเสี่ยงอื่น ๆ เป็นปัจจัยที่กลุ่มตัวอย่างให้ความสำคัญมากที่สุดโดยมีระดับนัยสำคัญทางสถิติ 0.10

เครือวัลย์ เนตรพนา [32] ได้ศึกษาเรื่องการวิเคราะห์ความเสี่ยงในการผิดนัดชำระของลูกหนี้บัตรเครดิตโดยการใช้อัลกอริทึมการเรียนรู้ของเครื่อง โดยมีวัตถุประสงค์เพื่อการศึกษาการทำนายลูกหนี้ที่มีโอกาสในการผิดนัดชำระกับทางธนาคารโดยใช้ชุดข้อมูล การทำธุรกรรมสินเชื่อบัตรเครดิตซึ่งประกอบด้วย ข้อมูลจำนวนทั้งหมด 307,511 แถว และคอลัมน์ทั้งหมด 122 คอลัมน์ ใช้เทคนิคการเรียนรู้ของเครื่องมาพัฒนาแบบจำลอง ได้แก่ Logistic Regression XG Boost Classifier K-nearest Neighbors Random Forest Support Vector Classifier (SVC) และ Gradient Boosting โดยแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มลูกหนี้ปกติ คือ กลุ่มลูกหนี้ที่ไม่ได้มีการผิดนัดชำระกับทางธนาคารและกลุ่มลูกหนี้ที่ไม่ปกติคือ กลุ่มลูกหนี้ที่มีการผิดนัดชำระกับทางธนาคาร โดยเอาผลลัพธ์ที่ได้ไปตรวจสอบกับชุดข้อมูลที่ใช้ในการทดสอบที่มีอยู่ว่าแบบจำลองที่ถูกพัฒนาขึ้นนั้น มีประสิทธิภาพและความถูกต้อง (Accuracy) มากน้อยเพียงใด จากผลการศึกษาสรุปได้ว่า การพัฒนาแบบจำลองโดยการใช้วิธี Gradient Boosting ให้ค่าความไว (Recall) ที่มากที่สุด ซึ่งมีค่าเท่ากับ 0.65 มีค่าความถูกต้อง (Accuracy) เท่ากับ 0.62 และมีค่า F1-Score ที่ใช้ในการวัดความสามารถของแบบจำลองเท่ากับ 0.54

สุดากาญจน์ ติดตะ [33] ศึกษางานวิจัยเกี่ยวกับระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทคนิควิธีการเรียนรู้ของเครื่อง โดยใช้เทคนิค Naive Bays K-nearest Neighbors Support Vector Machine และ Random Forest งานวิจัยนี้ได้นำแบบจำลองสำหรับพิจารณาอนุมัติเงินกู้สวัสดิการ เพื่อลดภาระการทำงานหนักของเจ้าหน้าที่ในการอนุมัติเอกสารขอกู้เงิน จากการรวบรวมปัจจัยที่มีผลต่อการทำนายการอนุมัติการขอกู้เงิน ได้ปัจจัยมาทั้งหมด 14 ปัจจัย โดยทำการแบ่งกลุ่มลูกค้าเป็น 2 กลุ่ม คือ กลุ่มผู้กู้ที่ได้รับการอนุมัติและกลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ ผลการศึกษาพบว่า เทคนิค Random Forest ให้ความแม่นยำสูงสุดที่ 99.56% โดยปัจจัยที่มีความสำคัญในการทำนายการอนุมัติเงินกู้ ได้แก่ ความเสี่ยงในการค้ำประกันและเงินเดือนพนักงานของผู้กู้และผู้ค้ำตามลำดับ

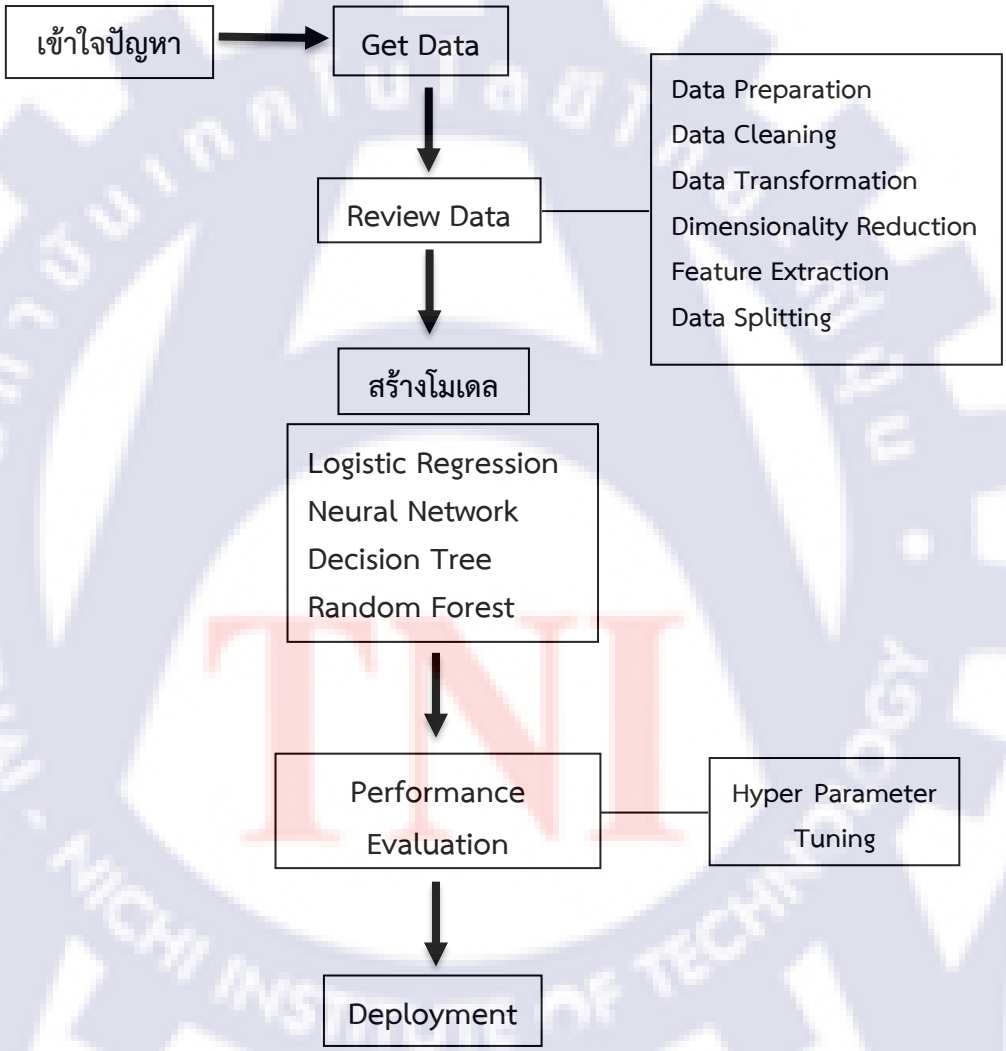
งานวิจัยของวริทธ์พล แสงทองรัตนโชติ [34] ศึกษาเรื่องการเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องเพื่อสร้างตัวแบบการจำแนกด้วยการปรับปรุงชุดข้อมูลอสมดุล โดยหลักการส่วนใหญ่จะมุ่งเน้นไปที่การเปรียบเทียบประสิทธิภาพตัวแบบการจำแนกมากกว่าการศึกษาการเลือกใช้เทคนิคการจำแนกอย่างไรให้เหมาะสมกับชุดข้อมูล เป้าหมายเพื่อศึกษาประสิทธิภาพเทคนิคการจำแนกกับชุดข้อมูลที่มีจำนวนของตัวแปรอิสระเชิงคุณภาพและเชิงปริมาณที่แตกต่างกันทั้งหมด 3 ชุดข้อมูล ได้แก่ ชุดข้อมูลสถาบันการเงินซึ่งเป็นชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงคุณภาพและเชิงปริมาณเท่ากัน ชุดข้อมูลสายพันธุ์ข้าว ซึ่งเป็นชุดข้อมูลที่มีตัวแปรอิสระเชิงปริมาณเท่านั้นและชุดข้อมูล

นักวิทยาศาสตร์ข้อมูล ซึ่งเป็นชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงคุณภาพมากกว่าเชิงปริมาณ วิธีการที่นำมาใช้ปรับปรุงชุดข้อมูลอสมดุลให้สมดุลใช้วิธีสุ่มลด 2 เทคนิค ได้แก่ การสุ่มตัวอย่างแบบง่ายและการแบ่งกลุ่มข้อมูลแบบเคมีน แบ่งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบด้วยหลักการ 5-Fold โดยนำชุดข้อมูลแต่ละชุดมาสร้างตัวแบบการจำแนกด้วยเทคนิคการจำแนกทั้งหมด 5 เทคนิค ได้แก่ การวิเคราะห์จำแนกกลุ่มเชิงเส้นโดยวิธีของฟิชเชอร์ เทคนิคนาอีฟเบย์ ต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 เทคนิคป่าสุ่มและโครงข่ายประสาทเทียม ซึ่งผลการวิจัยระบุว่าเทคนิคป่าสุ่มสามารถทำงานได้ดีภายใต้ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงคุณและปริมาณเท่ากัน เทคนิคการจำแนกการวิเคราะห์จำแนกกลุ่มเชิงเส้นโดยวิธีของฟิชเชอร์สามารถทำงานได้ดีภายใต้ชุดข้อมูลที่มีตัวแปรอิสระเชิงปริมาณทุกตัวและโครงข่ายประสาทเทียมสามารถทำงานได้ดีภายใต้ชุดข้อมูลที่มีจำนวนตัวแปรอิสระเชิงคุณภาพมากกว่าปริมาณและพบว่า การปรับปรุงชุดข้อมูลให้สมดุลด้วยการสุ่มตัวอย่างแบบง่ายให้ตัวแบบการจำแนกที่มีประสิทธิภาพสูงกว่าการแบ่งกลุ่มข้อมูลแบบเคมีน

งานวิจัยของพัชรียา ทองพูล และคณะ [35] ทำการศึกษาเกี่ยวกับการเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล จัดทำขึ้นเพื่อเปรียบเทียบวิธีการปรับข้อมูลที่ไม่สมดุล 4 วิธี คือ วิธีการสุ่มเกินวิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน โดยมาจากปัญหาการแบ่งกลุ่มข้อมูลที่ไม่สมดุลโดยข้อมูลที่อยู่ในกลุ่มส่วนน้อยถูกครอบงำหรือจะถูกจัดให้ไปอยู่ในกลุ่มส่วนมากทั้งหมด ซึ่งจะนำไปสู่ปัญหาที่เรียกว่าปัญหาการจำแนกข้อมูลผิดกลุ่ม จากงานวิจัยมีการเลือกใช้วิธีการสุ่มเกิน (Over Sampling) วิธีการสุ่มเกินโดยใช้เทคนิค SMOTE (Synthetic Minority Over Sampling) วิธีการสังเคราะห์ข้อมูลใหม่ วิธีสุ่มลด และวิธีผสมผสาน โดยใช้วิธีการจำแนก 4 วิธี คือ วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม และวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยเมื่อพิจารณาชุดข้อมูลการรับรู้ทางหูของเด็ก พบว่า เครื่องสนับสนุนเวกเตอร์โดยการแก้ไขข้อมูลอสมดุลด้วยเทคนิค SMOTE มีประสิทธิภาพในการจำแนกสูงที่สุด เมื่อพิจารณาชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าพบว่า เทคนิคเพื่อนบ้านใกล้ที่สุดโดยการแก้ไขข้อมูลอสมดุลด้วยเทคนิค SMOTE มีประสิทธิภาพในการจำแนกสูงที่สุด และเมื่อพิจารณาชุดข้อมูลคุณภาพไวน์แดง พบว่า เครื่องสนับสนุนเวกเตอร์โดยการปรับปรุงชุดข้อมูลอสมดุลให้สมดุลด้วยวิธีสุ่มลดมีประสิทธิภาพในการจำแนกสูงที่สุด

บทที่ 3
วิธีการดำเนินงาน

งานวิจัยนี้เป็นการเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมด้วยเทคนิคการเรียนรู้ของเครื่อง มีวัตถุประสงค์เพื่อสร้างแบบจำลองข้อมูลในการวิเคราะห์หาความสามารถในการชำระหนี้คืนโดยใช้เทคนิคการเรียนรู้ของเครื่องและเพื่อเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมาะสมกับการวิเคราะห์ความสามารถในการชำระหนี้ มีขั้นตอนดังนี้



รูปที่ 3.1 ภาพแสดง Conceptual Framework

3.1 ขั้นตอนการวิเคราะห์ข้อมูล

ขั้นตอนการวิเคราะห์ข้อมูล อ้างอิงการใช้กระบวนการ CRISP-DM ซึ่งเป็นการวิเคราะห์ข้อมูลด้าน Data Mining มี 6 ขั้นตอน ดังนี้

1. เข้าใจปัญหา (Business Understanding)

พิจารณาถึงปัจจัยอะไรบ้างที่มีผลต่อความสามารถในการชำระหนี้คืนของลูกค้าและปัจจัยอะไรบ้างที่ทำให้เกิดหนี้สูญ เมื่อเข้าใจปัญหาที่เกิดขึ้นแล้วผู้วิจัยจึงได้นำปัญหามาเป็นประเด็นในการทำวิจัย

2. การเก็บรวบรวมข้อมูลและการทำความเข้าใจกับข้อมูล
(Get Data & Data Understanding)

ข้อมูลเป็นปัจจัยที่สำคัญที่สุดที่ขาดไม่ได้ในการทำ Data Mining ในขั้นตอนนี้เป็นการรวบรวมข้อมูลที่เกี่ยวข้อง ข้อมูลถูกต้อง น่าเชื่อถือ เป็นข้อมูลที่มีปริมาณเพียงพอและมีความเหมาะสม มีรายละเอียดที่เป็นประโยชน์เพื่อนำไปใช้ในกระบวนการวิเคราะห์ โดยข้อมูลที่นำมาใช้เป็นข้อมูลการประเมินศักยภาพลูกหนี้ของลูกค้ารายเดิม

3. การเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนที่ใช้เวลานานที่สุดเนื่องจากแบบจำลองที่ได้จากการทำ Data Mining จะให้ผลลัพธ์ที่ถูกต้องหรือไม่นั้น ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้โดยการเตรียมข้อมูลนั้น ๆ การตรวจสอบและแก้ไขข้อมูล (Data Cleaning) โดยดูจากลักษณะรูปแบบของข้อมูล ดูความครบถ้วนสมบูรณ์ของข้อมูล ตรวจสอบข้อมูลขาดหาย ข้อมูลพร้อมนำไปใช้งานหรือไม่ เพื่อเป็นการตรวจสอบข้อมูลและแก้ไขข้อมูลให้มีความพร้อมต่อการนำไปวิเคราะห์ต่อไป ต่อไปเป็นการคัดเลือกข้อมูล (Data Selection) โดยคัดเลือกข้อมูลที่เกี่ยวข้องกับการนำไปวิเคราะห์ความสามารถในการชำระหนี้ เลือกปัจจัยที่จะนำไปใช้งาน ซึ่งข้อมูลที่คัดเลือกต้องเป็นปัจจัยที่มีผลต่อการพิจารณาอนุมัติสินเชื่อ ขั้นตอนต่อไปเป็นการเปลี่ยนรูปแบบข้อมูล (Data Transformation) โดยเปลี่ยนแปลงรูปแบบข้อมูลจากรูปแบบ String ให้เป็นรูปแบบที่จะนำใช้งานในรูปแบบ Numeric Data ซึ่งได้กำหนดให้เงินต้นค้างในบิล (The Principal Remains on the Bill) เป็นตัวแปรหลัก และทำการ Normalize Data เพื่อปรับสเกลของข้อมูลให้มีค่าช่วงที่กำหนด ทำให้แต่ละฟีเจอร์อยู่ในช่วงเดียวกัน ซึ่งจะช่วยให้โมเดล Machine Learning สามารถทำงานได้ดีขึ้น โดยเฉพาะอย่างยิ่งสำหรับโมเดลที่ใช้อัลกอริทึมที่อ่อนไหวต่อขนาดของข้อมูล เช่น Neural Networks Support Vector Machine K-Nearest Neighbors และ Logistic Regression เป็นต้น

ผู้วิจัยได้ทำการกำหนดแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มที่มีความสามารถในการชำระหนี้ (กลุ่มที่ผ่านการพิจารณาอนุมัติ) โดยกำหนดให้ 1 แทนลูกค้าชำระหนี้คืนยังไม่เสร็จสิ้น และกำหนด 0 แทนลูกค้าที่ชำระหนี้คืนเสร็จสิ้น

ข้อมูลที่นำมาใช้เป็นข้อมูลการประเมินศักยภาพลูกหนี้ของลูกค้ารายเดิมของธนาคารแห่งหนึ่ง จำนวน 4,334 ราย โดยทำงานแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ Training Data และ Testing Data จากชุดข้อมูลที่มีพบว่า ข้อมูลมีการกระจายข้อมูลอย่างกว้างและมีความแปรปรวนสูง กล่าวคือ ลูกค้าหลายรายมียอดหนี้คงเหลือสูงมากและลูกค้าบางรายมีหนี้คงเหลือต่ำ ในขั้นตอนต่อไปจะทำการลดมิติของข้อมูลโดยการใช้ Principal Component Analysis (PCA) และ Linear Discriminant Analysis (LDA) เพื่อต้องการที่จะลดมิติของข้อมูล ทำให้ข้อมูลมีประสิทธิภาพมากที่สุดและเป็นการเพิ่มความน่าเชื่อถือของโมเดลมากขึ้น

4. การสร้างแบบจำลอง (Modeling)

เมื่อทำการแบ่งข้อมูลออกเป็น Training Data และ Testing Data แล้วทำการปรับข้อมูลให้มีความสมดุลกัน โดยใช้เทคนิคต่าง ๆ เช่น เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) หลังจากนั้นนำไปพัฒนาแบบจำลองการวิเคราะห์ความสามารถในการชำระหนี้ของผู้กู้รายเดิมโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยใช้เทคนิคที่นำมาใช้ ได้แก่ โมเดล Logistic Regression Neural Network Decision Tree และ Random Forest เมื่อสร้างแบบจำลองแล้วนำผลลัพธ์ที่ได้มาเปรียบเทียบค่าความถูกต้องของแต่ละโมเดล เพื่อหาโมเดลที่ให้ค่าความถูกต้องมากที่สุดในการจำแนกข้อมูลที่ได้จากวิธีที่นำเสนอกับวิธีที่มีผู้เสนอแนะไว้ก่อนหน้าที่มีความเชื่อถือมากที่สุด โดยเทคนิคที่เลือกใช้ มีดังนี้

4.1 Logistic Regression

ขั้นตอนที่ 1 หาอัตราส่วนออดส์หรืออัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจต่อโอกาสไม่เกิดเหตุการณ์ที่สนใจ

ขั้นตอนที่ 2 ตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระ

ขั้นตอนที่ 3 ตรวจสอบความเหมาะสมของตัวแบบการถดถอยโลจิสติก

ขั้นตอนที่ 4 ทดสอบความมีนัยสำคัญของสัมประสิทธิ์ถดถอย

ขั้นตอนที่ 5 ทดสอบความน่าเชื่อถือของตัวแบบพยากรณ์

ขั้นตอนที่ 6 คัดเลือกตัวแปรอิสระทั้งหมด 4 วิธี คือ วิธีการเลือกแบบพื้นฐาน วิธีการเลือกแบบไปข้างหน้า วิธีการกำจัดแบบถอยหลัง และวิธีการถดถอยทีละขั้น กำหนดระดับนัยสำคัญเท่ากับ 0.05

4.2 Neural Network

การทำงานส่วนของโครงข่ายประสาทเทียมจะทำหน้าที่แยกหมวดหมู่ ต้องออกแบบลักษณะของโครงข่ายประสาทเทียมขึ้นมาก่อน จากนั้นทำการสอนให้โครงข่ายจดจำรูปแบบต่าง ๆ ตามที่ต้องการ ซึ่งการออกแบบโครงข่ายขึ้นกับลักษณะงาน ไม่สามารถกำหนดเฉพาะเจาะจงได้ ทำให้ต้องมีการทดลองและปรับเปลี่ยนลักษณะของโครงข่ายจนได้ลักษณะของโครงข่ายที่มีประสิทธิภาพ

การทำงานที่พึงพอใจ เช่น เริ่มจากการออกแบบโครงข่ายขนาดเล็กก่อน เมื่อพบว่า โครงข่ายเกิดค่าเฉลี่ยความผิดพลาดสูงในการสอนจึงค่อย ๆ เพิ่มโหนดในชั้นต่าง ๆ มากขึ้น แล้วทำการตัดโหนดที่มีการเปลี่ยนแปลงค่าน้ำหนักน้อยออกไป เนื่องจากโหลดเหล่านั้นไม่มีผลต่อการใช้งาน

4.3 Decision Tree

ต้นไม้การตัดสินใจเป็นขั้นตอนวิธีหนึ่งที่ได้รับความนิยมในการจำแนกหมวดหมู่เอกสาร กฎการจำแนกหมวดหมู่ข้อมูลที่ได้จากขั้นตอนวิธีนี้อยู่ในรูปแบบของลักษณะโครงสร้างต้นไม้ ซึ่งภายในประกอบไปด้วยโหนด (Node) และกิ่ง (Branch) แต่ละโหนดจะถูกแทนด้วยคุณลักษณะ (Feature) ของชุดข้อมูลที่นำมาเรียนรู้และทดสอบ แต่ละกิ่งของต้นไม้แสดงผลในการในการทดสอบ และลีฟโหนด (Leaf Node) แสดงหมวดหมู่ที่ผู้ใช้กำหนด ส่วนเกณฑ์การเลือกคุณลักษณะเพื่อนำมาเป็นโหนดของต้นไม้นั้นมาจากการคำนวณค่าเกนสารสนเทศ (Information Gain) โดยพิจารณาคุณลักษณะที่มีค่าเกนสารสนเทศหรือมีค่าเอ็นโทรพี (Entropy) ต่ำหมายความว่าคุณลักษณะนั้นมีความสามารถในการจำแนกหมวดหมู่สูง

4.4 Random Forest

แนวคิดของ Random Forest คือ การสร้างโมเดลด้วยวิธีการ Decision Tree ขึ้นมาหลาย ๆ โมเดล โดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละโมเดลมารวมกัน พร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย วิธีการของ Decision Tree คือ เทคนิคที่ให้ผลลัพธ์ในลักษณะเป็นโครงสร้างของต้นไม้ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดจะมีเงื่อนไขของคุณลักษณะเป็นตัวทดสอบ กิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบและใบ (Leaf) เป็นสิ่งที่อยู่อย่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (Class) ก็คือผลลัพธ์ที่ได้จากการพยากรณ์ ซึ่งข้อดีของวิธีการนี้คือให้ผลการพยากรณ์ที่แม่นยำและเกิดปัญหา Over fitting น้อย

ผู้วิจัยทำการกำหนด 5 ปัจจัย เพื่อหาปัจจัยที่มีผลต่อความสามารถในการชำระหนี้คืนมากที่สุด ประกอบด้วย เบี้ยปรับ (Late Charge) หนี้คงค้าง (Outstanding Debt) ดอกเบี้ยคงค้าง (Due Interest) เงินต้นค้างในบิล (The Principal Remains on the Bill) และดอกเบี้ยค้างในบิล (Interest Accrued on the Bill) ซึ่งผู้วิจัยทำการแบ่งข้อมูลลูกค้าออกเป็น 2 กลุ่ม คือ กลุ่มที่มีความสามารถในการชำระหนี้ (กลุ่มที่ผ่านการอนุมัติ) โดยกำหนดให้ 1 แทน ลูกค้ายังไม่ได้ชำระหนี้คืนเสร็จสิ้น และ 0 แทนลูกค้าชำระคืนเสร็จสิ้น

5. การประเมินผล (Evaluation Phase)

เป็นการประเมินผลลัพธ์ประสิทธิภาพของแต่ละอัลกอริทึมที่นำมาใช้ โดยวิเคราะห์ข้อมูลว่า ผลลัพธ์สามารถจำแนกได้ถูกต้องแม่นยำและสามารถตอบวัตถุประสงค์ที่กำหนดไว้หรือไม่ โดยในการประเมินแบบจำลองนั้นจะแบ่งตามลักษณะของการทำเหมืองข้อมูล กฎความสัมพันธ์

สามารถประเมินผลที่ได้จากการพิจารณาค่าความเชื่อมั่นและค่าสนับสนุนประเมินผล โดยการเปรียบเทียบผลพยากรณ์ที่ได้กับข้อมูลจริงที่เกิดขึ้น ซึ่งในการวัดประสิทธิภาพในการทำเหมืองข้อมูลนั้นจะประกอบไปด้วย

5.1 ค่าความถูกต้อง (Accuracy) คือ จำนวนข้อมูลที่ทำนายถูกของคลาส

5.2 ค่าความแม่นยำ (Precision) คือ ค่าตัวแบบทำนายที่ถูกต้อง

5.3 ค่าความครบถ้วน (Recall) คือ ค่าของตัวแบบที่ตรงกับค่าความเป็นจริง

5.4 ค่าวัดประสิทธิภาพโดยรวม (F-1 Score) คือ ค่าที่เกิดจากการเปรียบเทียบระหว่างค่าความแม่นยำและค่าความครบถ้วน

โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงาน คือ K-Fold Cross Validation โดยกำหนดให้มีจำนวน 5 กลุ่ม (K=5) และทำการทดสอบ 5 รอบ เพื่อใช้ทดสอบประสิทธิภาพของโมเดล การเลือกใช้ K-Fold Cross Validation มาใช้ในงานวิจัยจะทำให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือ การวัดประสิทธิภาพด้วยวิธี Cross-Validation นี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน ซึ่งมักจะแสดงด้วยค่า k โดยที่แต่ละส่วนกำหนดให้มีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล โดยจะเปรียบเทียบจากค่าความถูกต้อง (Accuracy) ของแต่ละโมเดล เพื่อหาโมเดลที่ได้ประสิทธิภาพที่ดีที่สุด เมื่อได้โมเดลที่ให้ค่าความถูกต้องสูงที่สุดแล้วจะนำโมเดลที่ได้จากการเปรียบเทียบมาทำการปรับค่าพารามิเตอร์ (Hyper parameter Tuning) เพื่อช่วยปรับโมเดลให้มีค่าผลลัพธ์ที่ดีที่สุดมากกว่าเดิม

6. การนำแบบจำลองไปใช้งาน (Deployment)

การใช้เกณฑ์การพิจารณาสินเชื่อแบบเดิมอาจไม่สามารถทำนายความเสี่ยงของผู้กู้ได้แม่นยำ อาจส่งผลกระทบให้เกิดหนี้เสีย (NPL) เพิ่มสูงขึ้น หากมีการนำ Machine Learning เข้ามาใช้งานควบคู่กัน จะช่วยให้ Credit Scoring Model สามารถวิเคราะห์ข้อมูลจำนวนมากได้อย่างมีประสิทธิภาพมากขึ้น อีกทั้งยังสามารถช่วยสร้างโมเดลที่ทำนายความเสี่ยงจากพฤติกรรมการชำระหนี้ของลูกค้าเดิมในอดีตได้ดียิ่งขึ้น ซึ่งการนำแบบจำลองข้อมูลการวิเคราะห์หาความสามารถในการชำระหนี้ของลูกค้าผู้กู้รายเดิมโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยใช้อัลกอริทึมที่มีความถูกต้องแม่นยำมากที่สุดนำมาใช้งาน ถือเป็นการช่วยสนับสนุนเครื่องมือของธนาคารให้มีประสิทธิภาพในการตัดสินใจปล่อยสินเชื่อให้กับลูกค้ารายเดิม จะดีอย่างยิ่งหากธนาคารทราบล่วงหน้าและมีการแจ้งเตือนให้พนักงานระมัดระวังการอนุมัติสินเชื่อให้ลูกค้ากลุ่มความเสี่ยงไหนหรือลูกค้าลักษณะเช่นไร โดยพิจารณาจากความสามารถในการชำระหนี้ของลูกค้าในอดีต ถือเป็นการป้องกันที่ต้นเหตุและช่วยเพิ่มความระมัดระวังในการปล่อยสินเชื่อสำหรับลูกค้ากลุ่มที่มีความเสี่ยงสูง เพื่อไม่ให้เกิดอัตราการผิดนัดชำระหนี้ที่จะเกิดขึ้นในอนาคต นอกจากนี้การมีเครื่องมือที่มีประสิทธิภาพและ

มีความน่าเชื่อถือ สามารถทำให้ธนาคารลดอัตราการเกิดหนี้เสียในกลุ่มลูกค้าเดิมที่มีประวัติเคยขอสินเชื่อที่อาจจะเกิดขึ้นในอนาคตได้เช่นกันและยังสามารถควบคุม ติดตามหนี้ในลูกค้าเดิมกลุ่มนี้ได้

บทที่ 4
ผลการวิจัย

งานวิจัยนี้เป็นงานวิจัยเพื่อศึกษาการเปรียบเทียบประสิทธิภาพการจำแนกความสามารถในการชำระหนี้สำหรับผู้กู้รายเดิมโดยใช้เทคนิคการเรียนรู้ของเครื่อง มีวัตถุประสงค์เพื่อสร้างแบบจำลองข้อมูลในการวิเคราะห์ความสามารถในการชำระหนี้คืน และเพื่อต้องการเปรียบเทียบประสิทธิภาพของแต่ละโมเดล และเลือกใช้โมเดลที่เหมาะสมในวิเคราะห์สามารถในการชำระหนี้คืนจากข้อมูลลูกค้ารายเดิม โดยเทคนิคการเรียนรู้ของเครื่องที่นำมาใช้ในงานวิจัย ได้แก่ Logistic Regression Neural Network Decision Tree และ Random Forest ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอนต่าง ๆ ดังนี้

4.1 ผลการเตรียมข้อมูล

4.2 ผลการพัฒนาแบบจำลอง

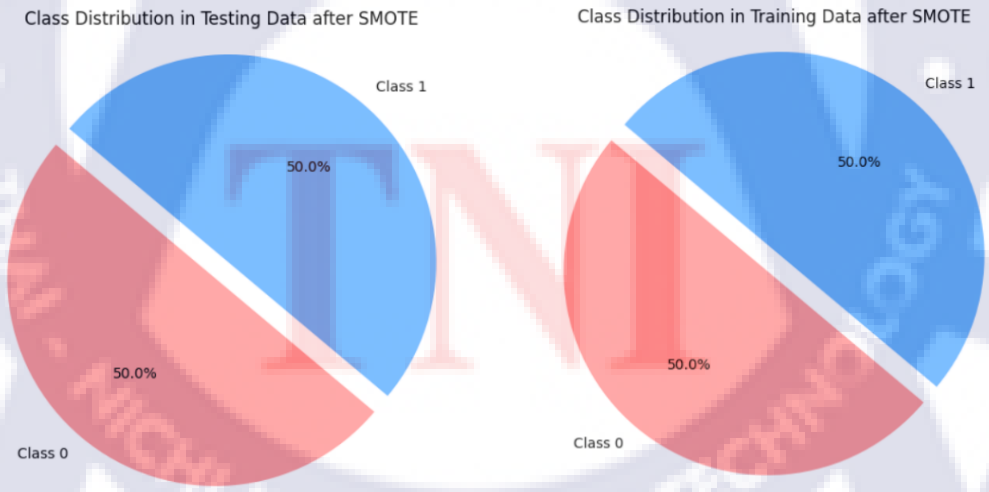
4.3 ผลการเปรียบเทียบประสิทธิภาพของแต่ละโมเดล

4.1 ผลการเตรียมข้อมูล

ข้อมูลที่นำมาใช้ในงานวิจัยเป็นข้อมูลประเมินศักยภาพลูกหนี้ของลูกค้าเดิมของธนาคารแห่งหนึ่ง จำนวน 4,334 ราย การเก็บรวมรวมข้อมูล (Data Collection) เพื่อนำไปใช้วิเคราะห์หาความสามารถในการชำระหนี้คืนของลูกค้ารายเดิม ใช้ข้อมูลในรูปแบบ CSV และ Excel เมื่อได้ข้อมูลที่ต้องการนำไปใช้งาน ขั้นตอนต่อไปเป็นการทำความสะอาดข้อมูล (Data Cleaning) โดยการดูความครบถ้วนสมบูรณ์ของข้อมูล ซึ่งจากข้อมูลพบว่า มีความสมบูรณ์ค่อนข้างสูง เนื่องจากมีจำนวนข้อมูลสูญหายค่อนข้างน้อย การแทนค่าข้อมูลสูญหายจึงใช้วิธีการลบแถวที่มีข้อมูลสูญหายออกไป ในส่วนของข้อมูลที่ซ้ำซ้อน (Duplication Data) จะทำการตรวจสอบและลบข้อมูลที่ซ้ำซ้อนออกไปเช่นกัน เพื่อไม่ให้เกิดปัญหาในขั้นตอนการฝึกโมเดล ต่อไปเป็นขั้นตอนการแปลงข้อมูล (Data Transformation) เป็นการเปลี่ยนแปลงรูปแบบข้อมูลที่จะใช้งานให้อยู่ในรูปแบบ Numeric Data โดยกำหนดเงินต้นค้างในบิล (The Principal Remains on the Bill) เป็นตัวแปรหลัก เนื่องจากการชำระหนี้จะต้องชำระดอกเบี้ยให้เสร็จสิ้นก่อนจะชำระเงินต้นได้หรือหากผู้กู้ไม่มีการชำระหนี้ ดอกเบี้ยก็จะสูงขึ้นจนไม่สามารถชำระต้นเงินได้ หากมีต้นเงินยังคงค้างอยู่ในบิล กล่าวคือ ลูกค้ายังไม่ได้ชำระหนี้คืนเสร็จสิ้น ซึ่งในกรณีนี้ค่าของคอลัมน์ได้กำหนดเป็นตัวแปร Binary โดยการแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มที่มีความสามารถในการชำระหนี้ (กลุ่มที่ผ่านการพิจารณาอนุมัติ) มีค่า 0 หรือ 1 โดยกำหนดให้ 0 แทน ลูกค้าชำระหนี้คืนเสร็จสิ้นแล้ว และ 1 แทน ลูกค้ายังไม่ได้ชำระหนี้คืนเสร็จสิ้น

ขั้นตอนต่อไปทำการแบ่งชุดข้อมูลออกเป็น 2 ชุด คือ ชุดข้อมูลสำหรับฝึกฝน (Training data) จำนวน 3,945 ชุด และชุดข้อมูลสำหรับทดสอบ (Testing data) จำนวน 500 ชุด ซึ่งได้มีการคำนวณสถิติพื้นฐาน เช่น ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าต่ำสุดและค่ามากสุด เพื่อดูลักษณะของข้อมูลเพิ่มเติม เมื่อทำการแบ่งข้อมูลออกเป็นชุดสำหรับฝึกฝนและชุดสำหรับทดสอบแล้ว ขั้นตอนต่อไปทำการ Normalize Data ข้อมูลทั้ง 5 ปัจจัยที่ได้กำหนดไว้ เพื่อปรับสเกลของข้อมูลให้อยู่ในช่วง [0,1] เนื่องจากข้อมูลมีช่วงที่ต่างกันมากเกินไป โดยจำนวนเงินกู้ของลูกค้าแต่ละรายไม่เหมือนกันจึงทำให้ต้นเงินหรือดอกเบี้ยของลูกค้าแต่ละรายไม่เท่ากัน ดอกเบี้ย เบี้ยปรับ ดอกเบี้ยคงค้างในบิลที่เกิดขึ้นจึงมีจำนวนยอดเงินที่ต่างกันเยอะ เพื่อเป็นการป้องกันการเกิดแนวโน้มการเอนเอียงไปทิศทางใดทางหนึ่งของโมเดล การทำ Normalization จะช่วยให้ทุกฟีเจอร์มีผลเท่ากัน และทำให้โมเดลเรียนรู้ได้ดีและมีประสิทธิภาพมากขึ้น

เนื่องด้วยชุดข้อมูลสำหรับฝึกฝนและชุดข้อมูลสำหรับทดสอบยังมีความอสมดุลกันระหว่างชุดข้อมูล เพื่อป้องกันการเกิด Over fitting จึงต้องทำการแก้ปัญหาชุดข้อมูลที่อสมดุลให้มีความสมดุลกัน โดยใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) ในการเพิ่มข้อมูลในคลาสขนาดเล็กโดยสร้างข้อมูลขึ้นมาใหม่จากชุดข้อมูลเดิม ทำให้ข้อมูลในคลาสขนาดเล็กมีการกระจายตัวที่ดีขึ้น และค่าความสำคัญของข้อมูลไม่สูญหาย นอกจากนี้การใช้เทคนิค SMOTE จะช่วยลดปัญหาการเกิด Over fitting ได้

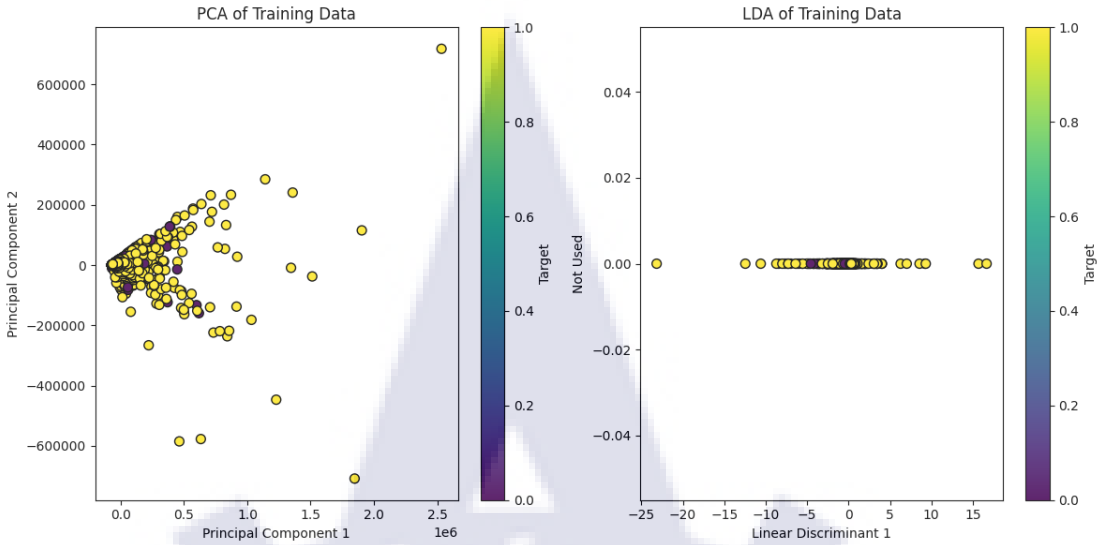


รูปที่ 4.1 ภาพแสดงอัตราส่วนของข้อมูลทั้งหมดหลังจากทำการปรับสมดุลข้อมูล

จากรูปที่ 4.1 ภาพแสดงอัตราส่วนของข้อมูลหลังจากทำการปรับข้อมูลให้สมดุลโดยใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) ในการปรับข้อมูลให้มี

ความสมดุลก่อนการสร้างแบบจำลอง ถือเป็นการเพิ่มความแม่นยำในการจำแนกคลาสที่มีจำนวนน้อย โดยการสร้างตัวอย่างข้อมูลเทียบเท่ากับข้อมูลที่มีอยู่ ซึ่งเป็นการเติมข้อมูลในส่วนที่ว่าง เพื่อให้ข้อมูลมีความครบถ้วนสมบูรณ์และมีความสมดุล หากไม่ทำการปรับข้อมูลให้สมดุลกัน อาจทำให้เกิดความเป็นไปได้ว่าผลลัพธ์จะมีความโน้มเอียงไปทำนายทางใดทางหนึ่งมากกว่าและเพื่อป้องกันการเกิดความเอนเอียงจากการทำนาย ผู้วิจัยเลือกใช้เทคนิค SMOTE มาปรับชุดข้อมูลให้สมดุลกัน ทำให้ข้อมูลที่ได้มีความสมดุลและช่วยปรับปรุงให้โมเดลมีประสิทธิภาพมากขึ้น นอกจากนี้ยังทำให้ข้อมูลแต่ละชุดเกิดการกระจายตัวที่ดีเพิ่มขึ้นและรักษาค่าความสำคัญของข้อมูลที่มีให้คงอยู่ ซึ่งจะช่วยลดปัญหาการเกิด Over fitting ได้

ในส่วนการลดมิติของข้อมูล (Dimensionality Reduction) โดยการคัดเลือกตัวแปรที่มีความสำคัญต่อการวิเคราะห์ โดยเทคนิค Machine Learning ได้มีการเลือกใช้ Recursive Feature Elimination (RFE) ตัวแปรที่เลือกโดย Recursive Feature Elimination (RFE) ประกอบด้วย 5 ปัจจัย ได้แก่ เบี้ยปรับ (Late Charge) หนี้คงค้าง (Outstanding Debt) ดอกเบี้ยคงค้าง (Due Interest) เงินต้นค้างในบิล (The Principal Remains on the Bill) และดอกเบี้ยค้างในบิล (Interest Accrued on the Bill) โดยทั้ง 5 ปัจจัย เป็นปัจจัยที่มีความเกี่ยวเนื่องในการดูประวัติการชำระหนี้ของลูกค้าเดิม และมีความเกี่ยวข้องในการพิจารณาการปล่อยสินเชื่อของธนาคาร การเลือกใช้ RFE ในงานวิจัย เพื่อประเมินความสำคัญของฟีเจอร์และคัดเลือกฟีเจอร์ที่มีผลดีที่สุดในการทำนาย ซึ่งจะช่วยให้ได้ฟีเจอร์ที่มีผลที่ดีที่สุดต่อโมเดล และมีแนวโน้มที่จะช่วยปรับปรุงความแม่นยำของโมเดลได้โดยการลดความซับซ้อนของข้อมูล เน้นเฉพาะฟีเจอร์ที่สำคัญในการทำนาย ในส่วนการลดมิติข้อมูล (Feature Extraction) เป็นการลดจำนวนมิติข้อมูล เพื่อให้ข้อมูลมีความกระชับและง่ายต่อการวิเคราะห์ โดยในงานวิจัยใช้ Principal Component Analysis (PCA) และ Linear Discriminant Analysis (LDA) โดย Principal Component Analysis (PCA) จะมุ่งเน้นที่การหาทิศทางที่มีความแปรผันสูงสุดในข้อมูลโดยรวม มักจะใช้เมื่อต้องการลดมิติของข้อมูลโดยไม่คำนึงการจำแนกกลุ่ม ช่วยรักษาความแปรปรวนของข้อมูลให้มากที่สุด ผลลัพธ์ที่ได้จาก PCA มักจะมีหลาย Principal Components ขึ้นอยู่กับจำนวนฟีเจอร์ที่เลือก (เช่นในที่นี้ 2 มิติ) ข้อมูลในแต่ละ Principal Component อาจจะไม่สอดคล้องกับการแยกกลุ่มข้อมูล และ Linear Discriminant Analysis (LDA) จะมุ่งเน้นที่การหาทิศทางที่ดีที่สุดในการแยกกลุ่มข้อมูล (Classes) จะใช้เมื่อต้องการลดมิติของข้อมูลและต้องการแยกกลุ่มข้อมูล ทำให้ข้อมูลแต่ละคลาสแยกออกจากกันได้ดีที่สุด โดยในกรณีนี้ LDA สร้างจำนวน Linear Discriminants ที่ได้เป็น n_classes - 1 (ในที่นี้เป็น 1 มิติ) ซึ่งแสดงถึงการแยกกลุ่มที่ดีที่สุดในมิติเดียว



รูปที่ 4.2 รูปแสดงผลลัพธ์การวิเคราะห์องค์ประกอบหลัก Principal Component Analysis (PCA) และ Linear Discriminant Analysis (LDA)

จากรูปที่ 4.2 แสดงถึง PCA (Principal Component Analysis) Explained Variance Ratio: [0.92794045 0.06871663] หมายความว่า Principal Component 1 อธิบาย ประมาณ 92.8% ของความแปรผันในข้อมูลทั้งหมด Shape of X_training after PCA: (3945, 2) ข้อมูลถูกลดมิติจากจำนวนฟีเจอร์ต้นฉบับเหลือ 2 มิติ และ Shape of X_testing after PCA: (500, 2) ข้อมูลทดสอบถูกลดมิติจากจำนวนฟีเจอร์ต้นฉบับเหลือ 2 มิติเช่นกัน ในส่วนของ Principal Component 2 อธิบายประมาณ 6.9% ของความแปรผัน และ LDA (Linear Discriminant Analysis) Explained Variance Ratio: [1.0] LDA อธิบาย 100% ของความแปรผันในข้อมูลที่เกี่ยวข้องกับการแยกกลุ่ม Shape of X_training after LDA: (3945, 1) ข้อมูลถูกลดมิติจากจำนวนฟีเจอร์ต้นฉบับเหลือ 1 มิติ และ Shape of X_testing after LDA: (500, 1) ข้อมูลทดสอบถูกลดมิติจากจำนวนฟีเจอร์ต้นฉบับเหลือ 1 มิติ

การจัดเตรียมข้อมูลสำหรับการทดสอบและการฝึกโมเดล โดยการแบ่งข้อมูล (Data Splitting) ใช้ K-Fold Cross Validation โดยกำหนด fold K=5 ในการทำ Cross-Validation Hold-out Validation และ Stratified Sampling เพื่อนำมาเปรียบเทียบค่าความแม่นยำของแต่ละโมเดล ได้ผลลัพธ์ ดังนี้

ตารางที่ 4.1 ตารางแสดงผลลัพธ์ K-Fold Cross Validation

	Mean Accuracy 5 – Fold	Standard Deviation
Logistic Regression	0.806660	0.0130258
Neural Network	0.747311	0.1049324
Decision Tree	0.944195	0.0064239
Random Forest	0.957751	0.0057695

จากตารางที่ 4.1 แสดงผลลัพธ์ในการทำ K-Fold Cross Validation แสดงให้เห็นค่าความแม่นยำเฉลี่ยของแต่ละโมเดล โดยได้กำหนด K=5 ซึ่งข้อมูลจะถูกแบ่งออกเป็น 5 ชุด แต่ละชุดโมเดลจะถูกฝึกและทดสอบกับชุดข้อมูลที่ต่างกัน เพื่อให้โมเดลแสดงค่าความแม่นยำเฉลี่ยของแต่ละโมเดลรวมถึงค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) พบว่า โมเดล Random Forest มีค่าความแม่นยำมากที่สุด คือ 95.77% และมีค่าเบี่ยงเบนมาตรฐานน้อยที่สุด คือ 0.57%

ตารางที่ 4.2 ตารางแสดงผลลัพธ์ Hold-out Validation

	Accuracy
Logistic Regression	0.8079
Neural Network	0.7657
Decision Tree	0.9407
Random Forest	0.9598

จากตารางที่ 4.2 แสดงผลลัพธ์การแบ่งข้อมูลแบบ Hold-out Validation แสดงให้เห็นค่าความแม่นยำของโมเดล Random Forest ให้ค่าความแม่นยำมากที่สุด คือ 95.98% โดยมีโมเดล Decision Tree ที่ให้ค่าความแม่นยำใกล้เคียงกัน อยู่ที่ 94.07% ในส่วนโมเดล Logistic Regression ให้ค่าความแม่นยำอยู่ที่ 80.79% และโมเดลที่ให้ค่าความแม่นยำน้อยสุด คือ Neural Network โดยให้ค่าความแม่นยำอยู่ที่ 76.57%

ตารางที่ 4.3 ตารางแสดงผลลัพธ์ Stratified Sampling

	Stratified K-Fold	Standard Deviation
Logistic Regression	0.806659	0.0049034
Neural Network	0.773776	0.0831185
Decision Tree	0.941169	0.0060185
Random Forest	0.958016	0.0063005

จากตารางที่ 4.3 แสดงผลลัพธ์ในการใช้ Stratified Sampling แสดงให้เห็นว่า โมเดล Random Forest มีค่าความแม่นยำมากที่สุด คือ 95.80% และโมเดล Decision Tree มีค่าเบี่ยงเบนมาตรฐานน้อยที่สุด คือ 0.60%

จากการเปรียบเทียบการใช้ K-Fold Cross Validation Hold-out Validation และ Stratified Sampling พบว่า โมเดล Random Forest และ Decision Tree แสดงผลลัพธ์ที่ดีในทุกวิธีการวัดความแม่นยำ โมเดล Logistic Regression มีความแม่นยำที่ค่อนข้างเสถียรไม่เปลี่ยนแปลงไปมากและโมเดล Neural Network แสดงค่าความแม่นยำที่ต่ำกว่าค่าที่ได้จากเทคนิคอื่น ๆ และมีความแปรปรวนสูง โดยการแบ่งกลุ่มข้อมูลจะทำให้ผลลัพธ์ของแต่ละโมเดลมีความน่าเชื่อถือเพิ่มมากขึ้น มีประสิทธิภาพมากขึ้นและช่วยป้องกันการเกิด Over fitting

4.2 ผลการพัฒนาแบบจำลอง

ผลการพัฒนาแบบจำลองเพื่อนำไปใช้ในการวิเคราะห์จากข้อมูลลูกค้าผู้กู้รายเดิมสำหรับใช้ในการอนุมัติการขอกู้สินเชื่อเพิ่ม โดยการใช้เทคนิคการเรียนรู้ของเครื่อง โมเดลที่นำมาใช้งานวิจัยประกอบด้วย Logistic Regression Neural Network Decision Tree และ Random Forest ซึ่งหลักการในการเลือกใช้แต่ละโมเดลจะมีความแตกต่างกัน โดยที่ Logistic Regression เป็นโมเดลที่เหมาะกับการจัดการข้อมูลที่มีความสัมพันธ์เชิงเส้นระหว่างฟีเจอร์และตัวแปรเป้าหมาย (Target Variable) แบบจำแนกประเภท เหมาะสำหรับกรณีที่ต้องการความเรียบง่ายและตีความที่ชัดเจน ประสิทธิภาพจะเกิดขึ้นได้ดีก็ต่อเมื่อปัญหามีความไม่ซับซ้อน ในส่วนของโมเดล Neural Network เป็นโมเดลที่มีความโดดเด่นในการเรียนรู้ความสัมพันธ์ที่ซับซ้อนระหว่างฟีเจอร์ได้ดี โมเดล Decision Tree และ Random Forest มีลักษณะที่คล้ายกัน โดยโมเดล Decision Tree จะจัดการข้อมูลที่มีความซับซ้อนได้ดี สามารถจัดการกับข้อมูลที่เป็นเชิงลึกและมีความหลากหลายได้ สามารถจัดการกับข้อมูลที่ไม่เป็นเชิงเส้นได้ และโมเดล Random Forest เป็นการสร้างโมเดลต้นไม้

หลาย ๆ ต้นแล้วรวมผลลัพธ์กัน โดยโมเดลนี้เหมาะกับข้อมูลที่มีลักษณะหลายมิติสามารถจัดการกับข้อมูลที่มีหลายฟีเจอร์และความสัมพันธ์ที่ซับซ้อนระหว่างฟีเจอร์ได้

ซึ่งโมเดลทั้ง 4 ที่เลือกใช้ในงานวิจัย ผู้วิจัยได้มีการเปรียบเทียบประสิทธิภาพของแบบจำลองโดยการเปรียบเทียบประสิทธิภาพแบบจำลองของแต่ละโมเดล โดยการนำแต่ละโมเดลมาเปรียบเทียบเพื่อหาค่าความแม่นยำ (Precision) ค่าความอ่อนไหว (Recall) ค่าประสิทธิภาพ (F1-score) และค่าความถูกต้อง (Accuracy) โดยใช้ค่าความถูกต้อง (Accuracy) เป็นหลักแล้วเลือกใช้โมเดลที่ให้ค่าความถูกต้องแม่นยำมากที่สุด เพื่อนำโมเดลไปใช้งานกับข้อมูลต่อไป ผลจากการวิเคราะห์แต่ละโมเดล โดยใช้โปรแกรม Visual Studio Code ได้ผลลัพธ์ ดังนี้

ตารางที่ 4.4 Classifier model โดยใช้ Logistic Regression

Logistic Regression	**0 (Paid)**	**1 (Not Paid)**
Precision	0.64	0.67
Recall	0.69	0.62
F1-Score	0.66	0.65
Accuracy	0.65	

จากตารางที่ 4.4 Classifier model โดยใช้ Logistic Regression ซึ่งกำหนด 1 แทนลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) 67% ค่าความอ่อนไหว (Recall) อยู่ที่ 62% และค่าประสิทธิภาพ (F1-score) 65% และ 0 แทนลูกค้าชำระคืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) 64% ค่าความอ่อนไหว (Recall) 69% และค่าประสิทธิภาพ (F1-score) อยู่ที่ 66% โมเดลสามารถทำนายข้อมูลถูกต้อง (Accuracy) คิดเป็น 65%

ตารางที่ 4.5 Classifier model โดยใช้ Neural Network

Neural Network	**0 (Paid)**	**1 (Not Paid)**
Precision	0.71	0.75
Recall	0.76	0.70
F1-Score	0.73	0.72
Accuracy	0.73	

ตารางที่ 4.5 Classifier model โดยใช้ Neural Network ซึ่งกำหนด 1 แทน ลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) 75% ค่าความอ่อนไหว (Recall) 70% และค่าประสิทธิภาพ (F1-score) 72% และ 0 แทนลูกค้าชำระคืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) คิดเป็น 71% ค่าความอ่อนไหว (Recall) คิดเป็น 76% และค่าประสิทธิภาพ (F1-score) คิดเป็น 73% โมเดลสามารถทำนายข้อมูลถูกต้อง (Accuracy) คิดเป็น 73%

ตารางที่ 4.6 Classifier model โดยใช้ Decision Tree

Decision Tree	0 (Paid)	1 (Not Paid)
Precision	0.94	0.95
Recall	0.95	0.94
F1-Score	0.94	0.94
Accuracy	0.94	

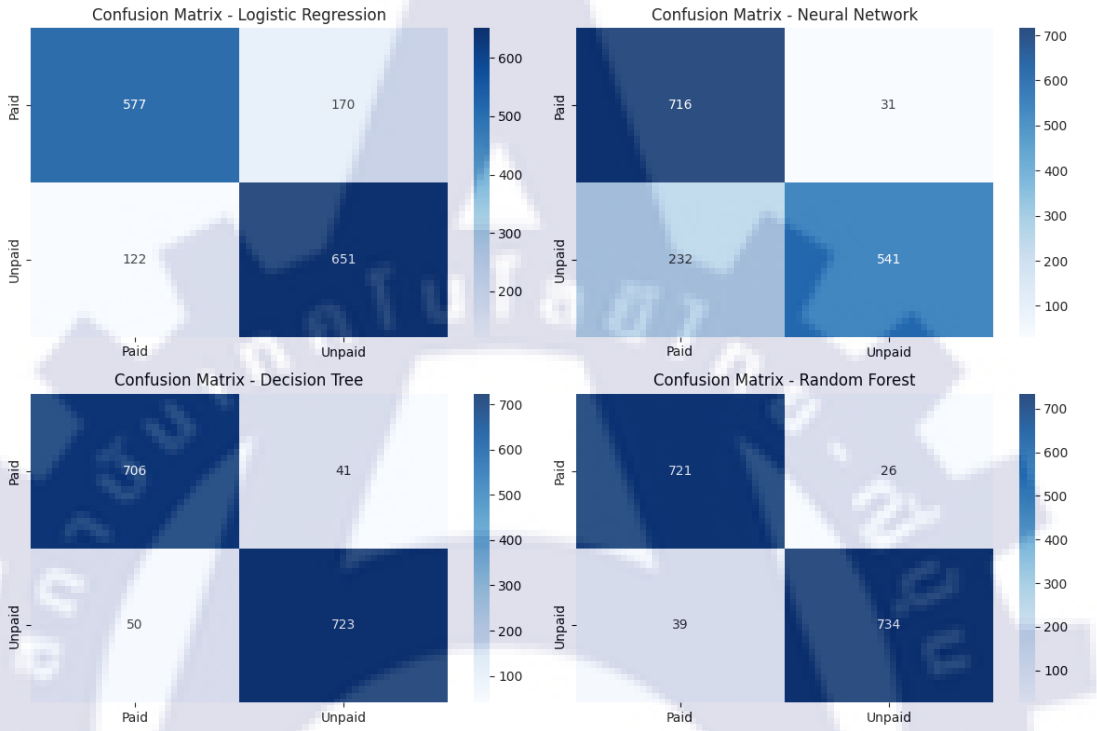
ตารางที่ 4.6 Classifier model โดยใช้ Decision Tree โดยมีโหนดราก (Root Node) คือ ต้นเงินค้างในบิล (The Principal Remains on the Bill) มีผลต่อความสามารถในการชำระหนี้คืน ซึ่งกำหนด 1 แทนลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) คิดเป็น 95% ค่าความอ่อนไหว (Recall) คิดเป็น 94% และค่าประสิทธิภาพ (F1-score) คิดเป็น 94% และ 0 แทนลูกค้าชำระคืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) 94% ค่าความอ่อนไหว (Recall) 95% และค่าประสิทธิภาพ (F1-score) 94% โมเดลสามารถทำนายข้อมูลถูกต้อง (Accuracy) คิดเป็น 94%

ตารางที่ 4.7 Classifier model โดยใช้ Random Forest

Random Forest	0 (Paid)	1 (Not Paid)
Precision	0.94	0.97
Recall	0.97	0.94
F1-Score	0.96	0.96
Accuracy	0.96	

ตารางที่ 4.7 Classifier model โดยใช้ Random Forest โดยมีโหนดราก (Root Node) คือ ต้นเงินค้างในบิล (The Principal Remains on the Bill) มีผลต่อความสามารถในการชำระหนี้คืน ซึ่งกำหนด 1 แทนลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) คิดเป็น 97%

ค่าความอ่อนไหว (Recall) คิดเป็น 94% และค่าประสิทธิภาพ (F1-score) คิดเป็น 96% และ 0 แทนลูกค้าชำระคืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) 94% ค่าความอ่อนไหว (Recall) คิดเป็น 97% และค่าประสิทธิภาพ (F1-score) 96% โมเดลสามารถทำนายข้อมูลถูกต้อง (Accuracy) คิดเป็น 96%



รูปที่ 4.3 ผลลัพธ์การสร้าง Confusion Matrix ของแต่ละโมเดล

จากภาพแสดงให้เห็นผลลัพธ์จากการทำ Confusion Matrix โดยพบว่า โมเดล Random Forest มีการทำนายได้ดีกว่าโมเดลอื่น ๆ ซึ่ง Random Forest ได้มีการทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น) แทนด้วย 0 ทำนายถูกต้อง 721 ราย ทำนายผิดไปเพียง 39 ราย และทำนาย Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) แทนด้วย 1 ทำนายถูกต้อง 734 ราย ทำนายผิดไปเพียง 26 ราย เมื่อเปรียบเทียบกับโมเดล Logistic Regression ทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น) แทนด้วย 0 ทำนายถูกต้อง 577 ราย ทำนายผิด 122 ราย และทำนาย Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) แทนด้วย 1 ทำนายถูกต้อง 651 ราย ทำนายผิด 170 ราย พบว่า การทำนายผิดยังมีค่อนข้างสูงสำหรับเทคนิคนี้ ในส่วนของโมเดล Neural Network ทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น) แทนด้วย 0 ทำนายถูกต้อง 716 ราย ทำนายผิด 232 ราย และทำนาย Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) แทนด้วย 1 ทำนายถูกต้อง 541 ราย ทำนายผิด 31 ราย พบว่า มีการทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น)

ถูกต้องสูงกว่าการทำนาย Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) และโมเดล Decision Tree มีการทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น) แทนด้วย 0 ทำนายถูกต้อง 706 ราย ทำนายผิดไปเพียง 50 ราย และทำนาย Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) แทนด้วย 1 ทำนายถูกต้อง 723 ราย ทำนายผิดไปเพียง 41 ราย พบว่า Decision Tree มีการทำนาย Paid (ชำระหนี้คืนเสร็จสิ้น) และ Unpaid (ชำระหนี้คืนยังไม่เสร็จสิ้น) ที่ถูกต้องสูงรองลงมาจากโมเดล Random Forest จากการทำ Confusion Matrix สรุปได้ว่า โมเดล Random Forest มีการคัดแยกประเภทและทำนายได้ดีกว่า มีความแม่นยำกว่าอย่างเห็นได้ชัด

4.3 ผลการเปรียบเทียบประสิทธิภาพของแต่ละโมเดล

จากการประเมินประสิทธิภาพของแต่ละโมเดล โดยทำการทดสอบโมเดล Logistic Regression Neural Network Decision Tree และ Random Forest ด้วย K-Fold Cross Validation และแสดงค่าความแม่นยำเฉลี่ยของแต่ละโมเดล โดยได้กำหนด K=5 ซึ่งข้อมูลจะถูกแบ่งเป็น 5 ชุด แต่ละชุดโมเดลจะถูกฝึกและทดสอบกับชุดข้อมูลที่ต่างกัน เพื่อให้โมเดลแสดงค่าความแม่นยำเฉลี่ยของแต่ละโมเดลรวมถึงค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าเฉลี่ยความคลาดเคลื่อนในการทำนาย (RMSE) ได้ผลการทดสอบ ดังนี้

1. โมเดล Logistic Regression มีค่าความแม่นยำเฉลี่ย 67.08% ค่า Standard Deviation 1.49% ค่า MAE 34.60% และค่า RMSE 58.82% กล่าวคือ โมเดลมีการทำนายถูกต้องน้อย การกระจายตัวของข้อมูลน้อย มีค่าคลาดเคลื่อนสูง

2. โมเดล Neural Network มีค่าความแม่นยำเฉลี่ย 71.84% แต่มีค่า Standard Deviation สูงที่สุดจากโมเดลอื่น ๆ อยู่ที่ 2.42% ค่า MAE 27.23% และค่า RMSE 52.18% กล่าวคือ โมเดลมีค่าความแปรปรวนสูงและมีค่าคลาดเคลื่อนสูง ทำให้การทำนายไม่ถูกต้องแม่นยำตามที่คาดหวัง

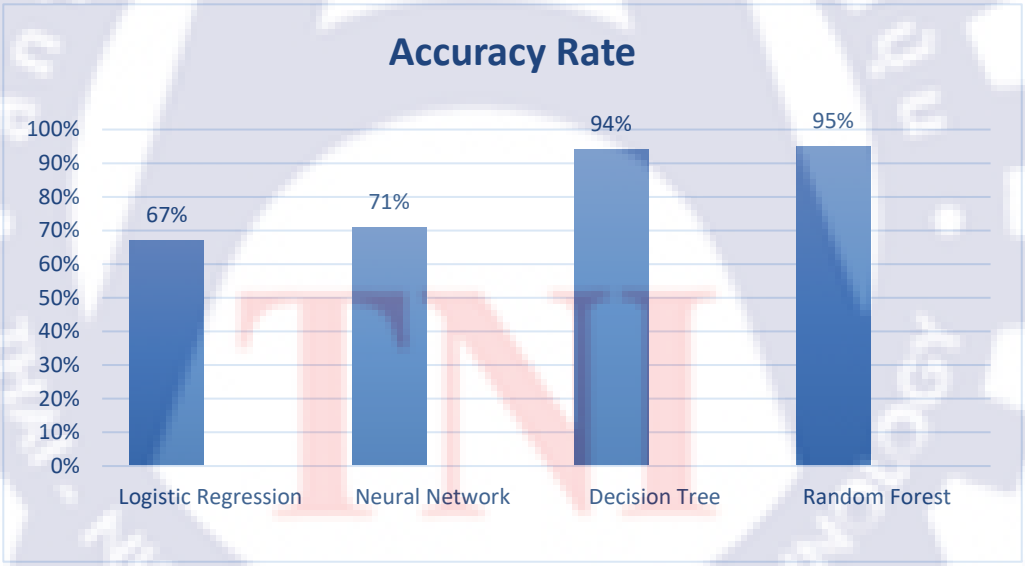
3. โมเดล Decision Tree มีค่าความแม่นยำเฉลี่ยสูงถึง 94.18% โดยมีค่า Standard Deviation 0.45% ค่า MAE 5.92% และค่า RMSE 24.33% กล่าวคือ โมเดลมีความแม่นยำสูงและความแปรปรวนต่ำ ดังนั้นอาจเป็นตัวเลือกที่ดีในการทำนาย

4. โมเดล Random Forest มีค่าความแม่นยำเฉลี่ยสูงถึง 95.91% โดยมีค่า Standard Deviation เพียง 0.41% ค่า MAE 4.40% และค่า RMSE 20.99% กล่าวคือ โมเดล Random Forest มีความแม่นยำสูงที่สุดจากการทดสอบ และมีความเสถียรที่ดีในการทำนาย โดยไม่มีความแปรปรวนมาก

ตารางที่ 4.8 ตารางแสดงผลการเปรียบเทียบโมเดล

ผลเปรียบเทียบ	Accuracy	Standard Deviation	MAE	RMSE
Logistic Regression	0.6708	0.0149	0.3460	0.5882
Neural Network	0.7184	0.0242	0.2723	0.5218
Decision Tree	0.9418	0.0045	0.0592	0.2433
Random Forest	0.9591	0.0041	0.0440	0.2099

จากตารางที่ 4.8 พบว่า ตารางเปรียบเทียบโมเดล Logistic Regression Neural Network Decision Tree และ Random Forest พิจารณาได้ว่า โมเดลที่ถูกสร้างขึ้นจากการใช้ Random Forest มีค่าความถูกต้องมากที่สุดเท่ากับ 0.95 ค่าความเบี่ยงเบนมาตรฐานเท่ากับ 0.004 ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) เท่ากับ 0.04 และค่าเฉลี่ยความคลาดเคลื่อนในการทำนาย (RMSE) เท่ากับ 0.20 ซึ่งมากกว่าโมเดลที่ถูกสร้างขึ้นจากการใช้ Logistic Regression Neural Network และ Decision Tree ซึ่งมีค่าความถูกต้องอยู่ที่ 0.67 0.71 และ 0.94 ตามลำดับ



รูปที่ 4.4 แผนภูมิแสดงการเปรียบเทียบค่าความถูกต้อง

จากรูปที่ 4.4 แผนภูมิแสดงการเปรียบเทียบค่าความถูกต้อง (Accuracy) ของแต่ละโมเดล จากการศึกษาพบว่า โมเดล Random Forest มีค่าความถูกต้องและเสถียรมากที่สุดจากโมเดลที่นำมาทดสอบทั้งหมด ซึ่งโมเดล Random Forest มีค่าความถูกต้องแม่นยำอยู่ที่ 95.91%

โดยมีความใกล้เคียงกับโมเดล Decision Tree มากที่สุด โดยโมเดล Decision Tree มีค่าความถูกต้องแม่นยำที่ 94.18% ส่วนโมเดล Neural Network มีค่าความถูกต้องแม่นยำอยู่ที่ 71.84% และในส่วนของโมเดล Logistic Regression มีค่าความถูกต้องแม่นยำน้อยที่สุดจากโมเดลที่ทำการทดสอบทั้งหมด ซึ่งมีค่าความถูกต้องน้อยสุด คือ 67.08%

เมื่อได้โมเดล Random Forest ซึ่งมีค่าความถูกต้องแม่นยำมากที่สุดจากโมเดลทั้งหมด โดยจะใช้ Grid Search สำหรับการปรับค่าพารามิเตอร์ (Hyper parameter Tuning) เพื่อเพิ่มประสิทธิภาพของโมเดลให้ดียิ่งขึ้น มีความถูกต้องแม่นยำยิ่งขึ้น ซึ่งการปรับค่าโดยใช้ Grid Search เป็นวิธีการค้นหาค่าที่ดีที่สุดสำหรับโมเดลโดยการทดสอบทุกค่าที่เป็นไปได้จากชุดของพารามิเตอร์ที่ได้มีการกำหนดไว้ล่วงหน้า จากนั้นจะเลือกค่าพารามิเตอร์ที่ทำให้โมเดลมีประสิทธิภาพสูงสุด จากการปรับพารามิเตอร์ของโมเดล Random Forest เพื่อให้ได้พารามิเตอร์ที่ดีที่สุด ผลลัพธ์ที่ได้จากปรับค่าพารามิเตอร์ เป็นดังนี้ max_depth': 40 'min_samples_leaf': 1 'min_samples_split': 2 'n_estimators': 200 จากการปรับพารามิเตอร์ พบว่า max_depth หากตั้งค่าที่สูงมากอาจทำให้ต้นไม้เรียนรู้ข้อมูลฝึกอบรมได้มากเกินไป (Over fitting) ขณะเดียวกันหากตั้งค่าที่ต่ำเกินไปอาจทำให้ต้นไม้เรียนรู้ข้อมูลได้น้อยเกินไป (Under fitting) ในส่วน min_samples_leaf เป็นการควบคุมความละเอียดของการแยกข้อมูลในแต่ละโหนด หากตั้งค่าค่าที่สูงจะทำให้มีโหนดที่เพิ่มมากขึ้น อาจช่วยลดการเกิด Over fitting ได้ min_samples_split เป็นการควบคุมจำนวนขั้นต่ำของตัวอย่างที่ต้องมีในโหนดก่อนที่จะทำการแยกออกเป็นสองโหนด การตั้งค่าค่าที่สูงจะทำให้การแยกโหนดเกิดขึ้นน้อยลง ซึ่งอาจช่วยลดการเกิด Over fitting และ n_estimators เป็นการกำหนดจำนวนต้นไม้ที่ใช้ใน Random Forest การเพิ่มจำนวนซึ่งจะช่วยให้โมเดลมีความแม่นยำสูงขึ้น ขนาดที่เหมาะสมสามารถช่วยลด Variance ของโมเดลและปรับปรุงผลลัพธ์ให้ดีมากขึ้น โดยได้คะแนนค่าความแม่นยำ (Accuracy) ที่ดีที่สุดจาก Grid Search คือ 0.9598

โดยรวมแล้วการปรับพารามิเตอร์ช่วยให้ Random Forest สามารถทำงานได้ดีขึ้นกับข้อมูลที่มีและเพิ่มประสิทธิภาพในการคัดแยกประเภทหรือการทำนาย จะเห็นว่าจากการใช้ Grid Search ได้พารามิเตอร์ที่เหมาะสมสำหรับโมเดล Random Forest ทำให้ได้ผลลัพธ์ที่ดีขึ้นจากการทดสอบเดิม โดยค่าความแม่นยำ (Accuracy) เพิ่มขึ้นเป็น 95.98% จากเดิมอยู่ที่ 95.91% เป็นการเพิ่มขึ้นจากเดิมเล็กน้อยแบบไม่มีนัยสำคัญ

บทที่ 5
ข้อสรุปและเสนอแนะ

งานวิจัยฉบับนี้มีความสนใจที่อยากจะหาเครื่องมือที่สามารถเข้ามาวิเคราะห์ความสามารถในการชำระหนี้จากข้อมูลผู้กู้รายเดิมโดยใช้ข้อมูลลูกค้าในอดีตมาวิเคราะห์ข้อมูลในอนาคต เพื่อช่วยธนาคารประกอบการตัดสินใจในการอนุมัติสินเชื่อ โดยมีวัตถุประสงค์เพื่อสร้างแบบจำลองข้อมูลการในการวิเคราะห์ความสามารถในการชำระหนี้คืนโดยใช้เทคนิคการเรียนรู้ของเครื่อง และเพื่อเปรียบเทียบประสิทธิภาพของเทคนิคที่เหมาะสมในการจำแนกความสามารถในการชำระหนี้คืนจากข้อมูลลูกค้าผู้กู้รายเดิม ในขั้นตอนการพัฒนาตัวแบบ การจำแนกจะแบ่งข้อมูลออกเป็นข้อมูลเรียนรู้ (Training Data) และชุดข้อมูลทดสอบ (Testing Data) ด้วยหลักการ 5-Fold จากนั้นนำชุดข้อมูลมาใช้ในการสร้างตัวแบบการจำแนกโดยใช้โมเดล Logistic Regression Neural Network Decision Tree และ Random Forest

งานวิจัยนี้มีวัตถุประสงค์หลักเพื่อพัฒนาเครื่องมือวิเคราะห์ความสามารถในการชำระหนี้ของผู้กู้ โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งสามารถนำไปประยุกต์ใช้กับระบบการอนุมัติสินเชื่อของธนาคาร ผลการวิจัยช่วยเพิ่มความแม่นยำในการทำนายความเสี่ยงจากข้อมูลผู้กู้ในอดีต ซึ่งนอกจากจะลดปัญหาหนี้เสียที่อาจเกิดขึ้นแล้ว ยังช่วยให้กระบวนการอนุมัติสินเชื่อมีประสิทธิภาพมากยิ่งขึ้น ในบทนี้จะสรุปผลการวิจัย พร้อมทั้งอภิปรายและนำเสนอข้อเสนอแนะสำหรับการพัฒนางานวิจัยในอนาคต

5.1 สรุปผลการวิจัย

จากการเปรียบเทียบประสิทธิภาพของเทคนิคเพื่อหาเทคนิคที่เหมาะสมกับการจำแนกความสามารถในการชำระหนี้คืนจากข้อมูลลูกค้าผู้กู้รายเดิม โดยการนำโมเดล Logistic Regression Neural Network Decision Tree และ Random Forest มาทดสอบกับกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) ที่ได้เตรียมไว้จำนวน 3,945 ชุด เพื่อสร้างโมเดลในโปรแกรม Visual Studio Code และนำโมเดลที่ให้ผลค่าความถูกต้องมากที่สุดไปใช้ในการทดสอบกับชุดข้อมูลสำหรับทดสอบ (Testing Data) จำนวน 500 ชุด เพื่อหาโมเดลที่มีค่าความถูกต้องแม่นยำสูงสุดในการวิเคราะห์ความสามารถในการชำระหนี้คืนของผู้กู้รายเดิม ผลที่ได้จากการศึกษาสามารถสรุปผลได้ดังนี้

5.1.1 โมเดล Random Forest โดยได้มีการกำหนดให้ตัวแปรต้นเงินค้างในบิล (The Principal Remains on the Bill) เป็นตัวแปรที่มีผลต่อความสามารถในการชำระหนี้คืนมากที่สุด มีค่าความแม่นยำสูงที่สุด (Accuracy) คิดเป็น 95.98% ค่าเบี่ยงเบนมาตรฐานคิดเป็น 0.41%

ค่า MAE (Mean Absolute Error) ที่ 4.40% และค่า RMSE (Root Mean Square Error) ที่ 20.99% ดังนั้นโมเดลที่ได้จึงเหมาะสมที่จะเอาไปใช้ช่วยสนับสนุนการวิเคราะห์ความสามารถในการชำระหนี้คืนของลูกค้าผู้กู้รายเดิมที่เคยมีประวัติการขอสินเชื่อจากสถาบันการเงินเดิมจากการแบ่งกลุ่มลูกค้าโดยแทนลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า มีค่าความแม่นยำ (Precision) คิดเป็น 97% ค่าความอ่อนไหว (Recall) คิดเป็น 94% และค่าประสิทธิภาพ (F1-score)คิดเป็น 96% และกลุ่มลูกค้าชำระคืนเสร็จสิ้น พบว่า มีค่าความแม่นยำ (Precision) คิดเป็น 94%ค่าความอ่อนไหว (Recall) คิดเป็น 97% และค่าประสิทธิภาพ (F1-score) คิดเป็น 96%

5.1.2 โมเดล Decision Tree เป็นโมเดลที่มีความแม่นยำสูงลำดับถัดมา โดยมีค่าความแม่นยำที่ 94.18% ค่าเบี่ยงเบนมาตรฐานอยู่ที่ 0.45% ค่า MAE (Mean Absolute Error) อยู่ที่ 5.92% และค่า RMSE (Root Mean Square Error) ที่ 24.33% แสดงถึงโมเดลมีความแม่นยำและมีความคลาดเคลื่อนน้อย จากการแบ่งกลุ่มลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) คิดเป็น 95% ค่าความอ่อนไหว (Recall) คิดเป็น 94% และค่าประสิทธิภาพ (F1-score) คิดเป็น 94% และกลุ่มแทนลูกค้าชำระคืนเสร็จสิ้น พบว่า มีค่าความแม่นยำ (Precision) คิดเป็น 94% ค่าความอ่อนไหว (Recall) คิดเป็น 95% และค่าประสิทธิภาพ (F1-score) คิดเป็น 94%

5.1.3 โมเดล Neural Network เป็นโมเดลที่มีความแม่นยำและในภาพรวม ซึ่งถือเป็นโมเดลที่มีความเสถียรในการทำนายที่ดี โดยโมเดล Neural Network มีค่าความแม่นยำเฉลี่ยอยู่ที่ 71.84% มีค่าเบี่ยงเบนมาตรฐาน 2.42% ค่า MAE (Mean Absolute Error) อยู่ที่ 27.23% และค่า RMSE (Root Mean Square Error) สูงถึง 52.18% จากการแบ่งลูกค้าออกเป็นกลุ่มลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น พบว่า ค่าความแม่นยำ (Precision) คิดเป็น 75% ค่าความอ่อนไหว (Recall) คิดเป็น 70% และค่าประสิทธิภาพ (F1-score) คิดเป็น 72% และกลุ่มลูกค้าที่แทนด้วยลูกค้าชำระคืนเสร็จสิ้น พบว่า มีค่าความแม่นยำ (Precision) คิดเป็น 71% ค่าความอ่อนไหว (Recall) คิดเป็น 76% และค่าประสิทธิภาพ (F1-score) คิดเป็น 73%

5.1.4 โมเดล Logistic Regression เป็นโมเดลที่มีความแม่นยำน้อยที่สุดเมื่อเปรียบเทียบกับโมเดลอื่น ๆ ที่เลือกใช้ในงานวิจัย โดยโมเดล Logistic Regression มีค่าความแม่นยำเฉลี่ยเพียง 67.08% ค่าเบี่ยงเบนมาตรฐานที่ 1.49% ค่า MAE (Mean Absolute Error) สูงถึง 34.60% และค่า RMSE (Root Mean Square Error) ที่สูงถึง 58.82% ถือเป็นโมเดลที่ให้ค่าความแม่นยำน้อยที่สุดจากโมเดลที่ได้ทำการทดสอบทั้งหมด โดยจากการแบ่งกลุ่มลูกค้ายังไม่ชำระหนี้คืนเสร็จสิ้น มีค่าความแม่นยำ (Precision) คิดเป็น 67% ค่าความอ่อนไหว (Recall) คิดเป็น 62% และค่าประสิทธิภาพ (F1-score) คิดเป็น 65% และกลุ่มลูกค้าชำระคืนเสร็จสิ้น มีค่าความแม่นยำ (Precision) คิดเป็น 64% ค่าความอ่อนไหว (Recall) คิดเป็น 62% และค่าประสิทธิภาพ (F1-score) คิดเป็น 65%

จากผลการทดสอบจากโมเดลที่เลือกใช้ พบว่า โมเดล Decision Tree และ Random Forest มีความแม่นยำสูงที่สุดและมีความเสถียรที่ดีในการทำนาย ส่วนโมเดล Logistic Regression มีค่าความถูกต้องที่ค่อนข้างต่ำเมื่อเทียบกับโมเดลอื่น ๆ แต่มีค่า MAE และ RMSE สูง ซึ่งแสดงว่ามีความคลาดเคลื่อนในการทำนายที่ค่อนข้างมาก ในขณะที่ Neural Network มีค่าความถูกต้องสูงกว่า Logistic Regression แต่ยังมีค่า RMSE และ MAE ที่สูงกว่าเมื่อเทียบกับ Decision Tree และ Random Forest ซึ่งแสดงถึงความแม่นยำในการทำนายที่น้อยกว่า

ปัจจุบันสถาบันการเงินยังคงใช้เครื่องมือ Credit Scoring Model โดยการใช้แบบจำลองทางคณิตศาสตร์และสถิติ (Statistical and mathematical programming) เช่น Logistic Regression ซึ่งเป็นวิธีที่ได้รับความนิยมนำใช้งานในสถาบันการเงินสำหรับการประเมินความน่าเชื่อถือและความสามารถในการชำระหนี้ของผู้ขอสินเชื่อ เนื่องจาก Logistic Regression สามารถอธิบายผลของการตอบอนุมัติการขอสินเชื่อได้ง่ายและชัดเจน แต่ยังมีข้อด้อย คือ ข้อมูลสินเชื่อมีลักษณะไม่ใช่เส้นตรงจึงทำให้ประสิทธิภาพโดยรวมด้อยกว่า หากเปรียบเทียบกับเทคนิค Machine Learning เนื่องจากเหมาะกับการวิเคราะห์ข้อมูลที่ไม่ใช่เส้นตรงมากกว่า หากสถาบันการเงินนำ Machine Learning เข้ามาใช้งานในการพิจารณาอนุมัติสินเชื่อ โดยการสร้างโมเดลที่ทำนายความเสี่ยงจากพฤติกรรมการชำระหนี้ของลูกหนี้จากในอดีต ซึ่งจะช่วยให้การทำนายความสามารถในการชำระหนี้ในอนาคตมีประสิทธิภาพและมีความแม่นยำมากยิ่งขึ้น

จากงานวิจัยสรุปผลได้ว่า โมเดล Random Forest มีความเหมาะสมที่จะนำไปใช้ในการพิจารณาอนุมัติสินเชื่อ โดยวัดจากค่าประสิทธิภาพความแม่นยำสูงสุดจากการสร้างโมเดลที่ทำนายจากพฤติกรรมการชำระหนี้ของผู้กู้รายเดิมในอดีต จะช่วยส่งผลให้การปล่อยอนุมัติสินเชื่อของธนาคารมีประสิทธิภาพและมีความแม่นยำมากยิ่งขึ้น

5.2 ปัญหาและข้อจำกัด

เนื่องจากชุดข้อมูลที่นำมาใช้ในงานวิจัยมีค่อยข้างน้อยและจำกัด ซึ่งข้อมูลที่นำมาใช้ในงานวิจัยเป็นข้อมูลส่วนบุคคล (Personal Data) ทำให้การได้มาซึ่งข้อมูลจำนวนมากเป็นไปได้ค่อนข้างยาก อาจจะส่งให้เกิดการหลุดไหลของข้อมูลและเผยแพร่ข้อมูลไปยังสาธารณะได้ ผู้วิจัยจึงเลือกเก็บรวบรวมข้อมูลส่วนบุคคลทั่วไปเท่าที่จำเป็นสำหรับใช้ในงานวิจัยเท่านั้น

ข้อจำกัดที่สำคัญในงานวิจัยนี้ คือ จำนวนชุดข้อมูลที่ค่อนข้างน้อย ซึ่งอาจส่งผลให้ผลลัพธ์มีความแม่นยำต่ำกว่าที่คาดไว้ หากมีข้อมูลเพิ่มเติม โดยเฉพาะข้อมูลจากแหล่งอื่น ๆ หรือข้อมูลที่เกี่ยวข้องกับปัจจัยทางเศรษฐกิจและสังคมที่กว้างขึ้น ผลลัพธ์อาจมีความน่าเชื่อถือมากยิ่งขึ้น อีกประการหนึ่ง คือ โมเดลบางประเภท เช่น Neural Network อาจให้ผลลัพธ์ที่ซับซ้อนและต้องการการปรับจูนโมเดลมากขึ้น เพื่อให้ได้ผลการทำนายที่แม่นยำสูงสุด ทั้งนี้งานวิจัยนี้ยังไม่ได้ศึกษา

ปัจจัยภายนอก เช่น สถานการณ์เศรษฐกิจโลกหรือการเปลี่ยนแปลงของตลาด ซึ่งอาจส่งผลต่อความสามารถในการชำระหนี้ของผู้กู้ด้วยเช่นกัน

5.3 การนำไปประยุกต์ใช้

ผลการวิจัยนี้สามารถนำไปประยุกต์ใช้กับระบบการอนุมัติสินเชื่อของธนาคารได้โดยตรง โดยโมเดล Random Forest เป็นโมเดลที่มีความแม่นยำสูง สามารถช่วยให้ธนาคารคัดกรองลูกค้าที่มีความเสี่ยงต่ำและสูงได้อย่างมีประสิทธิภาพ นอกจากนี้การใช้เทคนิค Machine Learning ยังช่วยให้ธนาคารสามารถลดการเกิดหนี้เสียและเพิ่มความมั่นใจในการตัดสินใจอนุมัติสินเชื่อ ผลการวิจัยชี้ให้เห็นว่า การปรับใช้เทคนิคนี้จะช่วยเพิ่มความรวดเร็วและความแม่นยำในการประเมินลูกค้าของธนาคาร ซึ่งเป็นสิ่งสำคัญในการแข่งขันในตลาดสินเชื่อที่มีความซับซ้อนมากขึ้นในอนาคต

นอกจากนี้ผลการวิจัยยังแสดงให้เห็นว่า การใช้เทคนิค Machine Learning ในการพยากรณ์ความสามารถในการชำระหนี้ของผู้กู้ มีศักยภาพสูงกว่าวิธีการดั้งเดิม เช่น Logistic Regression ซึ่งปัจจุบันยังคงใช้ในระบบ Credit Scoring Model การนำ Machine Learning เข้ามาใช้งานจะทำให้ธนาคารสามารถประเมินความเสี่ยงที่ซับซ้อนมากขึ้นได้ เช่น การพิจารณาพฤติกรรมการชำระหนี้ที่ไม่เป็นเชิงเส้นตรง หรือการประเมินผู้กู้ที่มีประวัติสินเชื่อไม่สมบูรณ์ ในทางปฏิบัติการนำโมเดลเหล่านี้ไปใช้จะช่วยให้กระบวนการพิจารณาสินเชื่อมีความรวดเร็วและแม่นยำมากขึ้น ลดขั้นตอนการประเมินด้วยมือ และช่วยให้ธนาคารสามารถจัดสรรทรัพยากรได้อย่างมีประสิทธิภาพมากขึ้น นอกจากนี้ยังสามารถนำมาใช้ในการวางแผนทางการเงินและจัดการพอร์ตสินเชื่อของธนาคารในเชิงกลยุทธ์ โดยโมเดลสามารถช่วยให้ธนาคารทำนายพฤติกรรมของลูกค้าหลังการอนุมัติสินเชื่อ ทำให้สามารถจัดการกับความเสี่ยงที่อาจเกิดขึ้นล่วงหน้า ตัวอย่างการนำไปประยุกต์ใช้เพิ่มเติมคือ ธนาคารสามารถนำโมเดลนี้ไปปรับใช้ในการพิจารณาสินเชื่อสำหรับกลุ่มลูกค้ารายย่อย เช่น การปล่อยสินเชื่อส่วนบุคคลหรือสินเชื่อบ้าน โดยระบบที่ขับเคลื่อนด้วย Machine Learning สามารถประเมินความสามารถในการชำระหนี้ของลูกค้าได้แบบเรียลไทม์ ช่วยให้ธนาคารสามารถพิจารณาอนุมัติสินเชื่อได้ภายในเวลาอันสั้น ซึ่งนอกจากจะลดภาระงานของพนักงานแล้ว ยังสามารถเพิ่มความพึงพอใจของลูกค้าได้ด้วย

ยิ่งไปกว่านั้นโมเดล Machine Learning ยังสามารถประยุกต์ใช้ในระบบการจัดการความเสี่ยงเชิงกลยุทธ์ของธนาคาร เช่น การคาดการณ์แนวโน้มการเกิดหนี้เสียในระดับพอร์ตสินเชื่อทั้งหมดของธนาคาร ซึ่งจะช่วยให้ธนาคารสามารถเตรียมพร้อมรับมือกับปัญหาทางการเงินที่อาจเกิดขึ้นได้ล่วงหน้า โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงทางเศรษฐกิจหรือปัจจัยภายนอกอื่น ๆ ซึ่งจากที่กล่าวมาการนำผลการวิจัยนี้ไปใช้ในเชิงปฏิบัติจะช่วยเสริมความสามารถในการแข่งขันของธนาคารในตลาดสินเชื่อที่มีความท้าทายและซับซ้อนมากยิ่งขึ้น

5.4 ข้อเสนอแนะ

เพื่อให้ระบบสนับสนุนความสามารถในการชำระหนี้คืนให้มีประสิทธิภาพมากขึ้น ทางผู้วิจัยเห็นควรได้รับการพัฒนา ดังต่อไปนี้

5.4.1 ศึกษาและประยุกต์การใช้โมเดลอื่น ๆ ที่เป็นการแบ่งกลุ่มเหมือนกันเพื่อช่วยสนับสนุนความสามารถในการชำระหนี้คืน

5.4.2 ข้อมูลที่ใช้ในการวิเคราะห์มีจำนวนน้อยเกินไป ดังนั้นในอนาคตควรมีการรวบรวมข้อมูลเพื่อทำการวิเคราะห์อีกครั้ง เพื่อให้ได้ผลของการวิเคราะห์ข้อมูลมีความสมบูรณ์มากขึ้น

5.4.3 การนำข้อมูลต่าง ๆ มาใช้ในการดำเนินการวิจัยเป็นเพียงบางส่วนเท่านั้น ในการศึกษาครั้งต่อไปอาจมีการเพิ่มข้อมูลเพื่อให้ครอบคลุมปัจจัยด้านต่าง ๆ ที่เกี่ยวข้องมากขึ้น ซึ่งจะช่วยให้การพิจารณาการปล่อยสินเชื่อมีประสิทธิภาพมากขึ้น

5.5 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

ในการวิจัยครั้งต่อไปควรพิจารณาใช้ข้อมูลที่มีความหลากหลายและครบถ้วนมากขึ้น รวมถึงการรวมปัจจัยภายนอก เช่น ข้อมูลเศรษฐกิจมหภาคและพฤติกรรมผู้บริโภคที่เกี่ยวข้องกับการใช้สินเชื่อ เช่น อัตราดอกเบี้ย อัตราเงินเฟ้อ หรือข้อมูลทางสังคมและการเมืองที่อาจมีผลกระทบต่อความสามารถในการชำระหนี้ของผู้กู้ เป็นต้น เพื่อให้โมเดลที่สร้างขึ้นสามารถทำนายความเสี่ยงได้แม่นยำมากยิ่งขึ้น

นอกจากนี้การประยุกต์ใช้ Deep Learning หรือ Ensemble Learning อาจช่วยเพิ่มประสิทธิภาพในการคาดการณ์ของโมเดลได้ดียิ่งขึ้น โดยเฉพาะในกรณีที่ข้อมูลมีความซับซ้อนสูง การนำเทคนิคใหม่ ๆ มาใช้ในการศึกษาจะช่วยเปิดโอกาสให้ได้ผลลัพธ์ที่ครอบคลุมและมีประโยชน์ต่ออุตสาหกรรมการเงินมากขึ้น

บรรณานุกรม

บรรณานุกรม

[1] วเรศ อุปปาติก, *เศรษฐศาสตร์การเงินและการธนาคาร*, กรุงเทพฯ: สมาคมเศรษฐศาสตร์ธรรมศาสตร์, 2539.

[2] ศุภภร อิ่มสุข, "*ปัจจัยที่มีอิทธิพลต่อการผิดนัดชำระหนี้ของลูกหนี้สินเชื่อสถาบันบริหารจัดการธนาคารที่ดิน (องค์การมหาชน),*" การค้นคว้าอิสระ ศศ.ม. (เศรษฐศาสตร์ธุรกิจ), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2561.

[3] ชุมพร ฉ่ำแสงและคณะ, "*ปัจจัยที่มีผลต่อคุณภาพชีวิตของบุคลากรฝ่ายการพยาบาลศูนย์การแพทย์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี จังหวัดนครนายก,*" การค้นคว้าอิสระ พท.บ. (แพทยศาสตร์), มหาวิทยาลัยศรีนครินทรวิโรฒ, นครนายก, 2555.

[4] G. V. Padilla and M. M. Grant, "Quality of life as a cancer nursing outcome variable," *Advances in Nursing Science,* vol. 8, no. 1, pp. 45-60, October 1985.

[5] ทิพย์วัลย์ เรืองขจร, *วิทยาศาสตร์เพื่อคุณภาพชีวิต*, สงขลา: มหาวิทยาลัยราชภัฏสงขลา, 2554.

[6] พิณณิน กิติพราภรณ์, "นานาทรรศนะธุรกิจเพื่อคุณภาพชีวิตของใคร," *วารสารเศรษฐกิจและบริหารธุรกิจ*, ปีที่ 15, ฉบับที่ 1, หน้า 42-46, มกราคม 2532.

[7] ศิริ ฮามสุโพธิ์, *ประชากรกับการพัฒนาคุณภาพชีวิต*, เชียงใหม่: โอ.เอส.พริ้นติ้งเฮ้าส์, 2543.

[8] M.J. Power and the WHOQOL Group, *The Universality of Quality of Life: An Empirical Approach Using the WHOQOL,* Netherland: Kluwer Academic Publishers, 2022.

[9] นิศารัตน์ ศิลปเดช, *ประชากรกับการพัฒนาคุณภาพชีวิต*, กรุงเทพฯ: พิศิษฐ์การพิมพ์, 2540.

[10] ชุดาภา ผิวเผือก, "*ปัจจัยที่ใช้ในการวิเคราะห์สินเชื่อของธนาคารพาณิชย์ในประเทศไทย,*" วิทยานิพนธ์ บธ.ม. (การบัญชี), มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพฯ, 2561.

[11] สุมิตรา ตุมาวรคุณ, "*ปัจจัยที่ส่งผลต่อการเกิดหนี้ที่ไม่ก่อให้เกิดรายได้ (NPLs) สินเชื่อธนาคารประชาชน กรณีศึกษาธนาคารออมสินสาขานครราชสีมา,*" วิทยานิพนธ์ บธ.บ. (การจัดการเงินและธนาคาร), มหาวิทยาลัยรามคำแหง, กรุงเทพฯ, 2562.

[12] C. Pittayawiwit, "*Credit Management Comprehensive Financial Institutions*," M.B.A. (Finance), Chulalongkorn University, Bangkok, 2007.

[13] พัชรี หล้าแหล่ง, *การวิเคราะห์โครงสร้างรายได้ ภาระหนี้สิน สภาพเศรษฐกิจและสังคมของเกษตรกรชาวสวนปาล์มน้ำมันในพื้นที่ภาคใต้*, เชียงใหม่: สำนักวิจัยและส่งเสริมวิชาการการเกษตร มหาวิทยาลัยแม่โจ้, 2555.

[14] บรรพต ตัณฑศรี, "*ปัจจัยที่ทำให้เกิดการค้างชำระของลูกหนี้ที่ได้รับการปรับปรุงโครงสร้างหนี้ ลูกค้า ธ.ก.ส. จังหวัดเชียงราย*," วิทยานิพนธ์ บธ.ม. (การจัดการทั่วไป), มหาวิทยาลัยราชภัฏเชียงราย, เชียงราย, 2549.

[15] กฤษฎา อึ้งทองหล่อ และ ประเวศ เพ็ญวุฒิกุล, "ความสามารถในการชำระหนี้ การวิเคราะห์สภาพคล่องทางการเงินและความรับผิดชอบต่อสังคมส่งผลต่อความสามารถในการทำกำไรของบริษัทที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยในกลุ่มอสังหาริมทรัพย์และการก่อสร้าง," *วารสารวิทยาการจัดการปริทัศน์*, ปีที่ 25, ฉบับที่ 2, หน้า 80-91, พฤษภาคม-สิงหาคม 2566.

[16] บุญเสริม กิจศิริกุล, "*อัลกอริทึมการทำเหมืองข้อมูล*," โครงการวิจัยร่วมภาครัฐและเอกชน ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2548.

[17] สงกรานต์ สมบุญและคณะ, "Credit Scoring Model เครื่องมือที่ต้องมีเพื่อช่วยบริหารความเสี่ยงสินเชื่อ," *FOCUSED AND QUICK (FAQ),* ปีที่ 10, ฉบับที่ 187, หน้า 1-9, พฤษภาคม 2021.

[18] C. Lawrence et al., "A machine learning approach to research curation for investment process," *Journal of Investment Management (JOIM)*, vol. 15, no. 1, no page, January 2017.

[19] R. J. Gray and K. S. Grove, *Burns & Grove's The Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence 9th Edition,* Texas: n.p., 2021.

[20] J. P. Stevens and K. A. Pituch, *Binary Logistic Regression,* New York: n.p., 2015.

[21] จิราภรณ์ เจริญยิ่ง, "*การพยากรณผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูลโดยใช้ Rapid Miner,*" สารนิพนธ์ วท.ม. (เทคโนโลยีสารสนเทศ), มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ, 2563.

[22] A. Verikas and M. Bacauskiene, "Feature selection with neural networks," *Pattern Recognition Letters*, vol. 2002, no. 23, pp. 1323–1335, May 2001.

[23] ภรัณยา ปาลวิสุทธ, "การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต," *วารสารเทคโนโลยีสารสนเทศ*, ปีที่ 12, ฉบับที่ 1, หน้า 54-63, มกราคม-มิถุนายน 2559.

[24] K. Dunlap and K. R. Popper, *"K-Fold Cross Validation การวัดประสิทธิภาพของโมเดล,"* [Online]. Available: https://shorturl.asia/dHli8. [Accessed: 5 มิถุนายน 2566].

[25] ระพีวรรณ อุดมรัตน์, "*การศึกษาลักษณะของลูกค้าธุรกิจขนาดกลางที่มีโอกาสเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ (NPL): กรณีศึกษาของสถาบันการเงินแห่งหนึ่ง*," การค้นคว้าอิสระ บธ.ม. (บริหารธุรกิจ), มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, 2561.

[26] กันทิมา มีนิล, "*การศึกษาเพื่อพยากรณ์ลูกค้า NPL สินเชื่อเพื่อที่อยู่อาศัยของธนาคารอาคารสงเคราะห์และจัดกลุ่มการเข้าสู่สภาวะความเสี่ยง*," การศึกษาค้นคว้าอิสระ บธ.ม. (วิศวกรรมคอมพิวเตอร์และเทคโนโลยีการเงิน), มหาวิทยาลัยหอการค้าไทย, กรุงเทพฯ, 2564.

[27] ปารดา ศัสตุระ, "การประยุกต์เหมืองข้อมูลในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ กรณีศึกษา ธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่," *การประชุมนำเสนอผลงานวิจัยบัณฑิตศึกษา ครั้งที่ 17 ประจำปีการศึกษา 2565 VIRTUAL CONFERENCE*, กรุงเทพฯ, 10 สิงหาคม 2565, ไม่ปรากฏเลขหน้า.

[28] เฉลิมชาติ ชัยวิลา และ ศิวาพร ฟองทอง, "ปัจจัยที่ส่งผลต่อการชำระหนี้ของลูกค้าธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร สาขาบ้านผือ จังหวัดอุดรธานี," *Journal of Modern Learning Development*, ปีที่ 5, ฉบับที่ 5, หน้า 197-211, กันยายน-ตุลาคม 2563.

[29] สุพัตรา มนัสชล และนลินี ทินนาม, "ศักยภาพในการชำระหนี้ของผู้กู้โครงการสินเชื่อฟื้นฟู SMEs จากอุทกภัยและภัยพิบัติ ปี 2560 ของธนาคารพัฒนาวิสาหกิจขนาดกลางและขนาดย่อมแห่งประเทศไทย สาขานครศรีธรรมราช," *การประชุมวิชาการระดับชาติ วลัยลักษณ์วิจัย ครั้งที่ 11*, จังหวัดนครศรีธรรมราช, 27-28 มีนาคม 2562, หน้า 1-8.

[30] ศิริลักษณ์ บุญชัยสุข และอุษณา แจ้งคล้อย, "ปัจจัยที่มีผลต่อพฤติกรรมการจ่ายชำระหนี้ของลูกหนี้กองทุนส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ ในเขตจังหวัดนครราชสีมา," *วารสารบริหารธุรกิจและการบัญชี มหาวิทยาลัยขอนแก่น*, ปีที่ 2, ฉบับที่ 2, ไม่ปรากฏเลขหน้า, พฤษภาคม- สิงหาคม 2561.

[31] ชลธิชา สุวรรณพิทักษ์, "ปัจจัยที่มีผลต่อการชำระหนี้ตามสัญญาปรับปรุงโครงสร้างหนี้ของลูกหนี้สินเชื่อบุคคล กรณีศึกษา: ธนาคารกสิกรไทย จำกัด มหาชน หน่วยบริหารคุณภาพสินทรัพย์ พัทยากลาง," *วารสารบัณฑิตศึกษา มหาวิทยาลัยราชภัฏสกลนคร*, ปีที่ 14, ฉบับที่ 64, หน้า 149-160, มกราคม-มีนาคม 2560.

[32] เครือวัลย์ เนตรพนา, "*การวิเคราะห์ความเสี่ยงในการผิดนัดชำระของลูกหนี้บัตรเครดิต โดยการใช้อัลกอริทึมการเรียนรู้ของเครื่อง*," สารนิพนธ์ วท.ม. (วิทยาการข้อมูล), มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ, 2566.

[33] สุดากาญจน์ ติดตะ, "*ระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทคนิควิธีการเรียนรู้ของเครื่อง*," สารนิพนธ์ วศ.ม. (วิศวกรรมข้อมูลขนาดใหญ่), มหาวิทยาลัยธุรกิจบัณฑิตย์, กรุงเทพฯ, 2564.

[34] วริทธิ์พล แสงทองรัตนโชติ, "*การเปรียบเทียบเทคนิคการเรียนรู้ของเครื่องเพื่อสร้างตัวแบบการจำแนกด้วยการปรับปรุงชุดข้อมูลอสมดุล*," วิทยานิพนธ์ วท.ม. (สถิติ), มหาวิทยาลัยนเรศวร, พิษณุโลก, 2565.

[35] พัชรียา ทองพูล และคณะ, "การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล," *Thai Journal of Science and Technology,* ปีที่ 8, ฉบับที่ 6, หน้า 565-584, พฤศจิกายน-ธันวาคม 2562.

ประวัติย่อผู้วิจัย

ชื่อ – สกุล นางสาวธันย์ชนก ขำประดิษฐ์

วัน เดือน ปีเกิด 10 มกราคม 2537

ที่อยู่ปัจจุบัน 75/121 ซ.ลาซาล 7 ถนนสุขุมวิท 105 แขวงบางนาใต้
เขตบางนา กทม. 10260
โทรศัพท์มือถือ : 09-4362-6691
E-mail : Thanchanok.tk@hotmail.com

ประวัติการศึกษา

พ.ศ. 2567 วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น

พ.ศ. 2559 บริหารธุรกิจบัณฑิต สาขาการเงิน
มหาวิทยาลัยสงขลานครินทร์

ประวัติการทำงาน

พ.ศ. 2565 – ปัจจุบัน พนักงานวิเคราะห์งานสินเชื่อ
ธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร

พ.ศ. 2560 – 2564 Ticketing and Reservation Helpdesk
Thai Smile Airways