การพยากรณ์การผิดนัดชำระหนี้เงินกู้ในแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (P2P) ด้วยเทคนิคการเรียนรู้ของเครื่อง

วิทยานิพนธ์

ของ

วฤทธิ วารีศรี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

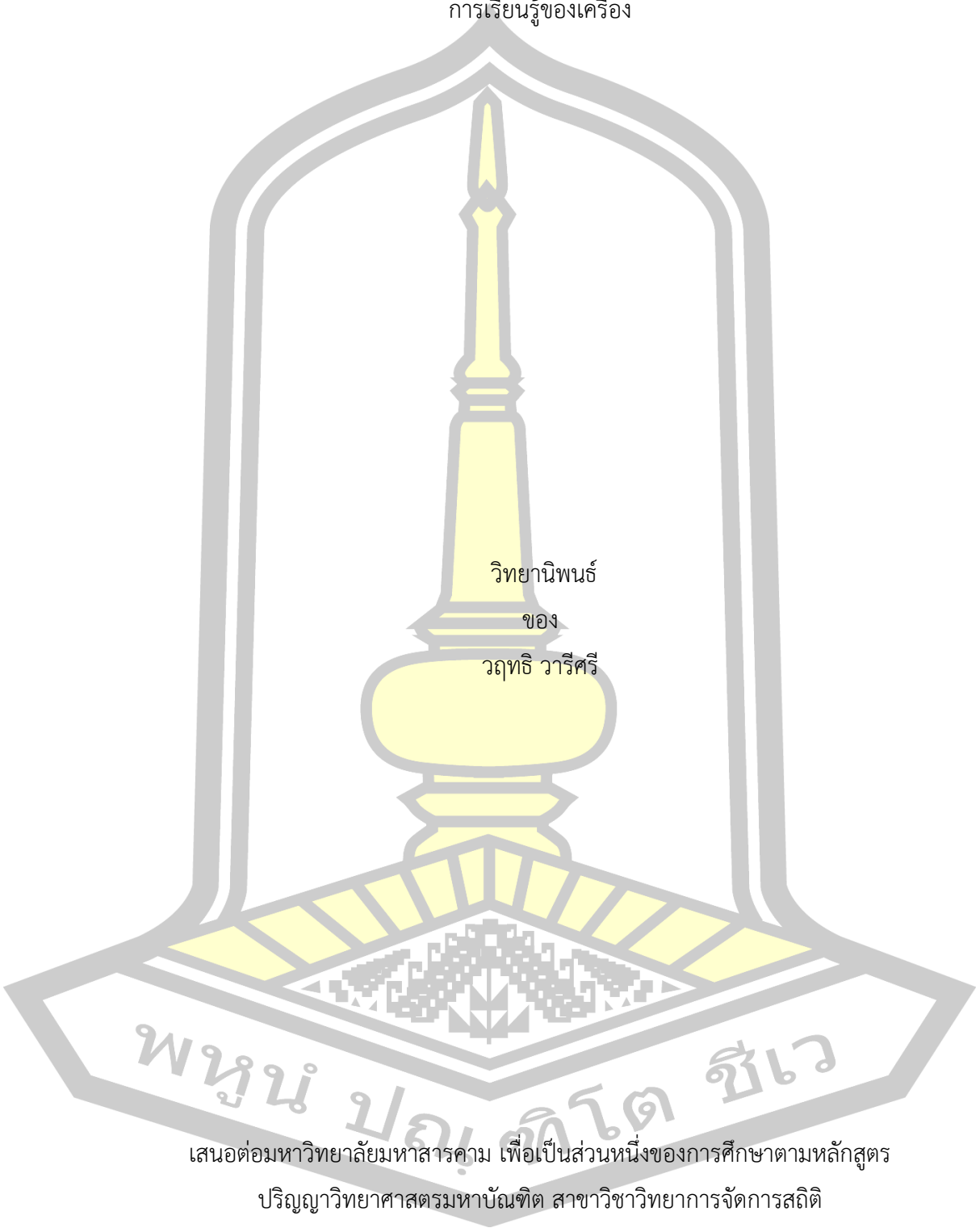
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

ธันวาคม 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

พหุน ปณฺ ฑิโต ชีเว

การพยากรณ์การผิดนัดชำระหนี้เงินกู้ในแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (P2P) ด้วยเทคนิคการเรียนรู้ของเครื่อง



วิทยานิพนธ์

ของ

วฤทธิ วารีศรี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ

ธันวาคม 2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Predicting Loan Default on Peer-to-Peer (P2P) Lending Platforms using Machine Learning Techniques

Warit Wareesri

A Thesis Submitted in Partial Fulfillment of Requirements

for Master of Science (Statistical Management Science)

December 2024

Copyright of Mahasarakham University

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนายวฤทธิ วารีศรี แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผศ. ดร. พรรณรัตน์ ก้วยเจริญพานิชก์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อ. ดร. สุภาวดี วิชิตชาญ)

.....กรรมการ
(ผศ. ดร. มนชยา เจียงประดิษฐ์)

.....กรรมการ
(รศ. ดร. นิภาพร ชุติมันต์)

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการจัดการสถิติ ของมหาวิทยาลัยมหาสารคาม

.....

(ศ. ดร. ไพโรจน์ ประมวล) (ผศ. ดร. กริสน์ ชัยมูล)
คณบดีคณะวิทยาศาสตร์ คณบดีบัณฑิตวิทยาลัย

พหุน ปณฺ ฑิโต ลภเต

ชื่อเรื่อง การพยากรณ์การผิดนัดชำระหนี้เงินกู้ในแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (P2P) ด้วยเทคนิคการเรียนรู้ของเครื่อง

ผู้วิจัย วฤทธิ วารีศรี

อาจารย์ที่ปรึกษา อาจารย์ ดร. สุภาวดี วิชิตชาญ

ปริญญา วิทยาศาสตรมหาบัณฑิต **สาขาวิชา** วิทยาการจัดการสถิติ

มหาวิทยาลัย มหาวิทยาลัยมหาสารคาม **ปีที่พิมพ์** 2567

บทคัดย่อ

การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (P2P) เป็นปัญหาที่ส่งผลกระทบต่อทั้งผู้ให้กู้และผู้กู้ อีกทั้งยังส่งผลต่อเสถียรภาพของตลาดสินเชื่อออนไลน์ ผู้กู้ที่เข้าถึงสินเชื่อได้ง่ายเกินไปอาจเผชิญกับภาระหนี้เกินกว่าที่จะชำระได้ การพยากรณ์การผิดนัดชำระหนี้จึงมีความสำคัญการบริหารความเสี่ยงทางการเงิน งานวิจัยนี้ได้นำเสนอการพัฒนาแบบจำลองการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ร่วมกับการวิเคราะห์ปัจจัยพื้นฐาน และปัจจัยเศรษฐศาสตร์มหภาคที่มีผลต่อการผิดนัดชำระหนี้ ข้อมูลที่ใช้ในการวิจัยได้มาจากแพลตฟอร์ม Bondora ประกอบด้วยข้อมูลสินเชื่อของกลุ่มผู้กู้ที่มีและไม่มีการผิดนัดชำระหนี้ในช่วงเดือนกุมภาพันธ์ พ.ศ. 2552 ถึงมกราคม พ.ศ. 2567 มีจำนวนข้อมูลทั้งหมด 358,092 รายการ 112 ตัวแปร และข้อมูลด้านเศรษฐศาสตร์มหภาคจาก World Economic Outlook (WEO) ผลการศึกษาพบว่าปัจจัยพื้นฐานที่สำคัญที่นำไปสู่การผิดนัดชำระหนี้ ได้แก่ อัตราดอกเบี้ยสูงสุดที่ยอมรับได้, รายได้รวม และจำนวนเงินโดยประมาณที่ต้องชำระรายเดือน ปัจจัยด้านเศรษฐศาสตร์มหภาคที่สำคัญ ได้แก่ ผลิตภัณฑ์มวลรวมในประเทศต่อหัว, อัตราเงินเฟ้อ, อัตราการว่างงาน และจำนวนประชากร ผลการวิจัยแสดงให้เห็นว่าอัลกอริทึม Random Forest และ XGBoost มีประสิทธิภาพสูงสุดในการพยากรณ์ โดยเฉพาะอย่างยิ่งเมื่อรวมตัวแปรเศรษฐศาสตร์มหภาคเข้ากับตัวแปรพื้นฐาน การศึกษานี้ยังชี้ให้เห็นถึงความสำคัญของการจัดการข้อมูลไม่สมดุล และการคัดเลือกคุณลักษณะในการเพิ่มประสิทธิภาพของแบบจำลอง

คำสำคัญ : การผิดนัดชำระหนี้, สินเชื่อออนไลน์ระหว่างบุคคล, การเรียนรู้ของเครื่อง, เศรษฐศาสตร์มหภาค

TITLE Predicting Loan Default on Peer-to-Peer (P2P) Lending Platforms using Machine Learning Techniques

AUTHOR Warit Wareesri

ADVISORS Supawadee Wichitchan , Ph.D.

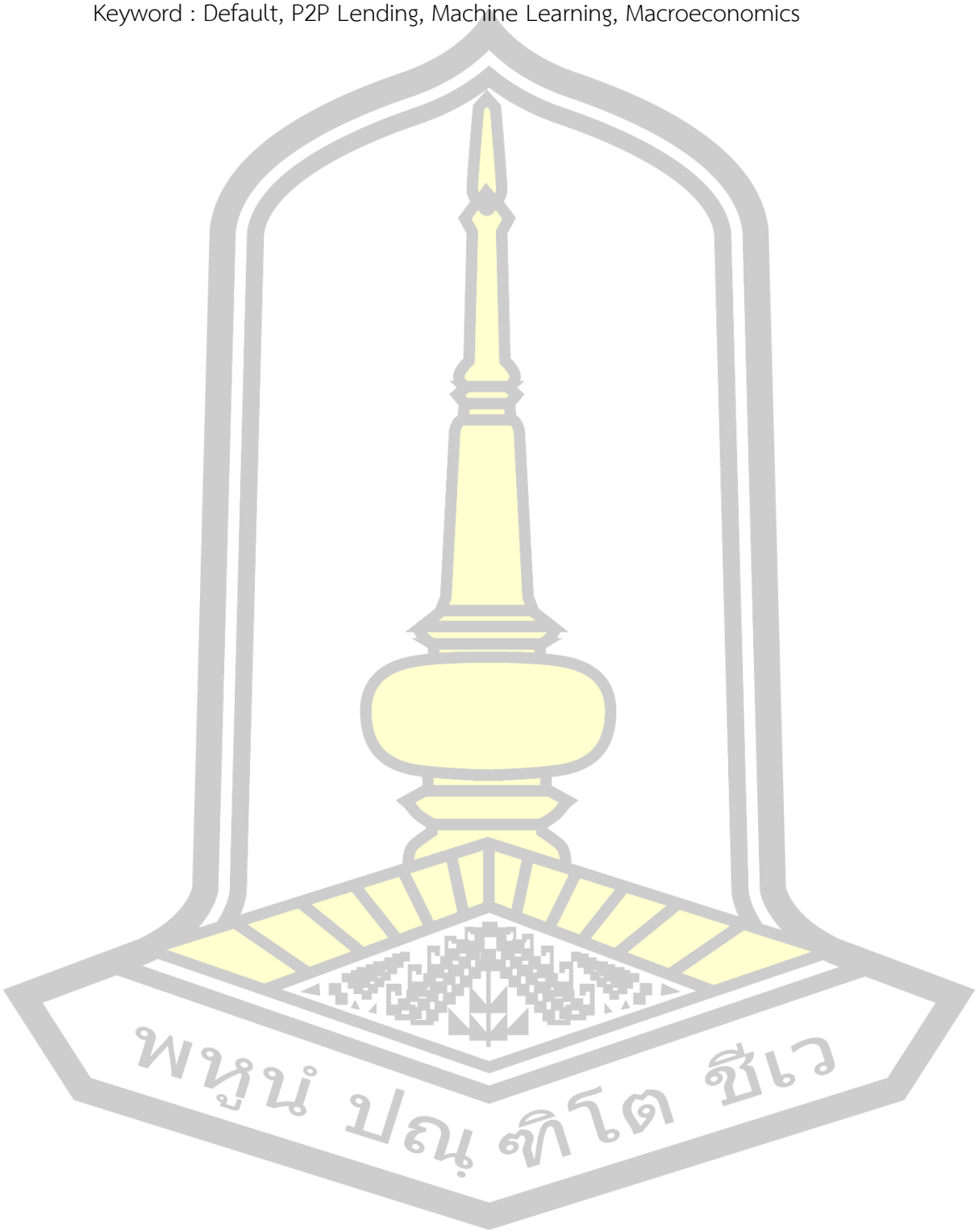
DEGREE Master of Science **MAJOR** Statistical Management Science

UNIVERSITY Mahasarakham University **YEAR** 2024

ABSTRACT

Loan default on peer-to-peer (P2P) lending platforms significantly impacts that affects both lenders, borrowers and market stability. Borrowers who get access to credit easily may face excessive debt beyond their repayment capacity. Therefore, accurately predicting loan default is essential for effective financial risk management. This study develops a loan default prediction model on P2P lending platforms using machine learning techniques along with fundamental factors analysis and macroeconomic factors influencing default. The data used in this research were obtained from the Bondora platform, containing loan information from borrowers who either defaulted or did not default between February 2009 and January 2024, comprises 358,092 records and 112 variables. Macroeconomic data were sourced from the World Economic Outlook (WEO). The study discovered that the key fundamental factors contributing to loan default include the interest rate, total income, and monthly payments. Macroeconomic factors include GDP per capita, inflation rate, unemployment rate, and population. Results indicate that the Random Forest and XGBoost algorithms provide the highest prediction accuracy, especially when integrating macroeconomic variables This study also highlights the importance of addressing imbalanced data and feature selection in enhancing model performance.

Keyword : Default, P2P Lending, Machine Learning, Macroeconomics



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ด้วยดี โดยได้รับความกรุณาและความช่วยเหลืออย่างดียิ่งจาก อาจารย์ ดร.สุภาวดี วิชิตชาญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. พรรณรัตน์ ก้วยเจริญพานิชก์ ประธานสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. นิภาพร ชุติมันต์ และ ผู้ช่วยศาสตราจารย์ ดร. มนชยา เจียงประดิษฐ์ กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำปรึกษา การวิเคราะห์ข้อมูล ตรวจสอบ และแก้ไขปรับปรุงข้อบกพร่องต่าง ๆ ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ด้วยดี ผู้วิจัยขอขอบพระคุณไว้ ณ โอกาสนี้

ขอขอบพระคุณคณาจารย์สาขาวิชาวิทยาการจัดการสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคามทุกท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ ให้แนวทางซึ่งเป็นประโยชน์ในการศึกษาและค้นคว้าวิจัย ตลอดระยะเวลาที่ผู้วิจัยได้ศึกษาอยู่ที่สาขาวิชาวิทยาการจัดการสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม

ขอขอบพระคุณบุพการีที่ให้การอบรมเลี้ยงดูและให้โอกาสในการศึกษา รวมถึงมิตรสหาย และผู้มีส่วนเกี่ยวข้องทุกท่านที่ให้ความช่วยเหลือ คำแนะนำ เตือนสติ และเป็นกำลังใจให้แก่ผู้วิจัยเสมอมาจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

วฤทธิ วารีศรี

พหูน ปณฺฑิโต ชีเว

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ซ

สารบัญรูปภาพ ..... ฎ

สารบัญตาราง ..... ฎ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 ความสำคัญของการวิจัย 4

1.4 ขอบเขตของการวิจัย 4

1.5 ประโยชน์ที่คาดว่าจะได้รับ 5

1.6 นิยามศัพท์เฉพาะ 5

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

2.1 สินเชื่อออนไลน์ระหว่างบุคคล (Peer-to-Peer Lending) 6

2.1.1 ประวัติความเป็นมา 6

2.1.2 สินเชื่อออนไลน์ระหว่างบุคคล 6

2.1.3 การทำงานของสินเชื่อออนไลน์ระหว่างบุคคล 7

2.1.4 ความเสี่ยงของสินเชื่อออนไลน์ระหว่างบุคคล 7

2.2 เศรษฐศาสตร์มหภาค (Macroeconomic) 9

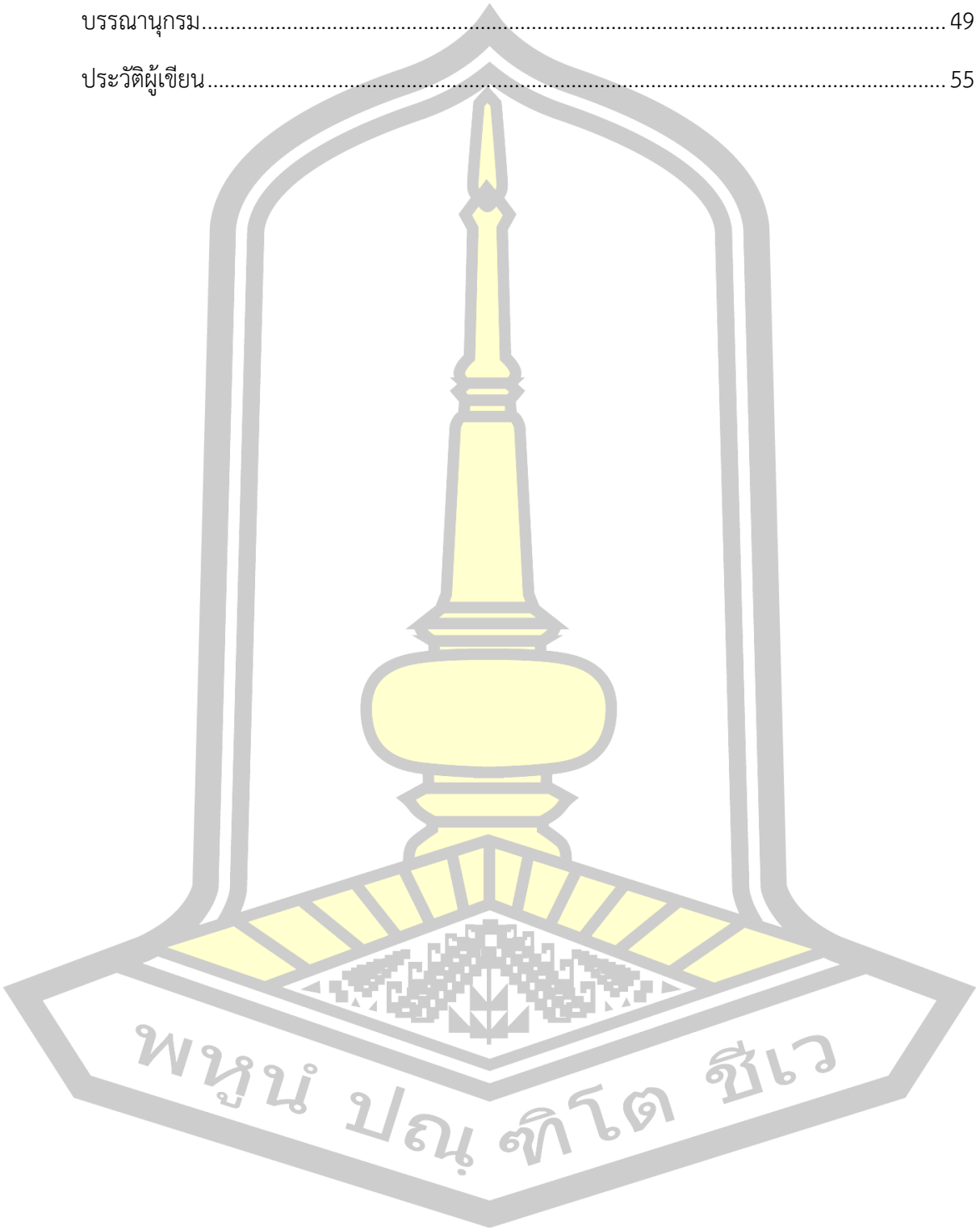
2.3 ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง (Workflow of Machine Learning) 9

2.4 อัลกอริทึม (Algorithm) 12
2.4.1 อัลกอริทึม LightGBM 12
2.4.2 อัลกอริทึม Random Forest 13
2.4.3 อัลกอริทึม XGBoost 14
2.5 การประเมินประสิทธิภาพของแบบจำลอง (Models Evaluation) 15
2.6 การวิเคราะห์ความสัมพันธ์ (Correlation Analysis) 16
2.6.1 วิธีการหาค่าสหสัมพันธ์แบบเพียร์สัน (Pearson Correlation) 17
2.6.2 วิธีการหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point Biserial Correlation) 17
2.7 งานวิจัยที่เกี่ยวข้อง 17
บทที่ 3 วิธีดำเนินการวิจัย 21
3.1 ชุดข้อมูล (Data set) 21
3.1.1 ข้อมูลสินเชื่อ Bondora 21
3.1.2 ข้อมูลเศรษฐศาสตร์มหภาค (Macroeconomic Dataset) 21
3.2 กรอบการดำเนินงาน 22
บทที่ 4 ผลการวิจัย 31
4.1 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการหาความสำคัญของคุณลักษณะ (Feature Importance) 31
4.2 ปัจจัยทางเศรษฐศาสตร์มหภาคที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการการวิเคราะห์ความสัมพันธ์ (Correlation) ของตัวแปรด้านเศรษฐศาสตร์มหภาคและตัวแปรตาม 35
4.3 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยชุดของตัวแปรที่แตกต่างกัน 4 รูปแบบ 36
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 44
5.1 สรุปผล 44
5.2 อภิปรายผล 46

5.3 ข้อเสนอแนะ 47

บรรณานุกรม 49

ประวัติผู้เขียน 55



สารบัญรูปภาพ

หน้า

รูปที่ 1 ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง ที่มา: Gangoda and Wanninayake (2023) 10
รูปที่ 2 วิธีการ Leaf-Wise และวิธีการ Level-Wise 13
รูปที่ 3 อัลกอริทึม Random Forest 14
รูปที่ 4 อัลกอริทึม XGBoost 15
รูปที่ 5 ตาราง Confusion Matrix 15
รูปที่ 6 แผนผังกรอบดำเนินงาน 23
รูปที่ 7 จำนวนลูกค้าที่มีการผิดนัดชำระหนี้และไม่ผิดนัดชำระหนี้ 24
รูปที่ 8 แสดงการวิเคราะห์ความสัมพันธ์ของตัวแปรเศรษฐศาสตร์มหภาคและตัวแปรตาม 25
รูปที่ 9 ตัวอย่างการเข้ารหัสข้อมูลด้วยวิธี Label Binarizer 29
รูปที่ 10 วิธีสุ่มลด (Undersampling) ที่มา: Sampaio (2023) 30
รูปที่ 11 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง LightGBM 32
รูปที่ 12 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง Random Forest 33
รูปที่ 13 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง XGBoost 34
รูปที่ 14 ตัวแปรด้านเศรษฐศาสตร์มหภาคที่มีค่า |Correlation| ≥ 0.2 กับตัวแปรตาม 35
รูปที่ 15 Normalized Confusion Matrix แบบจำลองที่ผึกอบรมโดยอัลกอริทึม LightGBM 39
รูปที่ 16 Normalized Confusion Matrix แบบจำลองที่ผึกอบรมโดยอัลกอริทึม Random Forest 41
รูปที่ 17 Normalized Confusion Matrix แบบจำลองที่ผึกอบรมโดยอัลกอริทึม XGBoost 42

สารบัญตาราง

หน้า

ตารางที่ 1 ตัวแปรด้านเศรษฐศาสตร์มหภาค 21

ตารางที่ 2 พจนานุกรมของข้อมูลสินเชื่อ 25

ตารางที่ 3 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยตัวแปรทั้งหมด 36

ตารางที่ 4 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยตัวแปรทั้งหมด และจัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling)..... 37

ตารางที่ 5 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) และคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) 37

ตารางที่ 6 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) ทำการคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) และรวมตัวแปรด้านเศรษฐศาสตร์มหภาค (Macroeconomic Dataset)..... 38

ตารางที่ 7 ผลการวัดประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องที่ฝึกอบรมด้วยชุดข้อมูลที่มีตัวแปรที่แตกต่างกัน 4 รูปแบบ 45

พหูน ปณฺฑิโต ชีเว

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

สินเชื่อส่วนบุคคล (Personal Loan) คือ จำนวนเงินที่ลูกค้าของผู้ให้บริการสินเชื่อสามารถกู้ยืมเพื่อนำไปใช้จ่ายส่วนตัวในวัตถุประสงค์ต่าง ๆ ตัวอย่างเช่น ลูกค้าอาจใช้สินเชื่อส่วนบุคคลเพื่อรวมหนี้ ใช้เป็นค่ารักษาพยาบาล การปรับปรุงหรือซ่อมแซมบ้าน การซื้อรถ ซื้อบ้าน หรือวางแผนการจัดงานแต่งงาน เป็นต้น ผู้ที่ต้องการขอสินเชื่อส่วนบุคคลสามารถขอรับบริการได้โดยผ่านทางธนาคาร สหกรณ์เครดิตยูเนี่ยน หรือผู้ให้กู้ผ่านช่องทางออนไลน์ เงินที่ลูกค้ายืมจะต้องได้รับการชำระคืนเมื่อเวลาผ่านไป โดยปกติแล้วจะมีดอกเบี้ย ผู้ให้กู้บางรายอาจเรียกเก็บค่าธรรมเนียมสำหรับสินเชื่อส่วนบุคคล (Lake, 2021) ในโลกยุคดิจิทัลที่เทคโนโลยีเข้ามามีบทบาทสำคัญเพื่อช่วยอำนวยความสะดวกในชีวิตประจำวันทั้งในแง่ความสะดวกสบายและความรวดเร็ว เห็นได้จากพฤติกรรมของผู้บริโภคที่เปลี่ยนไปใช้บริการผ่านช่องทางออนไลน์มากขึ้น เช่น การซื้อสินค้า การจองตั๋วภาพยนตร์ การจองที่พัก รวมไปถึงการทำธุรกรรมทางการเงิน ซึ่งปัจจุบันการกู้ยืมเงินก็เป็นอีกหนึ่งธุรกรรมทางการเงินที่สามารถทำผ่านช่องทางออนไลน์ได้ โดยผู้ที่ต้องการเงินทุนสามารถมาขอกู้เงินผ่านเว็บไซต์หรือแอปพลิเคชันที่ทำหน้าที่เป็นตัวกลางหรือช่องทาง หรือหากเป็นผู้ที่มีเงินเหลืออยากหาทางเลือกใหม่ในการลงทุนก็สามารถปล่อยกู้ได้เช่นกัน โดยดำเนินการผ่านผู้ให้บริการสินเชื่อออนไลน์ระหว่างบุคคล (ธนาคารแห่งประเทศไทย, 2563)

สินเชื่อออนไลน์ระหว่างบุคคล (Peer-to-Peer Lending หรือ P2P Lending) คือ ธุรกรรมการกู้ยืมเงินระหว่างบุคคลทั่วไปผ่านแพลตฟอร์มออนไลน์ โดยไม่จำเป็นต้องผ่านตัวกลางอย่างธนาคารหรือสถาบันการเงิน ซึ่งบุคคลทั่วไปสามารถกู้ยืมเงินกันเองได้โดยตรง โดยมีแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลทำหน้าที่เสมือนเป็นตลาดสินเชื่อออนไลน์ที่เชื่อมต่อระหว่าง ผู้กู้และผู้ให้กู้ให้มาเจอกัน ดังนั้นแนวคิดหลักของสินเชื่อออนไลน์ระหว่างบุคคล คือ การทำให้บุคคลทั่วไปสามารถเข้าถึงแหล่งเงินกู้ได้สะดวกขึ้น ในขณะเดียวกันก็มีทางเลือกในการลงทุนที่มากขึ้นกว่าเดิมเช่นกัน (หลักทรัพย์บัวหลวง, 2564) อย่างไรก็ตาม การทำธุรกรรมสินเชื่อผ่านแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลนั้นมีความเสี่ยง โดยที่ผู้กู้นั้นสามารถเข้าถึงแหล่งสินเชื่อได้ง่ายเกินไปจึงทำให้เกิดความเสี่ยงที่ผู้กู้นั้นจะก่อหนี้จนเกินความสามารถในการชำระหนี้ของตน หรือผู้ให้กู้อาจไม่ได้รับชำระหนี้หรือดอกเบี้ยตามสัญญา เนื่องจากไม่ใช่การฝากเงิน จึงไม่สามารถได้รับการคุ้มครองเช่นเดียวกับเงินฝาก และไม่อาจสามารถยกเลิกการให้สินเชื่อหรือเรียกให้ชำระหนี้ก่อนครบกำหนด ส่งผลให้ผู้ให้กู้ขาดสภาพคล่องทางการเงินเมื่อผู้กู้ผิดนัดชำระหนี้หรือไม่สามารถชำระหนี้ได้ครบตาม

กำหนดสัญญา และเนื่องจากสัญญาสินเชื่อนั้นสามารถเกิดขึ้นได้โดยที่ผู้กู้และผู้ให้กู้ไม่จำเป็นต้องแสดงตัวต่อกัน จึงมีโอกาสเกิดการหลอกลวงทั้งจากแพลตฟอร์มและผู้กู้ (รพีพัฒน์ อิงคสิทธิ์, 2563) การชำระคืน (Repayment) เป็นการจ่ายคืนเงินที่ยืมมาจากผู้ให้กู้ก่อนหน้านี้ โดยทั่วไป การคืนเงินจะเกิดขึ้นผ่านการชำระเงินเป็นงวด ซึ่งรวมถึงเงินต้นและดอกเบี้ย เงินต้น หมายถึง เงินก้อนเดิมที่ยืมมาในเงินกู้ ดอกเบี้ย หมายถึง ค่าธรรมเนียมหรือผลตอบแทนที่ผู้ให้กู้เรียกเก็บจากผู้กู้ ผู้กู้จะต้องจ่ายดอกเบี้ยสำหรับความสามารถในการใช้เงินที่ปล่อยผ่านเงินกู้ โดยปกติแล้วเงินกู้สามารถชำระเต็มจำนวนได้ตลอดเวลา แม้ว่าสัญญาบางฉบับอาจมีค่าธรรมเนียมการชำระคืนก่อนกำหนด การไม่สามารถชำระหนี้ได้ทันอาจนำไปสู่ปัญหาด้านเครดิต เช่น การถูกบังคับล้มละลาย การมีค่าใช้จ่ายที่เพิ่มขึ้นจากการชำระล่าช้า และการเปลี่ยนแปลงอันดับเครดิตในเชิงลบ (Twin, 2021) และเพื่อป้องกันปัญหาการผิดนัดชำระหนี้ในลำดับแรกจึงต้องมีการประเมินความเหมาะสมของผู้กู้ ในประเทศไทยมีการกำหนดหลักเกณฑ์ วิธีการ และเงื่อนไขในการประกอบธุรกิจระบบหรือเครือข่ายอิเล็กทรอนิกส์สำหรับธุรกรรมสินเชื่อระหว่างบุคคลกับบุคคล ไว้โดยธนาคารแห่งประเทศไทย ซึ่งได้กำหนดคุณสมบัติผู้กู้ เช่น เป็นบุคคลธรรมดาที่มีความสามารถในการชำระหนี้ ไม่เป็นผู้ประกอบธุรกิจระบบหรือเครือข่ายอิเล็กทรอนิกส์ เป็นต้น (ธนาคารแห่งประเทศไทย, 2562) แต่ทว่าการประเมินความเหมาะสมของผู้กู้ในเบื้องต้นนั้นก็ยังไม่เพียงพอที่จะป้องกันไม่ให้เกิดปัญหาการผิดนัดชำระหนี้ได้ เนื่องจากปัจจัยที่ส่งผลต่อการผิดนัดชำระหนี้มีได้หลายปัจจัย เช่น วัตถุประสงค์ของเงินกู้ รายได้ต่อปีของผู้กู้ สถานการณ์ที่อยู่อาศัยในปัจจุบัน ประวัติเครดิตและการเป็นหนี้ (Serrano-Cinca et al., 2015) นอกจากนี้ปัจจัยทางเศรษฐศาสตร์มหภาค (Bloomenthal, 2022) มีอิทธิพลซึ่งส่งผลกระทบในวงกว้างต่อเศรษฐกิจระดับภูมิภาคหรือระดับประเทศ และอาจส่งผลกระทบไปยังลูกค้าที่ต้องชำระหนี้ของสินเชื่อ (Nigmonov et al., 2022) ได้ทำการศึกษาปัจจัยทางเศรษฐศาสตร์ภาคที่สำคัญที่มีอิทธิพลต่อการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลโดยใช้การวิเคราะห์การถดถอย (Regression Analysis) ผลที่ได้พบว่าอัตราดอกเบี้ยและอัตราเงินเฟ้อที่สูงขึ้นจะเพิ่มอัตราการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

จากปัญหาข้างต้นทำให้ผู้ให้กู้เล็งเห็นว่าหากสามารถพยากรณ์ล่วงหน้าได้ว่าลูกค้ารายใดมีแนวโน้มที่จะผิดนัดชำระหนี้ได้ ก็จะสามารถใช้ประกอบการตัดสินใจ (Decision Making) สำหรับการอนุมัติสินเชื่อให้กับลูกค้า หรือใช้เป็นระบบเตือนภัยล่วงหน้าได้ (Early-Warning System) โดยการตรวจสอบพฤติกรรมของลูกค้า ซึ่งลูกค้าที่มีแนวโน้มที่จะผิดนัดชำระหนี้จะถูกตั้งสถานะไว้ เพื่อให้ผู้ให้กู้สามารถเข้าไปแก้ปัญหาได้ก่อนที่จะเกิดการผิดนัดชำระหนี้ (Profinit, 2022) ดังนั้นจึงมีนักวิจัยหลายคนได้นำเสนอวิธีการต่าง ๆ ในการพยากรณ์การผิดนัดชำระหนี้ ซึ่งวิธีการเรียนรู้ของเครื่อง (Machine Learning) เป็นอีกหนึ่งวิธีการที่นิยมนำมาใช้สร้างแบบจำลองสำหรับพยากรณ์การผิดนัดชำระหนี้ (Wu, 2022) ได้นำวิธีการเรียนรู้ของเครื่องมาใช้พยากรณ์การผิดนัดชำระหนี้ โดยใช้

อัลกอริทึม Random Forest และ XGBoost ซึ่งแบบจำลองทั้งสองมีความแม่นยำในการพยากรณ์สูงถึง 90% (J. Xu et al., 2021) ได้ใช้วิธีการเรียนรู้ของเครื่อง 4 วิธี ได้แก่ Random Forest, Extreme Gradient Boosting, Gradient Boosting Model และ Neural Network เพื่อทำนายปัจจัยที่ส่งผลต่อผิดนัดชำระหนี้ ผลการศึกษาพบว่าปัจจัยด้านการมีโทรศัพท์มือถือ อาชีพ ที่พักอาศัย และระดับการศึกษาเป็นปัจจัยที่สำคัญที่ส่งผลต่อการผิดนัดชำระหนี้ นอกจากนี้ยังพบว่าอัลกอริทึม Random Forest มีประสิทธิภาพดีที่สุดจากทั้ง 4 วิธี (Madaan et al., 2021) ได้นำเสนอแบบจำลองการเรียนรู้ของเครื่องสองแบบเพื่อพยากรณ์ว่าลูกค้าควรจะได้อนุมัติการขอสินเชื่อหรือไม่ เพื่อเป็นตัวช่วยของเจ้าหน้าที่ธนาคารในการลดขั้นตอนการคัดกรองลูกค้า อัลกอริทึมที่ใช้คือ Random Forest และ Decision Tree จากผลการศึกษาพบว่า Random Forest มีความแม่นยำในการพยากรณ์ดีที่สุด นอกจากนี้ยังมีการศึกษาการใช้อัลกอริทึม LightGBM (Dong, 2022) ในการพยากรณ์การผิดนัดชำระหนี้ โดยอาศัยข้อมูลจาก Tianchi Platform และใช้วิศวกรรมคุณลักษณะร่วมกับการฝึกแบบจำลอง พบว่า LightGBM มีค่า AUC อยู่ที่ประมาณ 0.73 ซึ่งแสดงให้เห็นว่ามีความแม่นยำในการพยากรณ์อยู่ในระดับดี โดยผลการศึกษายังพบว่าปัจจัยสำคัญที่ส่งผลต่อการผิดนัดชำระหนี้ ได้แก่ รายได้ต่อเดือน ระยะเวลาสินเชื่อ และอัตราดอกเบี้ย ซึ่งเป็นตัวแปรที่มีน้ำหนักสูงสุดในแบบจำลอง จากงานวิจัยที่กล่าวมาข้างต้นจะใช้ข้อมูลพื้นฐานของลูกค้าสำหรับการสร้างแบบจำลองในการพยากรณ์ แต่ปัจจัยทางเศรษฐศาสตร์มหภาคเป็นปัจจัยอีกด้านหนึ่งที่สำคัญต่อปัญหาการผิดนัดชำระหนี้ และยังเป็นข้อมูลที่สามารถเก็บรวบรวมได้ง่าย

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงต้องการนำเทคนิคการเรียนรู้ของเครื่องแบบต่าง ๆ เพื่อสร้างแบบจำลองสำหรับพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล และนำปัจจัยทางเศรษฐศาสตร์มหภาคที่สำคัญมาใช้เพื่อปรับปรุงความแม่นยำของแบบจำลอง จากนั้นทำการเปรียบเทียบประสิทธิภาพของแบบจำลอง

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 ศึกษาปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

1.2.2 ศึกษาปัจจัยทางเศรษฐศาสตร์มหภาคที่ส่งผลต่อการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลเพื่อปรับปรุงความแม่นยำของแบบจำลองการเรียนรู้ของเครื่อง

1.2.3 พัฒนาแบบจำลองการเรียนรู้ของเครื่องโดยใช้อัลกอริทึม LightGBM, Random Forest และ XGBoost และเปรียบเทียบประสิทธิภาพของแบบจำลองในแง่ของความสามารถในการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

1.3 ความสำคัญของการวิจัย

แพลตฟอร์มให้บริการสินเชื่อออนไลน์เป็นที่ได้รับความนิยมอย่างมากในปัจจุบันเนื่องจากผู้ต้องการสินเชื่อสามารถเข้าหาแหล่งเงินทุนได้ง่ายขึ้น ทำให้เกิดปัญหาการสร้างหนี้เกินตัวขึ้นตามมามากมาย ผู้ให้กู้เองก็ต้องประสบกับปัญหาทางการเงินตามมาเมื่อผู้กู้ไม่สามารถชำระหนี้เงินกู้ได้ตรงตามนัดหมาย การพัฒนาอัลกอริทึมที่สามารถช่วยพยากรณ์ได้ว่าผู้กู้มีโอกาสที่จะผิดนัดชำระหนี้ได้นั้นจึงมีความสำคัญมาก เพื่อให้ผู้ให้กู้สามารถใช้ประกอบการตัดสินใจในการอนุมัติสินเชื่อให้กับลูกค้าและสามารถใช้เป็นระบบเตือนเมื่อลูกค้าที่ได้รับสินเชื่ออาจมีการผิดนัดชำระหนี้เกิดขึ้นในช่วงระหว่างการชำระหนี้ แม้กระทั่งผู้กู้เองก็สามารถใช้ประเมินคุณสมบัติของตนเองก่อนการขอสินเชื่อได้

1.4 ขอบเขตของการวิจัย

1.4.1 ผู้วิจัยทำการศึกษาข้อมูลสินเชื่อของบริษัท Bondora ซึ่งเป็นผู้ให้บริการสินเชื่อบนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (P2P) ชั้นนำของยุโรป และข้อมูลเศรษฐศาสตร์มหภาคซึ่งรวบรวมโดยกองทุนการเงินระหว่างประเทศ หรือ IMF (International Monetary Fund)

1.4.2 รวบรวมข้อมูลสินเชื่อจากข้อมูลรายงานสาธารณะของบริษัท Bondora จากเว็บไซต์ https://www.bondora.com/en/public-reports ซึ่งครอบคลุมช่วงเวลาระหว่างเดือนกุมภาพันธ์ พ.ศ. 2552 ถึงมกราคม พ.ศ. 2567 มีจำนวนข้อมูลทั้งหมด 358,092 รายการ และรวบรวมข้อมูลด้านเศรษฐศาสตร์มหภาคจากเว็บไซต์ https://www.imf.org/external/datamapper/datasets/WEO ประกอบด้วยตัวบ่งชี้ทางเศรษฐศาสตร์ 15 ตัว ของ 237 ประเทศทั่วโลก ครอบคลุมตั้งแต่ปี พ.ศ. 2523 ถึง พ.ศ. 2570

1.4.3 ตัวแปรที่ใช้ในการวิจัย ประกอบด้วย ตัวแปรอิสระ (Independent Variable) ได้แก่ ข้อมูลเบื้องต้นของลูกค้าสินเชื่อของ Bondora และข้อมูลเศรษฐศาสตร์มหภาค ตัวแปรตาม (Dependent Variable) คือ สถานการณ์ผิดนัดชำระหนี้ ได้แก่ ลูกค้าที่มีการผิดนัดชำระหนี้ และลูกค้าที่ไม่มีการผิดนัดชำระหนี้

1.4.4 เปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ที่สร้างโดยอัลกอริทึม LightGBM, Random Forest และ XGBoost ซึ่งประเมินด้วย Accuracy, Precision, Recall และ F1-Score

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้ทราบถึงปัจจัยพื้นฐานที่ส่งผลให้เกิดการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

1.5.2 ได้ทราบถึงปัจจัยที่สำคัญทางเศรษศาสตร์มหภาคที่ส่งผลให้เกิดการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

1.5.3 ได้แบบจำลองการเรียนรู้ของเครื่องที่มีประสิทธิภาพ สำหรับการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

1.6 นิยามศัพท์เฉพาะ

1.6.1 การผิดนัดชำระหนี้ (Default) คือ สถานการณ์ที่ผู้กู้ไม่สามารถชำระหนี้ตามเงื่อนไขที่ตกลงไว้ในสัญญาเงินกู้ การผิดนัดชำระหนี้อาจส่งผลให้ผู้กู้ต้องเผชิญกับบทลงโทษต่าง ๆ เช่น การฟ้องร้องทางกฎหมายหรือการเสียเครดิต การผิดนัดชำระหนี้เป็นปัจจัยสำคัญในการบริหารความเสี่ยงของผู้ให้กู้ และส่งผลต่อเสถียรภาพทางการเงินของระบบสินเชื่อโดยรวม

1.6.2 สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending) คือ กระบวนการให้กู้ยืมเงินที่เกิดขึ้นผ่านแพลตฟอร์มออนไลน์ ซึ่งเชื่อมต่อผู้กู้และผู้ให้กู้โดยตรงโดยไม่ต้องผ่านสถาบันการเงินแบบดั้งเดิม แพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ช่วยอำนวยความสะดวกในการทำธุรกรรมโดยกำหนดอัตราดอกเบี้ย เงื่อนไขการชำระหนี้ และจัดการความเสี่ยงที่เกี่ยวข้อง สินเชื่อออนไลน์ระหว่างบุคคลได้รับความนิยมมากขึ้นเนื่องจากความสามารถในการเข้าถึงสินเชื่อของผู้กู้ และผลตอบแทนที่สูงขึ้นสำหรับผู้ให้กู้เมื่อเทียบกับการฝากเงินในระบบธนาคารทั่วไป

1.6.3 การเรียนรู้ของเครื่อง (Machine Learning) คือ อัลกอริทึมของคอมพิวเตอร์ที่มีการพัฒนาให้สามารถเรียนรู้สิ่งต่าง ๆ และพัฒนาการทำงานให้ดีขึ้นได้ด้วยตนเองจากข้อมูลและสภาพแวดล้อมที่ได้รับจากการเรียนรู้ของระบบ โดยไม่ต้องมีมนุษย์คอยกำกับหรือเขียนโปรแกรมเพิ่มเติม การเรียนรู้ของเครื่องถูกมองว่าเป็นส่วนหนึ่งของปัญญาประดิษฐ์ โดยอัลกอริทึมสร้างแบบจำลองทางคณิตศาสตร์จากข้อมูลตัวอย่าง (ข้อมูลชุดเรียนรู้) เพื่อที่จะพยากรณ์หรือตัดสินใจได้อย่างชัดเจน

1.6.4 เศรษฐศาสตร์มหภาค (Macroeconomics) คือ การศึกษาระบบเศรษฐกิจโดยรวมซึ่งอธิบายปรากฏการณ์ทางเศรษฐกิจที่เกิดขึ้น ไม่ว่าจะเป็นการจ้างงาน การว่างงาน อัตราค่าจ้าง อัตราดอกเบี้ย อัตราแลกเปลี่ยนเงินตราระหว่างประเทศ ภาวะเงินเฟ้อ เศรษฐกิจถดถอย เป็นต้น เพื่อจะสามารถวางแนวนโยบายเพื่อยกระดับเศรษฐกิจโดยรวมหรือบรรลุเป้าหมายทางเศรษฐกิจที่ตั้งไว้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพยากรณ์การผิดนัดชำระหนี้เงินกู้ โดยจะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับสินเชื่อออนไลน์ระหว่างบุคคล เศรษฐศาสตร์มหภาค ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง โดยนำเสนอเนื้อหาตามลำดับดังต่อไปนี้

2.1 สินเชื่อออนไลน์ระหว่างบุคคล (Peer-to-Peer Lending)

2.2 เศรษฐศาสตร์มหภาค (Macroeconomic)

2.3 ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง (Workflow of Machine Learning)

2.4 อัลกอริทึม (Algorithm)

2.5 การประเมินประสิทธิภาพของแบบจำลอง (Models Evaluation)

2.6 การวิเคราะห์ความสัมพันธ์ (Correlation Analysis)

2.7 งานวิจัยที่เกี่ยวข้อง

2.1 สินเชื่อออนไลน์ระหว่างบุคคล (Peer-to-Peer Lending)

2.1.1 ประวัติความเป็นมา

นวัตกรรมทางการเงินรูปแบบนี้เกิดขึ้นครั้งแรกเมื่อเดือนกุมภาพันธ์ พ.ศ. 2548 ในประเทศอังกฤษ โดยบริษัทที่มีชื่อว่า Zopa เป็นตัวกลางในการเชื่อมต่อ ผู้กู้ กับ ผู้ให้กู้ ที่มีความต้องการ และเงื่อนไขต่าง ๆ ที่ตรงกันมาเจอกัน หลังจากผู้บริหารเล็งเห็นถึงปัญหาของการเข้าถึงแหล่งเงินทุนของธุรกิจขนาดกลางและขนาดเล็ก แพลตฟอร์มออนไลน์ดังกล่าวถือว่าได้รับความนิยมอย่างมาก มีรายงานว่าบริษัทสามารถทำเงินจากการให้กู้ยืมได้มากกว่า 1.5 พันล้านปอนด์ (P2P Lending Sites, 2022)

2.1.2 สินเชื่อออนไลน์ระหว่างบุคคล

สินเชื่อออนไลน์ระหว่างบุคคล หรือ Peer-to-Peer Lending (P2P Lending) เป็นการกู้ยืมเงินรูปแบบหนึ่งที่ให้บริการโดยตรงแก่บุคคลหรือธุรกิจโดยไม่มีสถาบันการเงินเป็นตัวกลางในข้อตกลง โดยทั่วไปการให้กู้ยืมแบบ P2P จะทำผ่านแพลตฟอร์มออนไลน์ที่จับคู่ผู้ให้กู้กับผู้ที่มีศักยภาพในการกู้ (CFI Team, 2022) ดังนั้น แทนที่ผู้กู้จะกู้เงินแบบดั้งเดิมจากธนาคารหรือเครดิตยูเนี่ยน ผู้กู้สามารถกู้ยืมเงินจากนักลงทุนรายย่อยที่ลงทะเบียนในแพลตฟอร์มการให้ยืมแบบ P2P เงินกู้ที่ได้จาก P2P อาจมาจากบุคคลคนเดียวทั้งหมด หรือมาจากการรวมกันของนักลงทุนหลายคน

การให้กู้ยืมแบบ P2P มีทั้งสินเชื่อที่มีหลักประกันและไม่มีหลักประกัน อย่างไรก็ตาม สินเชื่อส่วนใหญ่ในการให้กู้ยืมแบบ P2P เป็นสินเชื่อส่วนบุคคลที่ไม่มีหลักประกัน สินเชื่อที่มีหลักประกันนั้นเป็นที่พบเห็นได้ยากสำหรับอุตสาหกรรมนี้ และมักจะค้ำประกันด้วยสินค้าฟุ่มเฟือย ตัวอย่างเช่น บุคคลหรือธุรกิจสามารถกู้ยืมโดยใช้สินค้าฟุ่มเฟือยเป็นหลักค้ำประกันได้หลากหลายประเภท เช่น ทองและเงิน เครื่องประดับ นาฬิกาหรู เครื่องเพชร รถยนต์หรู งานศิลปะ ของเก่า กระเป๋าหรู ไวน์รส เลิศ และเครื่องดนตรี (Haldar & Tavener, 2015)

การให้กู้ยืมแบบ P2P เป็นตัวเลือกที่ดีสำหรับผู้กู้ที่มีเครดิตที่ไม่ดีซึ่งไม่มีคุณสมบัติในการขอสินเชื่อผ่านธนาคาร ช่วยให้ผู้กู้สามารถเข้าถึงแหล่งเงินทุนที่ต้องการได้โดยไม่ต้องกู้เงินนอกระบบ หรือแม้แต่สำหรับเจ้าของธุรกิจที่ต้องการกู้เงินก็เป็นตัวเลือกที่ดี เนื่องจากมีขั้นตอนที่สะดวกและรวดเร็วกว่าการสมัครผ่านธนาคารมาก และสามารถรับเงินได้ภายในหนึ่งสัปดาห์หลังจากได้รับอนุมัติ (Johnson, 2022)

2.1.3 การทำงานของสินเชื่อออนไลน์ระหว่างบุคคล

วิธีการทำงานของ P2P (Capital One, 2022) มีวิธีการที่แตกต่างกันออกไปในแต่ละแพลตฟอร์ม แต่โดยทั่วไปแล้ว จะมีวิธีการดำเนินการดังนี้:

- ตรวจสอบคุณสมบัติสำหรับเงินกู้ที่ต้องการให้ตรงตามแพลตฟอร์ม P2P ที่เลือก
- ส่งใบสมัครขอสินเชื่อ P2P การสมัครขอสินเชื่ออาจมีการสอบถามเชิงลึกเกี่ยวกับรายงานเครดิตของผู้สมัครขอสินเชื่อ
- ตรวจสอบสถานะการอนุมัติสินเชื่อ
- หากใบสมัครได้รับการอนุมัติ นักลงทุนจะตรวจสอบเงินกู้เพื่อตัดสินใจว่าจะให้เงินทุนหรือไม่
- เมื่อเงินกู้ของผู้สมัครขอสินเชื่อได้รับเงินกู้เต็มจำนวน ผู้สมัครขอสินเชื่อจะได้รับการโอนเงินกู้ผ่านช่องทางอิเล็กทรอนิกส์
- ผู้กู้ทำการชำระคืนเงินกู้
- เงินกู้ที่ได้รับการชำระคืน ก็จะถูกส่งไปยังนักลงทุนเงินกู้ต่อไป

2.1.4 ความเสี่ยงของสินเชื่อออนไลน์ระหว่างบุคคล

ถึงแม้ว่าสินเชื่อออนไลน์ระหว่างบุคคลนั้นทำให้ผู้ต้องการเงินเพื่อไปใช้ในการลงทุนนั้นสามารถเข้าถึงเงินทุนได้ง่าย อนุมัติไว และได้ดอกเบี้ยที่ต่ำ หรือแม้แต่เป็นแหล่งลงทุนที่ง่ายสำหรับนักลงทุนที่ต้องการปล่อยสินเชื่อ แต่สินเชื่อออนไลน์ระหว่างบุคคลนั้นยังคงมีความเสี่ยงที่ตามมามากมายที่ส่งผลต่อทั้งผู้กู้และผู้ให้กู้เช่นกัน (Shadroch, 2022)

ความเสี่ยงของผู้กู้

1. ความปลอดภัย (Security)

มีความเสี่ยงด้านความปลอดภัยทางไซเบอร์ โดยเฉพาะอย่างยิ่งกับข้อมูลส่วนบุคคลที่แสดงเพื่อขอรับสินเชื่อ เนื่องจากแพลตฟอร์ม P2P ดูแลโดยบริษัทเอกชนทางออนไลน์ จึงไม่สามารถแน่ใจได้มากนักว่าจะไม่มีการละเมิดความปลอดภัยของเอกสารส่วนบุคคลที่สำคัญ

2. อัตราดอกเบี้ย (Interest Rates)

ไม่มีอัตราดอกเบี้ยคงที่ในทุกแพลตฟอร์ม P2P อัตราดอกเบี้ยถูกกำหนดโดยขึ้นอยู่กับคะแนนเครดิตและประวัติบริษัทของผู้กู้ ดังนั้นผู้กู้อาจเสี่ยงต่อการได้รับอัตราดอกเบี้ยที่สูงหากมีคะแนนเครดิตต่ำ

3. ขาดนักลงทุน (Lack of Investors)

แน่นอนว่าตอนนี้อาจมีนักลงทุนใน P2P มากกว่าในปี พ.ศ. 2548 แต่นั่นหมายความว่ามีการแข่งขันมากขึ้นเช่นกัน อาจเป็นไปได้ยากที่จะพบนักลงทุนที่ให้วงเงินกู้ได้ตามที่ต้องการ ผู้ที่หวังพึ่งเงินทุนผ่านช่องทาง P2P เพียงอย่างเดียวนั้นอาจเป็นความเสี่ยงจนเกินไป

4. ความบาดหมางกับผู้ให้กู้ (Feud with Lenders)

เกิดขึ้นในกรณีที่ผู้กู้มีการผิดนัดชำระหนี้เป็นประจำ ผู้ให้กู้มีสิทธิ์ที่จะดึงยอดเงินกู้ที่เหลือของผู้กู้ได้ แพลตฟอร์ม P2P บางแห่งจะส่งเรื่องไปยังศูนย์ติดตามหนี้ หรือแนะนำวิธีการนี้ให้กับผู้ให้กู้ อย่างไรก็ตาม ไม่มีการรับประกันว่าผู้ให้กู้จะต้องปฏิบัติตามวิธีการดังกล่าว ดังนั้นผู้ให้กู้บางรายอาจใช้วิธีการที่ผิดกฎหมายเพื่อติดตามหรือเรียกคืนจำนวนเงินกู้ที่เหลืออยู่จากผู้กู้ได้

ความเสี่ยงของผู้ให้กู้

1. ความปลอดภัย (Security)

ผู้ให้กู้อาจตกเป็นเหยื่อของการละเมิดความปลอดภัยทางไซเบอร์ แม้ว่าแพลตฟอร์ม P2P จะได้รับการตรวจสอบมาก่อน แต่ก็ยังมีกฎระเบียบที่น้อยกว่าธนาคารแบบดั้งเดิม ดังนั้น ความปลอดภัยของเงินทุนและข้อมูลลับของผู้ให้กู้จึงมีความเสี่ยงจากแพลตฟอร์ม P2P

2. การชำระคืน (Repayment)

ผู้กู้ไม่จำเป็นต้องเสนอหลักประกันเพื่อค้ำประกันเงินกู้ และผู้กู้อาจมีคะแนนเครดิตต่ำกว่าที่ธนาคารกำหนด ดังนั้นจึงมีความเสี่ยงสูงที่จะไม่ได้รับการชำระเงินคืน มีกฎระเบียบน้อยมากสำหรับการยืนยันว่าผู้กู้จะชำระเงินคืนได้ครบตามกำหนด นอกจากนี้ยังไม่มีการรับประกันเงินกู้ ดังนั้นผู้ให้กู้จะต้องรับภาระหากกรณีที่มีการผิดนัดชำระหนี้

3. สภาพคล่อง (Liquidity)

การรอผลตอบแทนของการลงทุนอาจนานขึ้นจากการให้กู้ยืมแบบ P2P เงินกู้จะได้รับการชำระคืนเป็นงวด ดังนั้นผู้ให้กู้จึงไม่สามารถเรียกคืนเงินก้อนใหญ่ได้จนกว่าจะครบกำหนด ดอกเบี้ยเป็นผลตอบแทนหลักของผู้ให้กู้ ซึ่งจะเกิดขึ้นเมื่อเวลาผ่านไปในแต่ละงวดเท่านั้น ดังนั้น การลงทุนแบบ P2P จึงไม่ใช่สิ่งที่ควรพิจารณา หากกำลังมองหาผลตอบแทนจากการลงทุนที่รวดเร็ว

4. การสูญเสียผลตอบแทน (Loss of Returns)

นอกเหนือจากที่อาจเกิดการสูญเสียจากการผิดนัดชำระหนี้ ยังมีโอกาสที่ผู้กู้จะชำระคืนเงินกู้ก่อนกำหนด เนื่องจากแพลตฟอร์ม P2P ส่วนใหญ่มีเงื่อนไขการชำระคืนที่ยืดหยุ่น หากสิ่งนี้เกิดขึ้นอาจทำให้สูญเสียผลตอบแทนจากดอกเบี้ยที่ควรได้รับหากเงินกู้นั้นเป็นเงินกู้ระยะยาว

5. แอปพลิเคชันปลอม (Fraudulent Applications)

แม้ว่าแพลตฟอร์ม P2P จะตรวจสอบความถูกต้องของทุกแอปพลิเคชัน แต่ก็ยังมีความเสี่ยงที่บางแอพพลิเคชันอาจเป็นของปลอมที่มิจฉาชีพจงใจสร้างขึ้นมา ไม่มีมาตรการความปลอดภัยรองรับเมื่อนำเงินไปลงทุนในแอพพลิเคชันปลอม จึงควรระมัดระวังและตรวจสอบความถูกต้องให้ดีก่อนการลงทุน

2.2 เศรษฐศาสตร์มหภาค (Macroeconomic)

เศรษฐศาสตร์มหภาค (Intrend, 2019) คือ พฤติกรรมทางด้านเศรษฐกิจและการใช้ทรัพยากรในระดับกว้าง ทั้งใน ระดับประเทศ ทวีป ภูมิภาค หรือระดับโลก ซึ่งรูปแบบการทำงานนั้นจะมีตั้งแต่ การศึกษาเรื่องราวของอัตราจ้างงาน รายได้ของประชาชาติ ราคาของสินค้าและบริการ ผลผลิตโดยรวม การนำเข้า การส่งออก พฤติกรรมการใช้จ่าย และ ดุลการค้าภายในประเทศ

เศรษฐศาสตร์มหภาค จะเน้นในด้านการศึกษาเศรษฐกิจ หรือเฝ้าดูในรูปแบบภาครวมเท่านั้น ทั้งอุปสงค์มวลรวม อุปทานมวลรวม อัตราเงินเฟ้อ อัตราการว่างงาน นโยบายการคลัง นโยบายการเงิน และการเติบโตของเศรษฐกิจภายในประเทศ เป็นหลัก

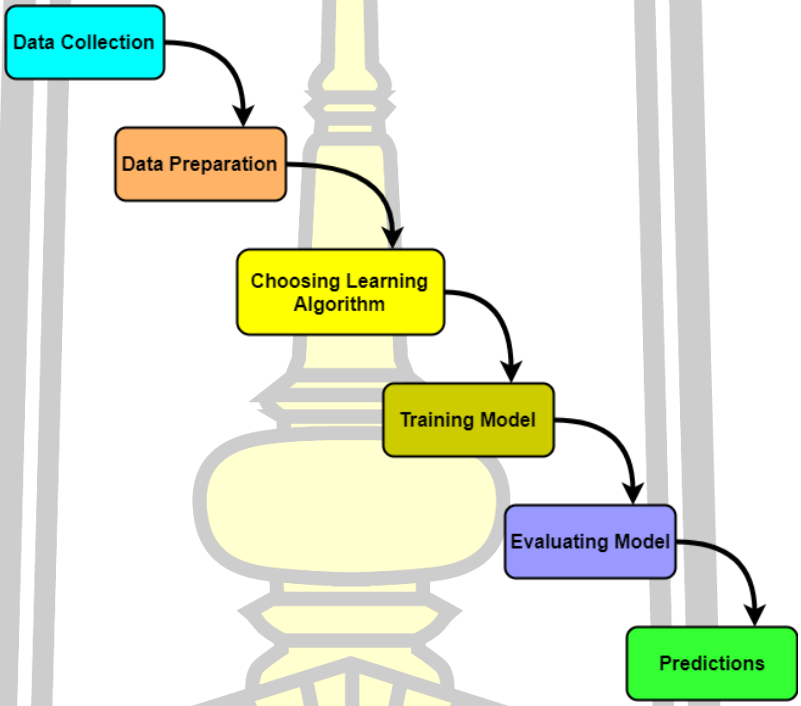
นอกจากนี้ยังจะต้องมีการรู้จักถึง สภาพการณ์ของฐานข้อมูลมวลรวมที่มีอยู่ เพื่อสะท้อนและนำมาวางนโยบายในด้านต่าง ๆ ได้ กล่าวคือการศึกษาทางด้าน เศรษฐศาสตร์ ย่อมเป็นการสร้างความได้เปรียบทางการแข่งขันในตลาดธุรกิจ รวมไปถึงความได้เปรียบในด้านการดำเนินชีวิตประจำวัน ในการจับจ่ายซื้อของบริโภคสินค้าที่มีอยู่อย่างจำกัด ให้เกิดประโยชน์สูงสุด ไปจนถึงการประเมินสถานการณ์ และปรับตัวเข้ากับสภาพเศรษฐกิจที่เปลี่ยนแปลงได้ อีกทั้งยังสามารถใช้ประโยชน์เพื่อการเตรียมการรับมือกับสภาวะต่าง ๆ ที่เกิดขึ้นจากการประเมินแนวโน้มความน่าจะเป็นภายในอนาคต อีกทั้งยังสามารถนำองค์ความรู้ไปศึกษาอดีต และวิเคราะห์เป็นบทเรียนได้อีกด้วยเช่นกัน

2.3 ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง (Workflow of Machine Learning)

การเรียนรู้ของเครื่อง (Great Learning Team, 2022) คือ การประยุกต์ใช้ปัญญาประดิษฐ์ที่ใช้เทคนิคทางสถิติเพื่อให้คอมพิวเตอร์สามารถเรียนรู้และตัดสินใจได้โดยไม่ต้องตั้งโปรแกรมไว้อย่าง

ชัดเจน มีความเชื่อที่ว่าคอมพิวเตอร์สามารถเรียนรู้จากข้อมูล สามารถระบุรูปแบบ และตัดสินใจได้โดยได้รับความช่วยเหลือเพียงเล็กน้อยจากมนุษย์

การเรียนรู้ของเครื่องเป็นสาขาย่อยของปัญญาประดิษฐ์ เป็นการศึกษาเกี่ยวกับการทำให้เครื่องจักรมีความคล้ายมนุษย์มากขึ้นในด้านพฤติกรรมและการตัดสินใจ โดยให้ความสามารถในการเรียนรู้และพัฒนาโปรแกรมของตนเอง สิ่งนี้ทำได้โดยมีการช่วยเหลือจากมนุษย์เพียงเล็กน้อย นั่นคือไม่มีการเขียนโปรแกรมที่ชัดเจน กระบวนการเรียนรู้เป็นแบบอัตโนมัติและปรับปรุงตามประสบการณ์ของเครื่องจักรตลอดกระบวนการ โดยสรุปขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่องได้เป็นขั้นตอนหลัก ๆ ดังแสดงในรูปที่ 1 ดังนี้



รูปที่ 1 ขั้นตอนการดำเนินงานของการเรียนรู้ของเครื่อง

ที่มา: Gangoda and Wanninayake (2023)

1. การรวบรวมข้อมูล (Data Collection) คือ รวบรวมข้อมูลทั้งหมดที่ต้องการสำหรับแบบจำลองที่ต้องการ ไม่ว่าจะมาจากองค์กร แหล่งข้อมูลสาธารณะหรือข้อมูลที่ต้องเสียค่าบริการ คุณภาพและปริมาณของข้อมูลที่รวบรวมมีผลโดยตรงต่อประสิทธิภาพของแบบจำลองที่ต้องการ ขั้นตอนนี้มีความสำคัญมากเนื่องจากคุณภาพและปริมาณของข้อมูลที่รวบรวมจะมีผลโดยตรงต่อประสิทธิภาพในการพยากรณ์ของแบบจำลอง

2. การเตรียมข้อมูล (Data Preparation) (ProjectPro, 2023) คือ กระบวนการจัดการกับข้อมูลดิบที่ถูกรวบรวมมาให้สามารถนำมาใช้วิเคราะห์ต่อได้ เนื่องจากข้อมูลดิบอาจมีค่าที่ขาดหายไป (Missing Values) ข้อมูลที่ไม่สอดคล้องกัน (Inconsistent Values) ข้อมูลที่ซ้ำกัน (Duplicate

Instances) และอัลกอริทึมการเรียนรู้ของเครื่องเป็นอัลกอริทึมทางคณิตศาสตร์ที่ใช้การดำเนินการทางคณิตศาสตร์เพื่อสร้างระบบการพยากรณ์ การดำเนินการต่าง ๆ สามารถทำได้กับข้อมูลที่เป็นตัวเลขเท่านั้น ดังนั้นชุดข้อมูลที่ใช้ควรเป็นตัวเลขเท่านั้นก่อนที่จะส่งเป็นข้อมูลนำเข้าให้กับอัลกอริทึม ขั้นตอนการเตรียมข้อมูลสำหรับกระบวนการเรียนรู้ของเครื่องมีดังนี้

กระบวนการสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis) คือ ขั้นตอนในการวิเคราะห์ข้อมูล โดยใช้พารามิเตอร์ทางสถิติ เช่น ค่าเฉลี่ย มัธยฐาน ฐานนิยม ส่วนเบี่ยงเบนมาตรฐาน ฯลฯ และการแสดงข้อมูลออกมาในรูปแบบกราฟ การวิเคราะห์ข้อมูลช่วยระบุค่าที่ขาดหายไป (Missing Values) ค่าผิดปกติ (Outliers) และความไม่สอดคล้องกันระหว่างตัวแปรต่าง ๆ นอกจากนี้ยังแสดงให้เห็นถึงการกระจายตัวของข้อมูล

การแปลงข้อมูล (Data Transformation) คือ การจัดการกับข้อมูลดิบให้อยู่ในรูปแบบที่รองรับกับวิธีการวิเคราะห์ข้อมูล อัลกอริทึมการเรียนรู้ของเครื่องเป็นการดำเนินการทางคณิตศาสตร์ ดังนั้นจึงจำเป็นต้องแปลงข้อมูลทั้งหมดเป็นรูปแบบตัวเลข ประเภทของข้อมูลที่ต้องมีการแปลงข้อมูล เช่น ข้อมูลที่เป็นตัวอักษร (String) ข้อมูลหมวดหมู่ (Categorical) หรือข้อมูลที่ไม่ใช่ตัวเลข เป็นต้น ซึ่งไม่สามารถดำเนินการทางคณิตศาสตร์ได้ จึงต้องมีการแปลงข้อมูลคุณลักษณะเหล่านั้น โดยส่วนใหญ่มักใช้การแทนที่ข้อมูลด้วยตัวเลข เช่น คุณลักษณะเพศ ประกอบด้วย เพศชาย และเพศหญิง ทำการแทนที่ข้อมูลด้วยตัวเลขโดยให้ 0 แทนเพศชาย 1 แทนเพศหญิง

การทำความสะอาดข้อมูล (Data Cleaning) คือ กระบวนการแก้ไขหรือลบข้อมูลที่ไม่ถูกต้อง เสียหาย จัดรูปแบบไม่ถูกต้อง ทำซ้ำ หรือไม่สมบูรณ์ภายในชุดข้อมูล เมื่อมีการรวมแหล่งข้อมูลจากหลายแหล่งเข้าด้วยกัน มีโอกาสสูงที่ข้อมูลจะถูกทำซ้ำหรือกำกับผิด หากข้อมูลไม่ถูกต้อง ผลลัพธ์และอัลกอริทึมจะไม่น่าเชื่อถือ แม้ว่าจะดูเหมือนถูกต้องก็ตาม ไม่มีวิธีการที่แน่นอนในการกำหนดขั้นตอนที่แน่นอนในกระบวนการล้างข้อมูล เนื่องจากกระบวนการจะแตกต่างกันไปในแต่ละชุดข้อมูล

การเลือกคุณลักษณะ (Feature Selection) เป็นการลดตัวแปรของข้อมูลนำเข้าให้กับแบบจำลองโดยเลือกใช้เฉพาะคุณลักษณะบางอย่างที่เกี่ยวข้องกับแบบจำลอง เพื่อลดความซับซ้อนในการคำนวณของอัลกอริทึม ในแรกอาจคัดเลือกโดยการใช้ตรรกะและเหตุผลพื้นฐานจากการพิจารณาชุดข้อมูล จากนั้นจึงใช้วิธีการเลือกคุณลักษณะที่ได้รับความนิยมในวิธีการเรียนรู้ของเครื่อง เช่น Intrinsic Methods Wrapper Methods หรือ Filter Methods

3. การเลือกอัลกอริทึม (Choosing Learning Algorithm) คือ การเลือกอัลกอริทึมการเรียนรู้ของเครื่องที่มีประสิทธิภาพและเหมาะสมกับประเภทของปัญหาที่ต้องแก้ไขและประเภทของข้อมูลที่มีอยู่ เนื่องจากอัลกอริทึมที่แตกต่างกันจะให้ผลลัพธ์ที่แตกต่างกันแม้จะใช้ชุดข้อมูลเดียวกันก็ตาม การเลือกอัลกอริทึมจึงมีเป้าหมายเพื่อที่จะสามารถฝึกแบบจำลองที่มีประสิทธิภาพดีที่สุด ภาพรวมของอัลกอริทึมการเรียนรู้ของเครื่องแสดงในรูปที่ 1

4. การฝึกแบบจำลอง (Training Model) คือ เป็นการป้อนข้อมูลให้กับแบบจำลองเพื่อเรียนรู้ โดยชุดข้อมูลที่ได้จากขั้นตอนการเตรียมข้อมูลจะถูกแบ่งออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลการเรียนรู้ (Training Set) และชุดข้อมูลการทดสอบ (Testing Set) โดยนิยมแบ่งออกเป็นสัดส่วน 80/20 หรือ 70/30 ตามลำดับ ซึ่งข้อมูลชุดการเรียนรู้จะถูกใช้สำหรับป้อนเข้าไปในแบบจำลองการเรียนรู้ของเครื่องเพื่อทำการเรียนรู้ ค้นหารูปแบบ ชุดข้อมูลการทดสอบจะถูกใช้เพื่อตรวจสอบความถูกต้องของแบบจำลองที่สร้างขึ้น

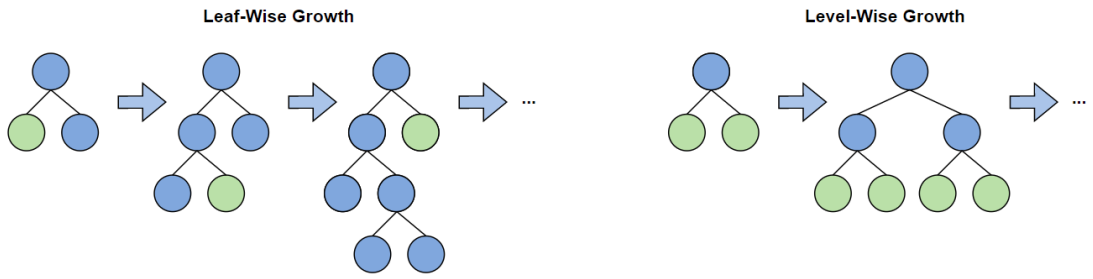
5. การประเมินแบบจำลอง (Evaluation Model) คือ เป็นกระบวนการของการใช้เมตริกการประเมินต่าง ๆ เพื่อประเมินประสิทธิภาพของแบบจำลอง ตลอดจนจุดแข็งและจุดอ่อน โดยใช้กับชุดข้อมูลการทดสอบ (Testing Set) การประเมินแบบจำลองเป็นสิ่งสำคัญในการประเมินประสิทธิภาพของแบบจำลองในระหว่างขั้นตอนการวิจัยเบื้องต้น และยังมีบทบาทในการตรวจสอบแบบจำลองอีกด้วย

6. การพยากรณ์ (Predictions) คือ การนำแบบจำลองที่ผ่านการประเมินไปใช้ประโยชน์เพื่อพยากรณ์ข้อมูลจริง หรือสถานการณ์จริง

2.4 อัลกอริทึม (Algorithm)

2.4.1 อัลกอริทึม LightGBM

LightGBM หรือ Light Gradient Boosting Machine (Ke et al., 2017) เป็นเฟรมเวิร์คของ Gradient Boosting โดยอิงตามแผนผังต้นไม้ตัดสินใจเพื่อเพิ่มประสิทธิภาพของแบบจำลองและลดการใช้หน่วยความจำในการคำนวณ สามารถใช้ได้กับทั้งปัญหาการจำแนกประเภทและปัญหาการถดถอย LightGBM ใช้เทคนิค 2 เทคนิค ได้แก่ Gradient-based One Side Sampling (GOSS) และ Exclusive Feature Bundling (EFB) ประกอบเข้าด้วยกันเพื่อทำให้แบบจำลองทำงานได้อย่างมีประสิทธิภาพ การดำเนินการบนต้นไม้ตัดสินใจนั้นจะใช้วิธีการ Leaf-Wise ซึ่งสามารถลดการสูญเสียได้มากกว่าวิธีการ Level-Wise ด้วยเหตุนี้จึงให้ผลลัพธ์ที่มีความแม่นยำที่ดีและมีความรวดเร็วในการคำนวณมากกว่า

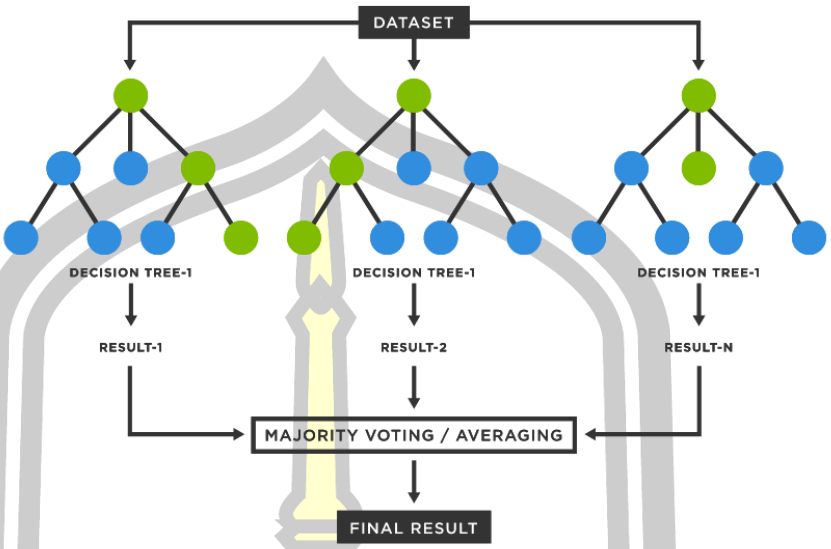


รูปที่ 2 วิธีการ Leaf-Wise และวิธีการ Level-Wise

ที่มา: (Wong, 2022)

2.4.2 อัลกอริทึม Random Forest

อัลกอริทึม Random Forest (Azhari et al., 2019) เป็นวิธีการหนึ่งของการเรียนรู้ของเครื่องที่นิยมใช้กันอย่างแพร่หลาย และใช้ได้ดีในหลาย ๆ ปัญหา ทั้งกับปัญหาแบบ Regression และ Classification โดย Random Forest เป็นขั้นตอนวิธีพัฒนาต่อยอดมาจาก Decision Tree ต่างกันที่ Random Forest เป็นการเพิ่มจำนวนต้นไม้ (Tree) เป็นหลาย ๆ ต้นทำให้มีประสิทธิภาพพยากรณ์สูงขึ้น Random Forest มีหลักการทำงาน คือ จะแบ่งข้อมูลออกเป็นต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ ต้น โดยแต่ละต้นจะได้รับคุณลักษณะ (Feature) และข้อมูล (Data) ที่ไม่เหมือนกันทั้งหมด ซึ่งทำการสุ่มมาจากชุดข้อมูลการเรียนรู้ เพื่อทำให้ได้ต้นไม้ที่มีความหลากหลายและมีความอิสระต่อกันมากขึ้น จากนั้นทำการหาค่าพยากรณ์ โดยค่าพยากรณ์ที่ได้จะเป็นการให้ต้นไม้ตัดสินใจแต่ละต้นหาค่าพยากรณ์ จากนั้นค่าพยากรณ์สุดท้าย ในกรณีที่เป็นปัญหาการจำแนกประเภท (Classification) จะใช้วิธีผลโหวตมากที่สุด (Majority Vote) โดยค่าพยากรณ์ของต้นไม้ตัดสินใจต้นได้รับค่าผลโหวตมากที่สุด จะถูกเลือกให้เป็นค่าพยากรณ์ของปัญหา แต่ถ้าเป็นปัญหาวิเคราะห์การถดถอย (Regression) จะใช้วิธีคำนวณหาค่าเฉลี่ย (Mean) โดยนำเอาค่าพยากรณ์ของทุกต้นไม้ตัดสินใจมาคำนวณหาค่าเฉลี่ยเพื่อแสดงเป็นค่าพยากรณ์ของปัญหา แสดงในรูปที่ 3



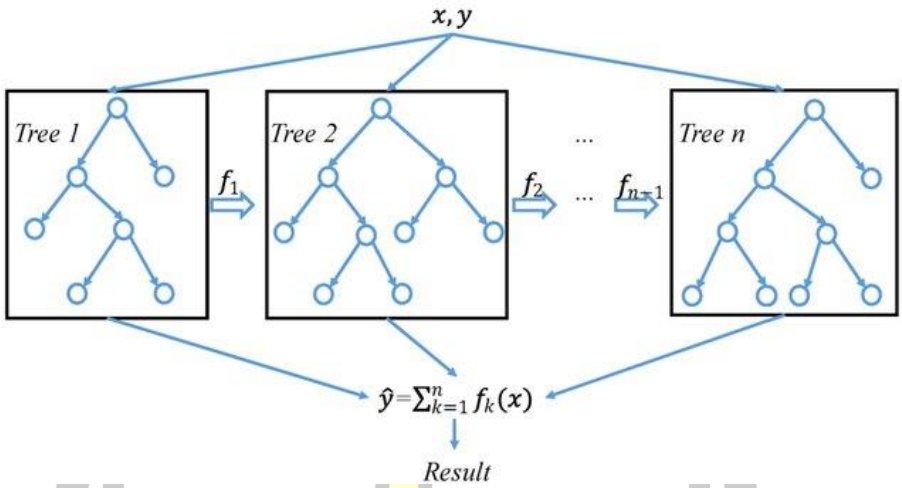
รูปที่ 3 อัลกอริทึม Random Forest

ที่มา : (TIBCO, 2022)

2.4.3 อัลกอริทึม XGBoost

XGBoost (Friedman, 2001) ซึ่งย่อมาจาก Extreme Gradient Boosting เป็นอัลกอริทึมการเรียนรู้ของเครื่องที่ได้รับความนิยมและทรงพลัง ซึ่งมักใช้ในการสร้างแบบจำลองเชิงพยากรณ์และการวิเคราะห์ข้อมูล เหมาะอย่างยิ่งสำหรับปัญหาที่มีคุณลักษณะอินพุตจำนวนมากและความสัมพันธ์ที่ซับซ้อนระหว่างกัน และสามารถพยากรณ์ตัวแปรตามที่เป็นตัวแปรที่มี 2 ค่า (Binary Variable) หรือตัวแปรต่อเนื่อง (Continuous Variable) ได้อย่างแม่นยำ

XGBoost มีพื้นฐานมาจากอัลกอริทึมการ Gradient Boosting ซึ่งทำงานโดยการเรียนรู้ซ้ำ ๆ และเพิ่มแผนผังต้นไม้ตัดสินใจที่ประสิทธิภาพต่ำให้กับแบบจำลอง ในการวนซ้ำแต่ละครั้ง อัลกอริทึมจะคำนวณความคลาดเคลื่อนของแบบจำลองก่อนหน้าและฝึกโมเดลใหม่เพื่อแก้ไขความคลาดเคลื่อนเหล่านั้น XGBoost ใช้วิธีการหาค่าที่เหมาะสมที่สุดซึ่งรวมแผนผังต้นไม้ตัดสินใจหลาย ๆ แบบเข้าด้วยกัน และไม่รวมกันมากเกินไปเพื่อไม่ให้เกิดความซับซ้อนของโมเดล



รูปที่ 4 อัลกอริทึม XGBoost
ที่มา : (Wang et al., 2019)

2.5 การประเมินประสิทธิภาพของแบบจำลอง (Models Evaluation)

เกณฑ์การประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องมีอยู่หลากหลายวิธี (Hossin & Sulaiman, 2015) ในงานวิจัยนี้ จะใช้เกณฑ์ในการประเมินประสิทธิภาพที่เป็นที่ยอมรับโดยทั่วไป ได้แก่ Accuracy, Precision, Recall และ F1-score เกณฑ์ในการวัดแต่ละตัวนั้นมาจาก Confusion Matrix ที่แสดงความสัมพันธ์ระหว่างข้อมูลจริง (Actual Class) และข้อมูลที่ได้จากการพยากรณ์ (Predicted Class) ซึ่งแสดงในรูปที่ 5

		Predicted Class	
		Predicted Positive Class	Predicted Negative Class
Actual Class	Actual Positive Class	True Positive (TP)	False Negative (FN)
	Actual Negative Class	False Positive (FP)	True Negative (TN)

รูปที่ 5 ตาราง Confusion Matrix

- True Positive (TP): ผลลัพธ์ที่ข้อมูลเป็นจริง และแบบจำลองพยากรณ์ว่าเป็นจริง
- True Negative (TN): ผลลัพธ์ที่ข้อมูลเป็นไม่เป็นจริง และแบบจำลองพยากรณ์ว่าไม่เป็นจริง

- False Positive (FP): ผลลัพธ์ที่ข้อมูลไม่เป็นจริง แต่แบบจำลองพยากรณ์ว่าเป็นจริง
- False Negative (FN): ผลลัพธ์ที่ข้อมูลเป็นจริง แต่แบบจำลองพยากรณ์ว่าไม่เป็นจริง

โดยทั่วไป Accuracy จะวัดอัตราส่วนของการพยากรณ์ที่ถูกต้อง จากจำนวนข้อมูลทั้งหมดที่ประเมิน โดยมีสูตรในการคำนวณ ดังนี้

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

Precision ใช้ในการวัดรูปแบบ Positive ที่พยากรณ์ได้อย่างถูกต้อง จากจำนวนข้อมูลทั้งหมดที่พยากรณ์ได้เป็นรูปแบบ Positive โดยมีสูตรในการคำนวณ ดังนี้

$$Precision = \frac{TP}{TP + FP}$$

Recall ใช้ในการวัดรูปแบบ Positive ที่พยากรณ์ได้อย่างถูกต้อง จากจำนวนข้อมูลจริงทั้งหมดที่เป็นรูปแบบ Positive โดยมีสูตรในการคำนวณ ดังนี้

$$Recall = \frac{TP}{TP + FN}$$

F1-score แสดงถึงค่าเฉลี่ยฮาร์มอนิกระหว่าง Recall และ Precision

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

2.6 การวิเคราะห์ความสัมพันธ์ (Correlation Analysis)

การวิเคราะห์ความสัมพันธ์เป็นกระบวนการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวหรือมากกว่า เพื่อตรวจสอบว่าตัวแปรเหล่านั้นว่ามีความแปรผันต่อกันมากน้อยเพียงใด ซึ่งเป็นวิธีการที่มีความสำคัญที่ช่วยให้นักวิทยาศาสตร์ข้อมูลสามารถเลือกใช้ตัวแปรในการพัฒนาแบบจำลองการเรียนรู้ของเครื่องได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ใช้วิธีการหาค่าสหสัมพันธ์แบบเพียร์สัน (Pearson Correlation) และวิธีการหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point Biserial Correlation) ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้านเศรษฐศาสตร์มหภาค และตัวแปรตาม เพื่อคัดเลือกตัวแปรด้านเศรษฐศาสตร์มหภาคที่เหมาะสมสำหรับการพัฒนาแบบจำลองการเรียนรู้ของเครื่องเพื่อพยากรณ์การผิดนัดชำระหนี้ บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล โดยค่าสหสัมพันธ์จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 หมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้าม หากมีค่าใกล้ +1.0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก (LeBlanc & Cox, 2017)

2.6.1 วิธีการหาค่าสหสัมพันธ์แบบเพียร์สัน (Pearson Correlation)

สหสัมพันธ์แบบเพียร์สัน (r) เป็นการหาความสัมพันธ์ระหว่างสองตัวแปร โดยที่ตัวแปรทั้งสองเป็นตัวแปรต่อเนื่อง (Continuous) โดยมีสูตรในการคำนวณ ดังนี้

$$r = \frac{n\sum(x_i y_i) - \sum x_i \cdot \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

เมื่อ n แทน จำนวนคู่ของข้อมูลทั้งหมด

x_i แทน ค่าของตัวแปร x ลำดับที่ i โดยที่ $i = 1,2,\dots,n$

y_i แทน ค่าของตัวแปร y ลำดับที่ i โดยที่ $i = 1,2,\dots,n$

2.6.2 วิธีการหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point Biserial Correlation)

สหสัมพันธ์แบบพอยท์ไบซีเรียล (r_{pb}) เป็นการหาความสัมพันธ์ระหว่างสองตัวแปร โดยที่ตัวแปรหนึ่งเป็นตัวแปรต่อเนื่อง (Continuous) และอีกตัวแปรหนึ่งมี 2 ลักษณะจริง (True Dichotomous) ที่มี 2 กลุ่ม เช่น 0 คือ กลุ่มลูกค้าที่มีการผิดนัดชำระหนี้ และ 1 คือ กลุ่มลูกค้าที่ไม่มีการผิดนัดชำระหนี้ในงานวิจัยนี้ โดยมีสูตรในการคำนวณ ดังนี้

$$r_{pb} = \sqrt{pq} \cdot \frac{\mu_1 - \mu_0}{\sigma}$$

เมื่อ p แทน สัดส่วนของจำนวนข้อมูลกลุ่มในกลุ่มหนึ่งกับจำนวนข้อมูลทั้งหมด

q แทน สัดส่วนของจำนวนข้อมูลกลุ่มในอีกกลุ่มหนึ่งกับจำนวนข้อมูลทั้งหมด

μ_1 แทน ค่าเฉลี่ยของตัวแปรต่อเนื่องในกลุ่มที่กำหนดให้เป็น 1

μ_0 แทน ค่าเฉลี่ยของตัวแปรต่อเนื่องในกลุ่มที่กำหนดให้เป็น 0

σ แทน ส่วนเบี่ยงเบนมาตรฐานของตัวแปรต่อเนื่อง

2.7 งานวิจัยที่เกี่ยวข้อง

Çığşar and Ünal (2019) ได้ประยุกต์ใช้อัลกอริทึมการจำแนกประเภทเพื่อศึกษาผลกระทบของลักษณะทางประชากรศาสตร์ เศรษฐกิจ และสังคมบางประการที่ส่งผลต่อการผิดนัดชำระหนี้ ชุดข้อมูลที่ใช้ได้มาจากแบบสำรวจของสำนักงานสถิติประเทศตุรกี (Turkish Statistical Institute) ในปี พ.ศ. 2558 ข้อมูลดิบประกอบด้วยการแบบสำรวจ 59,663 รายการ ซึ่งใช้เพียง 20,275 รายการและคุณลักษณะ 12 รายการโดยคัดเลือกเฉพาะหัวหน้าครัวเรือนที่มีอายุมากกว่า 15 ปี ผู้วิจัยได้ทำการสร้างแบบจำลองในการพยากรณ์และทำการเปรียบเทียบประสิทธิภาพโดยใช้อัลกอริทึมที่แตกต่างกัน 6 อัลกอริทึม ได้แก่ Bayes networks, Naïve Bayes, Logistic Regression, Decision Tree J48,

Random Forest และ Multilayer Perceptron ซึ่งประเมินประสิทธิภาพของแต่ละแบบจำลองโดยใช้ Accuracy, RMSE, Precision, Recall , ROC Curve และ F-measure ผลการวิจัยพบว่า Logistic Regression มีประสิทธิภาพในการพยากรณ์สูงที่สุด

Sabilla et al. (2019) ได้ศึกษาการคัดเลือกคุณลักษณะสำหรับการวิเคราะห์กลิ่นด้วยระบบ e-nose ซึ่งใช้ข้อมูลที่รวบรวมจากเซนเซอร์วัดกลิ่น 10 ตัว เพื่อจำแนกกลิ่นอาหาร โดยใช้ Pearson Correlation ในการประเมินความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยใช้ Threshold = 0.95 เป็นค่าเริ่มต้น และจะคัดเลือกตัวแปรที่มีค่าความสัมพันธ์มากกว่า Threshold จากนั้นปรับค่า Threshold ลดลงครั้งละ 0.05 เพื่อคัดเลือกเฉพาะคุณลักษณะที่สำคัญที่สุด หลังจากการคัดเลือกคุณลักษณะจำนวนตัวแปรลดลงจาก 10 ตัว เหลือเพียง 5 ตัว ซึ่งทำให้ระบบมีความเร็วในการคำนวณเพิ่มขึ้นและยังคงความแม่นยำในการจำแนกกลิ่นได้ถึง 92% ใกล้เคียงกับกรณีที่ใช้ตัวแปรทั้งหมด

Terko et al. (2019) ได้พัฒนาโมเดลการพยากรณ์การผิดนัดชำระหนี้โดยใช้ข้อมูลจากสถาบันไมโครไฟแนนซ์ในประเทศบอสเนียและเฮอร์เซโกวีนา ซึ่งประกอบไปด้วย 75,386 รายการและตัวแปรทั้งหมด 49 ตัว ผู้วิจัยได้ใช้ XGBoost ในการคำนวณค่า Feature Importance ของตัวแปรทั้งหมด จากนั้นเลือกตัวแปรที่มีคะแนนสะสม ≥ 0.99 เหลือเพียง 22 ตัวแปร พบว่าการลดจำนวนตัวแปรช่วยลดระยะเวลาการประมวลผลของโมเดลได้อย่างมีประสิทธิภาพ และยังคงความแม่นยำของโมเดลไว้ที่ 81.8% เมื่อเปรียบเทียบกับกรณีที่ใช้ตัวแปรทั้งหมด

Junhui Xu et al. (2021) ได้เล็งเห็นว่ายังมีการศึกษาค่อนข้างน้อยในการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลของประเทศจีน ซึ่งการพยากรณ์แบบเดิมในประเทศจีนนั้นมักนิยมใช้วิธีการการสถิติแบบเดิม งานวิจัยนี้จึงได้นำการเรียนรู้ของเครื่องมาประยุกต์ใช้ในการสร้างแบบจำลองในการพยากรณ์การผิดนัดชำระหนี้สำหรับข้อมูลของแพลตฟอร์ม P2P ประเทศจีน ข้อมูลที่ใช้ในงานวิจัยนี้มาจาก Renrendai.com ซึ่งเป็น 1 ใน 10 แพลตฟอร์มการให้กู้ยืมแบบ P2P ที่ใหญ่ที่สุดในประเทศจีน ข้อมูลครอบคลุมตั้งแต่วันพฤหัสบดีที่ 1 มกราคม พ.ศ. 2558 ถึงวันอังคารที่ 30 มิถุนายน พ.ศ. 2558 เนื่องจากข้อมูลดังกล่าวมีตัวแปรตามสำหรับการพยากรณ์ที่ไม่สมดุล ผู้วิจัยจึงนำวิธีการ SMOTE มาแปลงข้อมูลให้เป็นชุดข้อมูลที่สมดุล จากนั้นผู้วิจัยทำการสร้างแบบจำลองสำหรับพยากรณ์การผิดนัดชำระหนี้และเปรียบเทียบอัลกอริทึมที่ใช้สร้างแบบจำลองทั้ง 5 อัลกอริทึม ได้แก่ Random Forest (RF), Extreme Gradient Boosting Tree (XGBT), Gradient Boosting (GBM), Neural Network (NN) และ Generalized Linear model (GLM) จากผลการวิจัยพบว่า Random Forest สามารถให้ผลการพยากรณ์ได้ดีที่สุดโดยประเมินจากค่า Accuracy ที่มีค่าสูงถึง 98.4% อย่างไรก็ตามแบบจำลองจากอัลกอริทึมการเรียนรู้ของเครื่องอื่น ๆ นั้นล้วนมีความแม่นยำในการทำนายสูงเกิน 90% ทั้งหมด และสูงกว่าแบบจำลอง GLM แบบดั้งเดิมทั้งหมด

Chang et al. (2022) ได้ทำการศึกษาเชิงเปรียบเทียบเกี่ยวกับอัลกอริทึมที่ใช้ในการพยากรณ์การผิดนัดชำระหนี้ เพื่อหาอัลกอริทึมที่เหมาะสม งานวิจัยนี้ได้นำวิธีการ Under-Sampling มาใช้เพื่อแก้ปัญหาความไม่สมดุลของข้อมูล ข้อมูลที่ใช้เป็นข้อมูลสินเชื่อจาก Lending Club ตั้งแต่เดือนมกราคม พ.ศ. 2556 ถึง ธันวาคม พ.ศ. 2557 จำนวน 282,763 รายการ 151 คุณลักษณะ เพื่อลดจำนวนของคุณลักษณะที่มีจำนวนมาก ผู้วิจัยได้ทำการคัดเลือกคุณสมบัติโดยใช้วิธีการ XGBoost ในการหาค่าความสำคัญของแต่ละคุณสมบัติ จากนั้นทำการคัดเลือกากคุณสมบัติที่มีค่าความสำคัญอยู่ในอันดับต้น ๆ และทำการหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยคุณสมบัติคู่ใดที่มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำกว่า 0.7 จะถูกตัดทิ้ง การลดจำนวนคุณสมบัติจะทำให้การทำงานของคอมพิวเตอร์ที่ใช้คำนวณอัลกอริทึมนั้นรวดเร็วขึ้น โดยอัลกอริทึมที่ใช้ในงานวิจัยนี้ประกอบไปด้วย Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, XGBoost, LightGBM และ Neural Network จากนั้นทำการประเมินประสิทธิภาพในการพยากรณ์ของแต่ละแบบจำลองด้วย Accuracy, Recall, Precision, F1 score และ ระยะเวลาในการคำนวณ (Runtime) จากผลการศึกษาพบว่าอัลกอริทึม XGBoost และ LightGBM มีประสิทธิภาพดีที่สุด

Nemade (2022) ได้พบว่าปัญหาที่สำคัญในการศึกษาการพยากรณ์การผิดนัดชำระหนี้นั้นมาจากชุดข้อมูลที่มีความไม่สมดุลของตัวแปรตาม ผู้วิจัยจึงได้นำวิธีการ Undersampling, Oversampling และ Cost-Sensitive Learning มาประยุกต์ใช้เพื่อแก้ปัญหาความไม่สมดุลของข้อมูล และทำการเปรียบเทียบผลลัพธ์จากวิธีการทั้ง 3 เพื่อตรวจสอบวิธีการที่เหมาะสมในการแก้ปัญหาความไม่สมดุลของข้อมูล ซึ่งข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลจาก Lending Club ซึ่งเป็นข้อมูลประวัติเงินกู้ 887,379 รายการ ที่รวบรวมไว้ตั้งแต่ปี พ.ศ. 2550 ถึง 2558 ผู้วิจัยได้ทำการแบ่งชุดข้อมูลเป็น 70% สำหรับชุดการเรียนรู้ และ 30% สำหรับชุดข้อมูลการทดสอบ อัลกอริทึมที่ใช้ในการสร้างแบบจำลองประกอบด้วย Logistic Regression, Random Forest และ Neural Networks จากนั้นทำการประเมินประสิทธิภาพในการพยากรณ์ด้วย Accuracy ผลการศึกษาพบว่าการแก้ปัญหาความไม่สมดุลของข้อมูลโดยใช้วิธีการ Undersampling ทำให้สามารถสร้างแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ได้ดีที่สุด

Wu (2022) ศึกษาการสร้างแบบจำลองในการพยากรณ์การผิดนัดชำระหนี้เพื่อใช้เป็นเครื่องมือในการประเมินผู้สมัครขอรับสินเชื่อ ชุดข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลที่จัดทำโดย Imperial College London มีจำนวนข้อมูล 105,471 รายการกับ 771 คอลัมน์ ด้วยจำนวนคอลัมน์ของข้อมูลที่เยอะ ผู้วิจัยจึงได้ใช้วิธีการคัดเลือกคุณลักษณะ (Feature Selection) เพื่อลดจำนวนข้อมูลที่ไม่จำเป็นก่อนนำไปสร้างแบบจำลอง ขั้นแรกจะใช้วิธีกำหนดเกณฑ์ค่าความแปรปรวน (Variance) ของแต่ละคอลัมน์ โดยคอลัมน์ที่มีค่าความแปรปรวนเป็น 0 ก็จะถูกตัดทิ้ง เนื่องจากคอลัมน์ดังกล่าวไม่มีข้อมูลหรือข้อมูลเป็นค่าเดียวกันหมด ขั้นที่สองทำการวิเคราะห์ Variance

Inflation Factor (VIF) ซึ่งคุณลักษณะใดที่มีค่า VIF จะถูกตัดทิ้ง ผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็น 80% สำหรับชุดข้อมูลการเรียนรู้ (Train Set) 20% สำหรับชุดข้อมูลการทดสอบ (Test Set) หลังจากนั้นทำการสร้างแบบจำลองโดยใช้อัลกอริทึม Random Forest และ XGBoost จากนั้นทำการประเมินประสิทธิภาพของแบบจำลองทั้งสองด้วย Accuracy ผลการศึกษาพบว่าแบบจำลองจากอัลกอริทึม Random Forest และ XGBoost มีประสิทธิภาพที่ใกล้เคียงกันที่ค่า Accuracy เป็น 0.90657 และ 0.90635 ตามลำดับ

Sharma et al. (2023) ได้ศึกษาการสร้างแบบจำลองสำหรับพยากรณ์ความเสี่ยงในการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลเงินกู้จาก Lending Club ซึ่งเป็นแพลตฟอร์มผู้ให้บริการสินเชื่อออนไลน์ระหว่างบุคคล ข้อมูลครอบครุมตั้งแต่ช่วง มกราคม พ.ศ. 2561 ถึง ธันวาคม พ.ศ. 2562 มีจำนวนข้อมูลทั้งสิ้น 272,401 รายการ ประกอบไปด้วยคุณลักษณะจำนวน 142 ตัว หลังจากนั้นจึงทำการทำความสะอาดข้อมูลและทำการคัดเลือกคุณลักษณะที่สำคัญของข้อมูลเพื่อนำไปใช้ในการสร้างแบบจำลองโดยอิงจากงานวิจัยอื่น ๆ จะได้คุณลักษณะที่สำคัญ 26 ตัว คุณลักษณะ loan status คือ สถานการณ์ชำระเงินคืนของบัญชีนั้น ๆ ซึ่งเป็นตัวแปรตามในการพยากรณ์ จากการตรวจสอบพบว่า 90% ของข้อมูลทั้งหมดมีตัวแปรตามเป็น “สถานะปกติ (Current)” และ “ชำระครบเต็มจำนวน (Fully Paid)” ทำให้คุณลักษณะเป้าหมายในการพยากรณ์นั้นไม่สมดุล เพื่อแก้ไขปัญหาดังกล่าวจึงทำการสุ่มเลือกข้อมูลมา 20,000 รายการ โดยข้อมูลกลุ่มตัวอย่างนี้จะประกอบด้วยข้อมูลจำนวน 10,000 รายการที่มีสถานะเป็น “ชำระครบเต็มจำนวน (Fully Paid)” และอีก 10,000 รายการมีสถานะเป็น “บัญชีถูกตัด (Charged Off)” จากนั้นแบ่งข้อมูลออกเป็น 2 ชุด คือ 80% สำหรับข้อมูลชุดฝึก (Train data) และ 20% สำหรับข้อมูลชุดทดสอบ (Test data) สำหรับสร้างแบบจำลองสำหรับพยากรณ์การผิดนัดชำระหนี้ซึ่งในงานวิจัยนี้ได้ทำการเปรียบเทียบประสิทธิภาพในการพยากรณ์ของ 2 อัลกอริทึม ได้แก่ Random Forest (RF) และ XGBoost โดยประเมินประสิทธิภาพในการพยากรณ์ของแต่ละแบบจำลองด้วย AUC, Accuracy, Precision, Recall และ F1-score ผลการวิจัยพบว่า Random Forest มีประสิทธิภาพการพยากรณ์โดยประเมินจาก AUC อยู่ที่ 0.99 สูงกว่า XGBoost ที่มีค่า AUC อยู่ที่ 0.9

พหุน ปณฺ ฑิโต ชีเว

บทที่ 3
วิธีดำเนินการวิจัย

ในส่วนนี้ผู้วิจัยได้นำเสนอแผนการดำเนินงานวิจัย ชุดข้อมูลที่ใช้ในการวิจัย กระบวนการดำเนินงานวิจัย และกระบวนการการสำรวจข้อมูล ซึ่งมีดังนี้

3.1 ชุดข้อมูล (Data set)

3.1.1 ข้อมูลสินเชื่อ Bondora

ชุดข้อมูลนี้ได้รับมาจากชุดข้อมูลสาธารณะของ Bondora ซึ่งเป็นแพลตฟอร์มการให้กู้ยืมแบบ P2P ชั้นนำของยุโรป จากเว็บไซต์ https://www.bondora.com/en/public-reports เป็นข้อมูลสินเชื่อของกลุ่มลูกค้าที่มีการผิดนัดชำระหนี้และไม่ผิดนัดชำระหนี้จากช่วงเวลาระหว่างเดือนกุมภาพันธ์ พ.ศ. 2552 ถึงมกราคม พ.ศ. 2567 ข้อมูลประกอบด้วยข้อมูลทางประชากรศาสตร์และการเงินของผู้กู้และธุรกรรมการกู้ยืม มีจำนวนข้อมูลทั้งหมด 358,092 รายการ 112 ตัวแปร

3.1.2 ข้อมูลเศรษฐศาสตร์มหภาค (Macroeconomic Dataset)

ชุดข้อมูลด้านเศรษฐศาสตร์มหภาคที่ใช้ในงานวิจัยนี้เป็นชุดข้อมูลที่รวบรวมมาจากเว็บไซต์ https://www.imf.org/external/datamapper/datasets/WEO ซึ่งเป็นข้อมูล World Economic Outlook (WEO) ที่รวบรวมโดยเจ้าหน้าที่ของ IMF เป็นข้อมูลในอดีตและข้อมูลที่ประมาณการซึ่งวิเคราะห์ผ่านสถานการณ์ที่เปลี่ยนแปลงไปในแต่ละประเทศ ประกอบด้วยตัวบ่งชี้ทางเศรษฐศาสตร์มหภาค 15 ตัว ของ 237 ประเทศทั่วโลก ครอบคลุมตั้งแต่ปี พ.ศ. 2523 ถึง พ.ศ. 2570 แสดงในตารางที่ 1

ตารางที่ 1 ตัวแปรด้านเศรษฐศาสตร์มหภาค

ตัวแปร	คำอธิบายข้อมูล
GDP_G	อัตราการเติบโตทางเศรษฐกิจ ณ มูลค่าตามราคาคงที่
GDP_CP	ผลิตภัณฑ์มวลรวมในประเทศ ณ ราคาปัจจุบัน
GDP_PC	ผลิตภัณฑ์มวลรวมในประเทศต่อหัว
GDP_PW	ผลิตภัณฑ์มวลรวมในประเทศ ณ ราคาปัจจุบันเทียบกับความเท่าเทียมกันของกำลังซื้อ
GDP_ID	ผลิตภัณฑ์มวลรวมในประเทศต่อหัวเทียบกับความเท่าเทียมกันของกำลังซื้อดอลลาร์ระหว่างประเทศ
GDP_PPP	ผลิตภัณฑ์มวลรวมภายในประเทศตามความเท่าเทียมกันของกำลังซื้อ
I_PPP	อัตราการแปลงสกุลเงินตามความเท่าเทียมกันของกำลังซื้อ
IR_AC	อัตราเงินเฟ้อเทียบกับดัชนีราคาผู้บริโภคโดยเฉลี่ย
IR_EC	อัตราเงินเฟ้อเทียบกับมูลค่าสิ้นงวดของดัชนีราคาผู้บริโภค

ตัวแปร	คำอธิบายข้อมูล
POP	จำนวนประชากร
UR	อัตราการว่างงาน
AC_US	ดุลบัญชีเดินสะพัดเทียบกับดอลล่าร์สหรัฐ
AC_GDP	ดุลบัญชีเดินสะพัดเทียบกับเปอร์เซ็นต์ของผลิตภัณฑ์มวลรวมในประเทศ
GG_NET	ดุลการให้กู้ยืมสุทธิ/กู้ยืมสุทธิ
GG_GD	หนี้โดยรวมทั่วไปของรัฐบาล

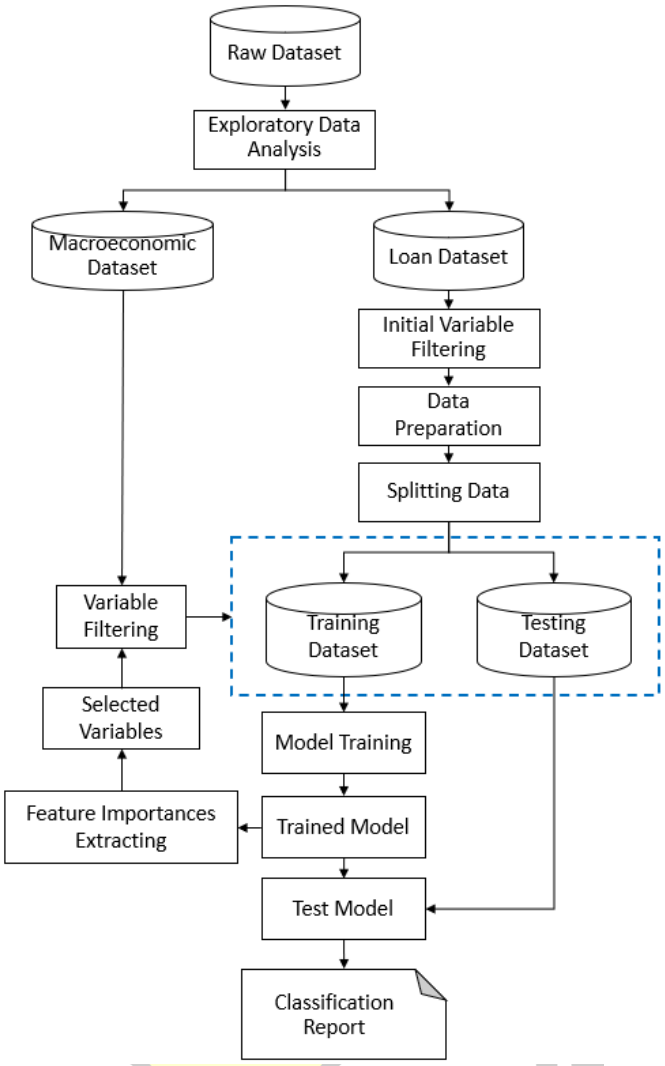
สำหรับการใช้งานข้อมูลเศรษฐศาสตร์มหภาค จะต้องทำการ Mapping (จับคู่ข้อมูล) กับข้อมูลสินเชื่อที่มีอยู่ โดยใช้ วันที่ที่สินเชื่อได้รับการอนุมัติ (LoanDate) เป็นตัวเชื่อมโยง ซึ่งผู้วิจัยได้ใช้ค่า "ปี" ของวันที่ดังกล่าว จากนั้นจึงนำไปจับคู่กับข้อมูลเศรษฐศาสตร์มหภาคที่ตรงกับปี ก่อนหน้า เช่น หากสินเชื่อได้รับการอนุมัติในปี พ.ศ. 2565 ข้อมูลเศรษฐศาสตร์มหภาคที่ใช้จะเป็นของปี พ.ศ. 2564

แนวทางการ Mapping นี้มีจุดมุ่งหมายเพื่อให้ข้อมูลเศรษฐศาสตร์มหภาคสะท้อนถึงสภาพเศรษฐกิจที่อาจส่งผลต่อการพิจารณาสินเชื่อ ทั้งนี้การวิเคราะห์ข้อมูลเศรษฐศาสตร์มหภาคดังกล่าวมีบทบาทสำคัญในการระบุปัจจัยภายนอกที่อาจส่งผลต่อปัญหาการผิดนัดชำระหนี้ในแบบจำลองที่พัฒนาขึ้นในงานวิจัยนี้

3.2 กรอบการดำเนินงาน

สำหรับส่วนต่อไปนี้เป็นแผนผังกรอบการดำเนินงาน ดังรูปที่ 6

พหุน ปณฺ ฑิโต ชีเว

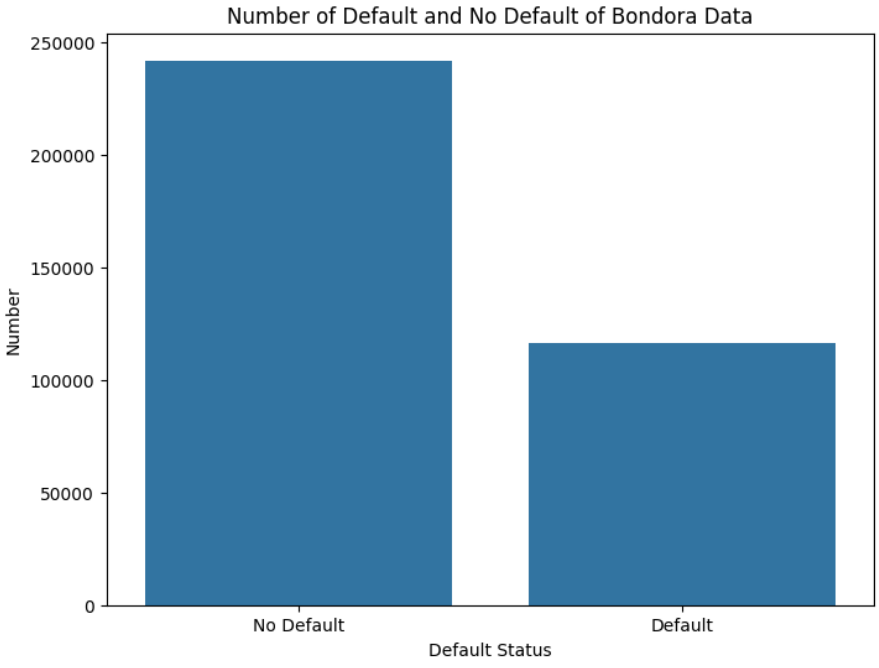


รูปที่ 6 แผนผังกรอบดำเนินงาน

1. ข้อมูลดิบ (Raw dataset) ประกอบด้วยชุดข้อมูลสินเชื่อของ Bondora และข้อมูลด้านเศรษฐศาสตร์มหภาคที่ได้ทำการรวบรวมมา

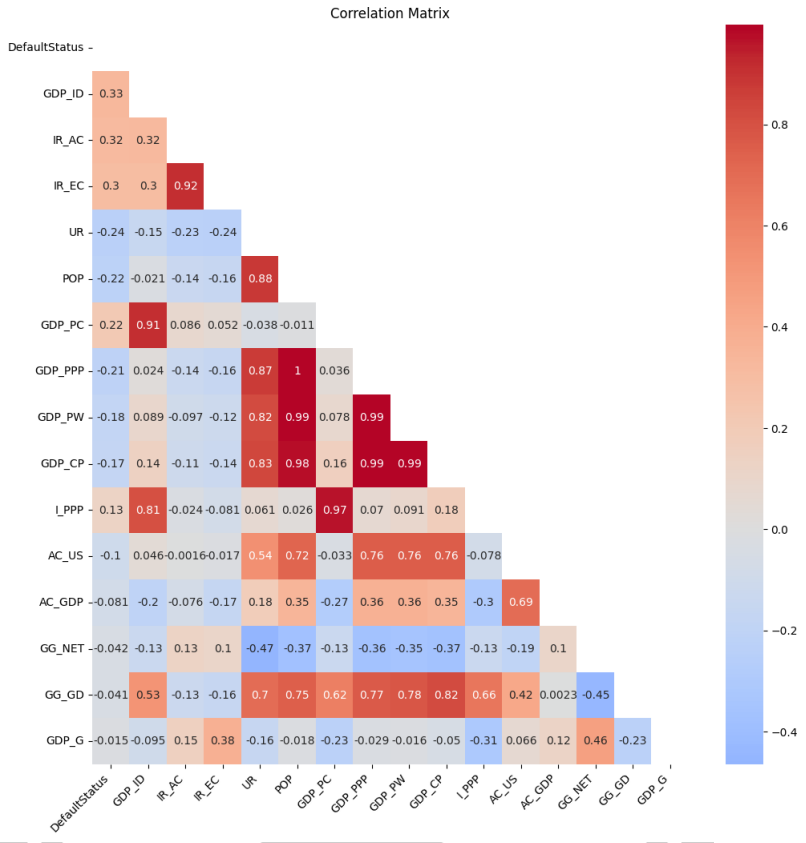
2. กระบวนการการสำรวจข้อมูล (Exploratory Data Analysis) เป็นขั้นตอนสำคัญในการวิเคราะห์ข้อมูล ซึ่งมีบทบาทสำคัญในการทำความเข้าใจลักษณะและโครงสร้างของข้อมูลก่อนที่จะนำไปใช้ในกระบวนการสร้างแบบจำลอง ตัวแปรตาม (Dependent Variable) ที่ใช้ในงานวิจัยนี้ ถูกสร้างขึ้นโดยเพิ่มคอลัมน์ใหม่มีชื่อว่า DefaultStatus โดยอ้างอิงจากตัวแปร DefaultDate ซึ่งระบุวันที่ที่ลูกค้ามีการผิดนัดชำระหนี้ โดยกำหนดให้ลูกค้าที่มีประวัติการผิดนัดชำระหนี้เป็น 0 (Default) และลูกค้าที่ไม่มีประวัติการผิดนัดชำระหนี้เป็น 1 (No Default) การตรวจสอบการกระจายตัวข้อมูลของตัวแปรตามแสดงในรูปที่ 7 พบว่าชุดข้อมูลประกอบไปด้วยลูกค้าที่ไม่มีการผิดนัดชำระหนี้จำนวน

241,932 ราย คิดเป็นประมาณ 67.56% และลูกค้าที่มีการผิดนัดชำระหนี้จำนวน 116,160 ราย คิดเป็นประมาณ 32.44% ซึ่งเป็นข้อมูลไม่สมดุล (Imbalanced Data)



รูปที่ 7 จำนวนลูกค้าที่มีการผิดนัดชำระหนี้และไม่ผิดนัดชำระหนี้

ในงานวิจัยนี้ ผู้วิจัยได้นำตัวแปรด้านเศรษฐศาสตร์มหภาคกับตัวแปรตามมาวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้วยวิธีการ Point Biserial Correlation เนื่องจากประเภทของตัวแปรเศรษฐศาสตร์มหภาคเป็นตัวแปรต่อเนื่อง (Continuous) และประเภทของตัวแปรตามเป็นตัวแปรที่มี 2 ลักษณะจริง (True Dichotomous) ได้แก่ 0 คือ ลูกค้าที่มีประวัติการผิดนัดชำระหนี้เป็น และ 1 คือ ลูกค้าที่ไม่มีประวัติการผิดนัดชำระหนี้ สำหรับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรด้านเศรษฐศาสตร์จะใช้วิธีการ Pearson Correlation ผลลัพธ์ของการวิเคราะห์ความสัมพันธ์แสดงในรูปที่ 8 หลังจากนั้น ผู้วิจัยได้ทำการคัดเลือกเฉพาะตัวแปรที่มีค่าความสัมพันธ์แบบสัมบูรณ์ (|Correlation|) สูงกว่าเกณฑ์ที่กำหนด (Threshold) ซึ่งในกรณีนี้กำหนดให้ Threshold = 0.2 เนื่องจากพบว่าความสัมพันธ์ระหว่างตัวแปรทางเศรษฐศาสตร์มหภาคกับตัวแปรตามมีระดับต่ำ โดยแนวทางนี้เป็นแนวทางเดียวกับงานวิจัยของ Sabilla et al. (2019) ดังนั้น ตัวแปรที่ถูกคัดเลือกจะต้องมีค่า |Correlation| ≥ 0.2 กับตัวแปรตาม ซึ่งหมายถึงตัวแปรที่มีความสัมพันธ์ทางสถิติต่อตัวแปรตามเพียงพอ เพื่อใช้ในการฝึกอบรมแบบจำลองสำหรับงานวิจัยนี้



รูปที่ 8 แสดงการวิเคราะห์ความสัมพันธ์ของตัวแปรเศรษฐศาสตร์มหภาคและตัวแปรตาม

3. การคัดกรองตัวแปรเบื้องต้น (Initial Variable Filtering) เนื่องจากงานวิจัยนี้มุ่งเน้นศึกษาที่ตัวแปรที่เกี่ยวกับข้อมูลเบื้องต้นของผู้ขอสินเชื่อ ดังนั้นข้อมูลที่เกี่ยวข้องกับการชำระหนี้จะไม่ถูกนำเข้ามารวมในกระบวนการนี้ หลังจากการคัดกรองตัวแปรเบื้องต้นจะมีตัวแปรที่เกี่ยวข้องทั้งหมด 45 ตัวแปร พจนานุกรมของข้อมูล (Data Dictionary) แสดงในตารางที่ 2

ตารางที่ 2 พจนานุกรมของข้อมูลสินเชื่อ

ตัวแปร	ชนิดของข้อมูล	คำอธิบาย
NewCreditCustomer	Boolean	ผู้ขอสินเชื่อมีประวัติเครดิตใน Bondora ก่อนหน้านี้หรือไม่ 0: มีประวัติเครดิตอย่างน้อย 3 เดือน, 1: ไม่มีประวัติเครดิตมาก่อน
LoanDate	Date	วันที่ที่สินเชื่อได้รับการอนุมัติ

ตัวแปร	ชนิดของข้อมูล	คำอธิบาย
VerificationType	Integer	วิธีที่ใช้ในการตรวจสอบข้อมูลคำขอสินเชื่อ 0: ไม่ระบุ, 1: รายได้ไม่ได้รับการยืนยัน, 2: รายได้ไม่ได้รับการยืนยันแต่มีการตรวจสอบผ่านการสอบถามทางโทรศัพท์, 3: รายได้ได้รับการยืนยัน, 4: รายได้และค่าใช้จ่ายได้รับการตรวจสอบ
LanguageCode	Integer	ภาษาที่ใช้ในการขอสินเชื่อ 1: เอสโตเนีย, 2: อังกฤษ, 3: รัสเซีย, 4: ฟินแลนด์, 5: เยอรมัน, 6: สเปน, 9: สโลวาเกีย
Age	Integer	อายุของผู้ขอสินเชื่อ
Gender	Integer	เพศของผู้ขอสินเชื่อ 0: ชาย, 1: หญิง, 2: ไม่ระบุ
Country	Text	ประเทศที่อยู่อาศัยของผู้ขอสินเชื่อ
AppliedAmount	Float	จำนวนเงินที่ผู้ขอสินเชื่อระบุในคำขอสินเชื่อ
Amount	Float	จำนวนเงินที่ผู้ขอสินเชื่อได้รับ
Interest	Float	อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ
LoanDuration	Integer	ระยะเวลาของสินเชื่อหน่วยเป็นเดือน
MonthlyPayment	Integer	จำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือน
UseOfLoan	Integer	วัตถุประสงค์ของการขอสินเชื่อ 0: การรวมสินเชื่อ, 1: อสังหาริมทรัพย์, 2: ปรับปรุงบ้าน, 3: ธุรกิจ, 4: การศึกษา, 5: การท่องเที่ยว, 6: ยานพาหนะ, 7: อื่น ๆ , 8: สุขภาพ, 101: การจัดหาเงินทุนหมุนเวียน, 102: การซื้อเครื่องจักรและอุปกรณ์, 103: การปรับปรุงอสังหาริมทรัพย์, 104: การจัดหาเงินทุนลูกหนี้, 105: การซื้อยานพาหนะ, 106: การเงินก่อสร้าง, 107: การซื้อหุ้น, 108: การซื้ออสังหาริมทรัพย์, 109: การค้ำประกันภาระผูกพัน, 110: ธุรกิจอื่นๆโค้ดรูปแบบ
Education	Integer	ระดับการศึกษาของผู้ขอสินเชื่อ 1: การศึกษาระดับประถมศึกษา, 2: การศึกษาขั้นพื้นฐานทั่วไป, 3: การศึกษาวิชาชีพ, 4: การศึกษาระดับมัธยม, 5: การศึกษาระดับอุดมศึกษา
MaritalStatus	Integer	สถานภาพของผู้ขอสินเชื่อ 1: สมรส, 2: อยู่ร่วมกัน, 3: โสด, 4: หย่าร้าง, 5: หม้าย
NrOfDependants	Text	จำนวนบุตรหรือผู้ที่อยู่ในความอุปการะ
EmploymentStatus	Integer	สถานะการทำงาน 1: ว่างงาน, 2: งานพาร์ทไทม์, 3: งานประจำ, 4: งานอิสระ, 5: ผู้ประกอบการ, 6: เกษียณอายุ
EmploymentDurationCurrentEmployer	Text	ระยะเวลาการในการทำงานของอาชีพปัจจุบัน

ตัวแปร	ชนิดของข้อมูล	คำอธิบาย
WorkExperience	Text	ประสบการณ์ทำงานหน่วยเป็นปี
OccupationArea	Integer	ประเภทของอาชีพ 1: อื่นๆ, 2: เหมืองแร่, 3: การแปรรูป, 4: พลังงาน, 5: สาธารณูปโภค, 6: ก่อสร้าง, 7: ค้าปลีกและค้าส่ง, 8: ขนส่งและคลังสินค้า, 9: การบริการและการจัดเลี้ยง, 10: ข้อมูลและโทรคมนาคม, 11: การเงินและการประกันภัย, 12: อสังหาริมทรัพย์, 13: การวิจัย, 14: การบริหาร, 15: ราชการและทหาร, 16: การศึกษา, 17: การดูแลสุขภาพและช่วยเหลือทางสังคม, 18: ศิลปะและความบันเทิง, 19: การเกษตรป่าไม้และการประมง
HomeOwnershipType	Integer	ประเภทของที่อยู่อาศัย 0: ไม่มีที่อยู่อาศัย, 1: เจ้าของ, 2: อยู่กับพ่อแม่, 3: ผู้เช่าทรัพย์สินที่มีเฟอร์นิเจอร์, 4: ผู้เช่าทรัพย์สินที่ไม่มีเฟอร์นิเจอร์, 5: บ้านเช่าจากเทศบาล, 6: ผู้เช่าร่วม, 7: กรรมสิทธิ์ร่วม, 8: จำนอง, 9: เจ้าของที่มีภาระผูกพัน, 10: อื่นๆ
IncomeFromPrincipalEmployer	Float	รายได้จากการทำงาน
IncomeFromPension	Float	รายได้จากเงินบำนาญ
IncomeFromFamilyAllowance	Float	รายได้จากเงินช่วยเหลือครอบครัว
IncomeFromSocialWelfare	Float	รายได้จากการสนับสนุนทางสังคม
IncomeFromLeavePay	Float	รายได้จากการลาคลอด
IncomeFromChildSupport	Float	รายได้จากการจ่ายค่าเลี้ยงดูบุตร
IncomeOther	Float	รายได้จากแหล่งอื่น
IncomeTotal	Float	รายได้รวม
ExistingLiabilities	Integer	จำนวนหนี้สินที่มีอยู่
LiabilitiesTotal	Float	หนี้สินรวมต่อเดือน
RefinanceLiabilities	Integer	จำนวนหนี้สินรวมหลังการรีไฟแนนซ์
DebtToIncome	Float	อัตราส่วนของรายได้รวมต่อเดือนที่ใช้ในการชำระหนี้
FreeCash	Float	รายได้ที่สามารถใช้จ่ายได้หลังหักหนี้สินรายเดือน
MonthlyPaymentDay	Integer	วันของเดือนที่เป็นกำหนดการชำระเงินสินเชื่อ
DefaultDate	Date	วันที่สินเชื่อเข้าสู่สถานะผิดนัดชำระหนี้และเริ่มกระบวนการเรียกเก็บเงิน

ตัวแปร	ชนิดของข้อมูล	คำอธิบาย
CreditScoreEsMicroL	Text	คะแนนที่ออกแบบมาโดยเฉพาะสำหรับการจัดความเสี่ยงผู้ขอสินเชื่อ subprime (ที่กำหนดโดย Equifax ว่าเป็นผู้กู้ที่ไม่สามารถเข้าถึงสินเชื่อของธนาคาร) แสดงความน่าจะเป็นของการผิดนัดชำระล่วงหน้า 1 เดือน ลำดับคะแนนจาก M1 - M10 เรียงจากคะแนนที่น้อยที่สุดไปมากที่สุด
CreditScoreEsEquifaxRisk	Text	คะแนนความเสี่ยงของ Equifax สำหรับผู้ขอสินเชื่อเป็นคะแนนทั่วไปสำหรับผู้ขอสินเชื่อที่ไม่มีการดำเนินการผิดนัดชำระอยู่ในฐานข้อมูล ASNEF แสดงถึงความน่าจะเป็นของการผิดนัดชำระหนี้ในช่วงเวลา 1 ปีข้างหน้า คะแนนจะถูกให้ในระดับ 6 ระดับ ได้แก่ AAA (ต่ำมาก), AA (ต่ำ), A (ปานกลาง), B (ปานกลางสูง), C (สูง), และ D (สูงมาก)
CreditScoreFiAsiakasTietoRiskGrade	Text	โมเดลการให้คะแนนเครดิต FinnishAsiakastietoRL1 ความเสี่ยงต่ำมาก: 01-20RL2, ความเสี่ยงต่ำ: 21-40RL3, ความเสี่ยงปานกลาง 41-60RL4, ความเสี่ยงสูง: 61-80RL5, ความเสี่ยงสูงมาก: 81-100
CreditScoreEeMini	Integer	ข้อมูลคะแนนเครดิต 1000: ไม่มีปัญหาการชำระเงินก่อนหน้านี้, 900: ปัญหาการชำระเงินสิ้นสุดลง 24-36 เดือนที่แล้ว, 800: ปัญหาการชำระเงินสิ้นสุดลง 12-23 เดือนที่แล้ว, 700: ปัญหาการชำระเงินสิ้นสุดลง 6-11 เดือนที่แล้ว, 600: ปัญหาการชำระเงินสิ้นสุดลง <6 เดือนที่แล้ว, 500: อยู่ระหว่างประสบในปัญหาการชำระเงิน
NoOfPreviousLoansBeforeLoan	Integer	จำนวนสินเชื่อก่อนหน้านี้
AmountOfPreviousLoansBeforeLoan	Float	มูลค่าสินเชื่อก่อนหน้านี้
PreviousRepaymentsBeforeLoan	Float	จำนวนเงินที่ผู้ขอสินเชื่อได้ชำระก่อนหน้านี้
PreviousEarlyRepaymentsBefoleLoan	Float	จำนวนเงินที่ชำระล่วงหน้าก่อนหน้านี้
PreviousEarlyRepaymentsCountBeforeLoan	Float	จำนวนครั้งที่ผู้ขอสินเชื่อได้ชำระคืนก่อนกำหนด

3. การเตรียมข้อมูล (Data Preparation)

- การแปลงข้อมูล (Data Transformation) จากการตรวจสอบพจนานุกรมข้อมูล พบว่ามีชนิดของข้อมูลจำนวนเต็มที่เป็นข้อมูลเชิงหมวดหมู่ (Categorical Data) ที่ระบุกลุ่มหรือประเภทของข้อมูล ผู้วิจัยได้ทำการแปลงข้อมูลดังกล่าวตามคำอธิบายของข้อมูล
- จัดการกับค่าที่ขาดหายไป (Missing Value) พบว่ามีข้อมูลที่ขาดหายไปจำนวนมาก โดยสำหรับกลุ่มข้อมูลเชิงหมวดหมู่ (Categorical Data) ผู้วิจัยได้ทำการแทนที่ค่าที่ขาดหายไปโดยระบุ “ไม่ระบุ” (Undefined) ส่วนกลุ่มข้อมูลเชิงตัวเลข (Numerical Data) จะทำการแทนที่ค่าที่ขาดหายไปด้วยมัธยฐาน (Median) ของตัวแปรเหล่านั้น
- วิศวกรรมคุณลักษณะ (Feature Engineering) จากชุดข้อมูลที่มีข้อมูลเชิงหมวดหมู่ ซึ่งไม่สามารถนำเข้าสู่กระบวนการเรียนรู้ของเครื่องได้โดยตรง เนื่องจากข้อมูลนำเข้า (Input) จะต้องเป็นข้อมูลเชิงตัวเลข ในกระบวนการนี้ ผู้วิจัยได้ทำการเข้ารหัสข้อมูลเชิงหมวดหมู่โดยใช้วิธี Label Binarizer เพื่อสร้างตัวแปรใหม่ให้อยู่ในรูปแบบข้อมูลเชิงตัวเลข โดย Label Binarizer จะสร้างตัวแปรใหม่ที่มีจำนวนเท่ากับหมวดหมู่ของตัวแปรเหล่านั้น ซึ่งจะมีค่าเป็น 0 เพื่อระบุว่าข้อมูลดังกล่าวไม่จัดอยู่ในหมวดหมู่ดังกล่าว และ 1 เพื่อระบุว่าข้อมูลดังกล่าวจัดอยู่ในหมวดหมู่ดังกล่าว ตัวอย่างดังรูปที่ 9

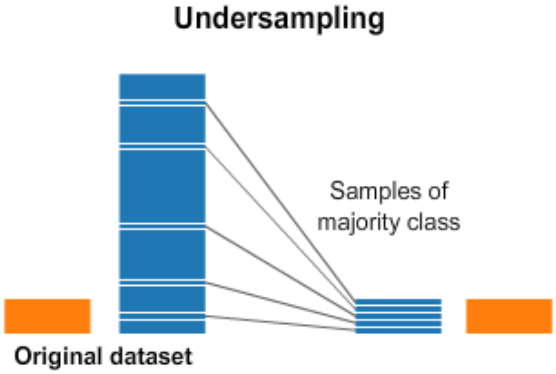
DefaultStatus	Gender
0	Male
1	Female
0	Undefined

Label Binarizer →

DefaultStatus	Gender_Male	Gender_Female	Gender_Undefined
0	1	0	0
1	0	1	0
0	0	0	1

รูปที่ 9 ตัวอย่างการเข้ารหัสข้อมูลด้วยวิธี Label Binarizer

- การจัดการข้อมูลไม่สมดุล (Imbalanced data) จากการสร้างตัวแปรตาม พบว่าชุดข้อมูลมีความไม่สมดุล เมื่อนำข้อมูลเข้าสู่กระบวนการเรียนรู้ของเครื่องจะทำให้ประสิทธิภาพในการพยากรณ์ของแบบจำลองมีการเอนเอียง (Bias) ผู้วิจัยได้ทำการจัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) เป็นการลดจำนวนข้อมูลของกลุ่มส่วนมาก (Majority Class) ให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลของกลุ่มส่วนน้อย (Minority Class) ดังที่แสดงในรูปที่ 10 จากชุดข้อมูลของงานวิจัยนี้ กลุ่มลูกค้าที่ไม่มีการผิดนัดชำระหนี้จำนวน 241,932 ราย จัดเป็นกลุ่มส่วนมาก และลูกค้าที่มีการผิดนัดชำระหนี้จำนวน 116,160 ราย จัดเป็นกลุ่มส่วนน้อย ความแตกต่างในแง่ของประสิทธิภาพในการพยากรณ์ของแบบจำลองระหว่างข้อมูลที่สมดุลและข้อมูลไม่สมดุล ผู้วิจัยจะแสดงในบทที่ 4 ต่อไป



รูปที่ 10 วิธีสุ่มลด (Undersampling)

ที่มา: Sampaio (2023)

- หลังจากกระบวนการเตรียมข้อมูลจะได้ว่า ข้อมูลสินเชื่อที่จะใช้สำหรับกระบวนการเรียนรู้ของเครื่องจะมีจำนวนข้อมูลทั้งหมด 232,302 รายการ 183 ตัวแปร ตัวแปรตามประกอบไปด้วย ลูกค้าที่ไม่มีการผิดนัดชำระหนี้จำนวน 116,160 ราย และลูกค้าที่มีการผิดนัดชำระหนี้จำนวน 116,160 ราย

4. แบ่งข้อมูล (Spitting Data) ในงานวิจัยนี้จะทำการแบ่งชุดข้อมูลที่ผ่านกระบวณการเตรียมข้อมูลออกเป็นชุดข้อมูลการเรียนรู้ (Training set) และชุดข้อมูลการทดสอบ (Testing set) ด้วยอัตราส่วน 80:20

5. การฝึกแบบจำลอง (Model Training) ทำการฝึกแบบจำลองด้วยชุดข้อมูลการเรียนรู้โดยใช้อัลกอริทึม LightGBM, Random Forest และ XGBoost

6. การทดสอบแบบจำลอง (Test Model) แบบจำลองที่ผ่านการฝึกอบรม (Trained Model) จะได้รับการทดสอบประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลการทดสอบ (Testing Set) โดยใช้ Accuracy, Precision, Recall และ F1-score และบันทึกผลการทดสอบแบบจำลองในรูปแบบของ Classification Report เพื่อใช้ในการเปรียบเทียบประสิทธิภาพของแบบจำลองต่อไป

7. การคัดเลือกคุณลักษณะ (Feature Selection) เนื่องจากงานวิจัยนี้มุ่งเน้นที่จะศึกษาปัจจัยที่ส่งผลต่อการผิดนัดชำระหนี้ หลังจากที่ได้รับแบบจำลองที่ผ่านการฝึกอบรมโดยใช้ตัวแปรทั้งหมดแล้ว ผู้วิจัยจะทำการหาความสำคัญของคุณลักษณะ (Feature Importance) จากแบบจำลองและเรียงลำดับคะแนนความสำคัญของตัวแปรตามแบบจำลองที่ต่างกัน จากนั้นทำการคำนวณผลรวมสะสม (Cumulative Sum) ของคะแนนความสำคัญ และคัดเลือกเฉพาะตัวแปรกลุ่มที่ผลรวมของคะแนนความสำคัญสูงกว่าเกณฑ์ที่กำหนด โดยเป็นแนวทางเดียวกับงานวิจัยของ Terko et al. (2019) ซึ่งในงานวิจัยนี้กำหนดเกณฑ์ที่ผลรวมสะสม ≥ 0.8 เพื่อคัดเลือกเฉพาะตัวแปรที่สำคัญที่สุดและรวมตัวแปรด้านเศรษฐศาสตร์มหภาคเข้าด้วยกัน จากนั้นทำซ้ำขั้นตอนการฝึกแบบจำลอง และการทดสอบแบบจำลองอีกครั้ง

บทที่ 4
ผลการวิจัย

ในส่วนนี้ผู้วิจัยได้นำเสนอผลการวิจัยในการพัฒนาแบบจำลองการเรียนรู้ของเครื่องเพื่อพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล โดยใช้อัลกอริทึม LightGBM, Random Forest และ XGBoost ซึ่งจะแบ่งการนำเสนอออกเป็น 3 ส่วนดังนี้

4.1 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการหาความสำคัญของคุณลักษณะ (Feature Importance)

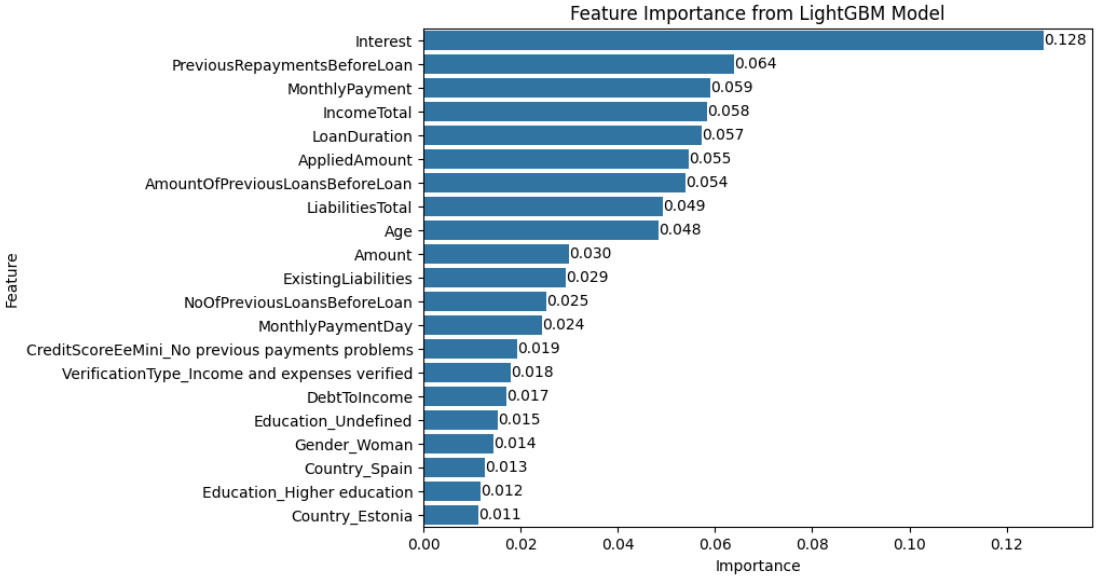
4.2 ปัจจัยทางเศรษฐศาสตร์มหภาคที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการการวิเคราะห์ความสัมพันธ์ (Correlation) ของตัวแปรด้านเศรษฐศาสตร์มหภาคและตัวแปรตาม

4.3 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยชุดของตัวแปรที่แตกต่างกัน 4 รูปแบบ ดังนี้

1. การฝึกอบรมแบบจำลองด้วยตัวแปรทั้งหมด
2. การฝึกอบรมแบบจำลองด้วยตัวแปรทั้งหมด และจัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling)
3. การฝึกอบรมแบบจำลองด้วยชุดข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) และคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection)
4. การฝึกอบรมแบบจำลองด้วยชุดข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) ทำการคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) และรวมตัวแปรด้านเศรษฐศาสตร์มหภาค (Macroeconomic Dataset)

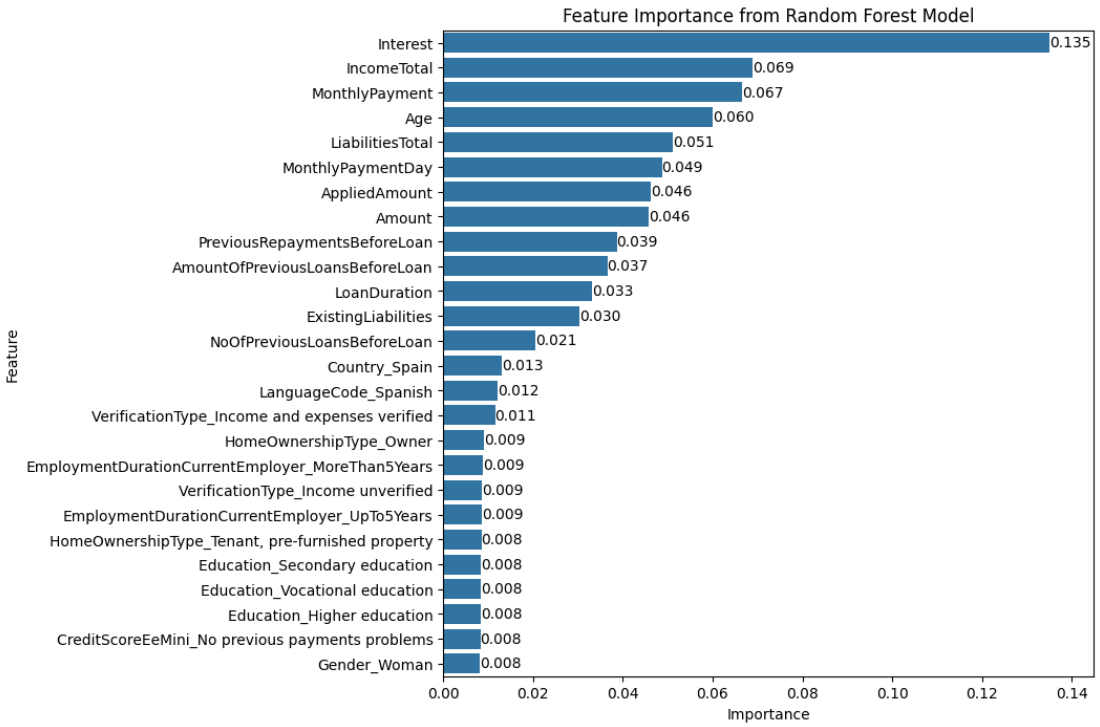
4.1 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการหาความสำคัญของคุณลักษณะ (Feature Importance)

ในการตรวจสอบปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ผู้วิจัยได้นำแบบจำลองที่ได้รับการฝึกอบรมด้วยชุดข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) เพื่อหาความสำคัญของคุณลักษณะ (Feature Importance) จากแบบจำลอง จากนั้นคำนวณผลรวมสะสม (Cumulative Sum) ของคะแนนความสำคัญ โดยมีเกณฑ์ที่ผลรวมสะสม ≥ 0.8 เพื่อคัดเลือกเฉพาะตัวแปรที่มีความสำคัญที่สุดสำหรับแต่ละแบบจำลอง ซึ่งแสดงผลลัพธ์ในรูปที่ 11 - รูปที่ 13



รูปที่ 11 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง LightGBM

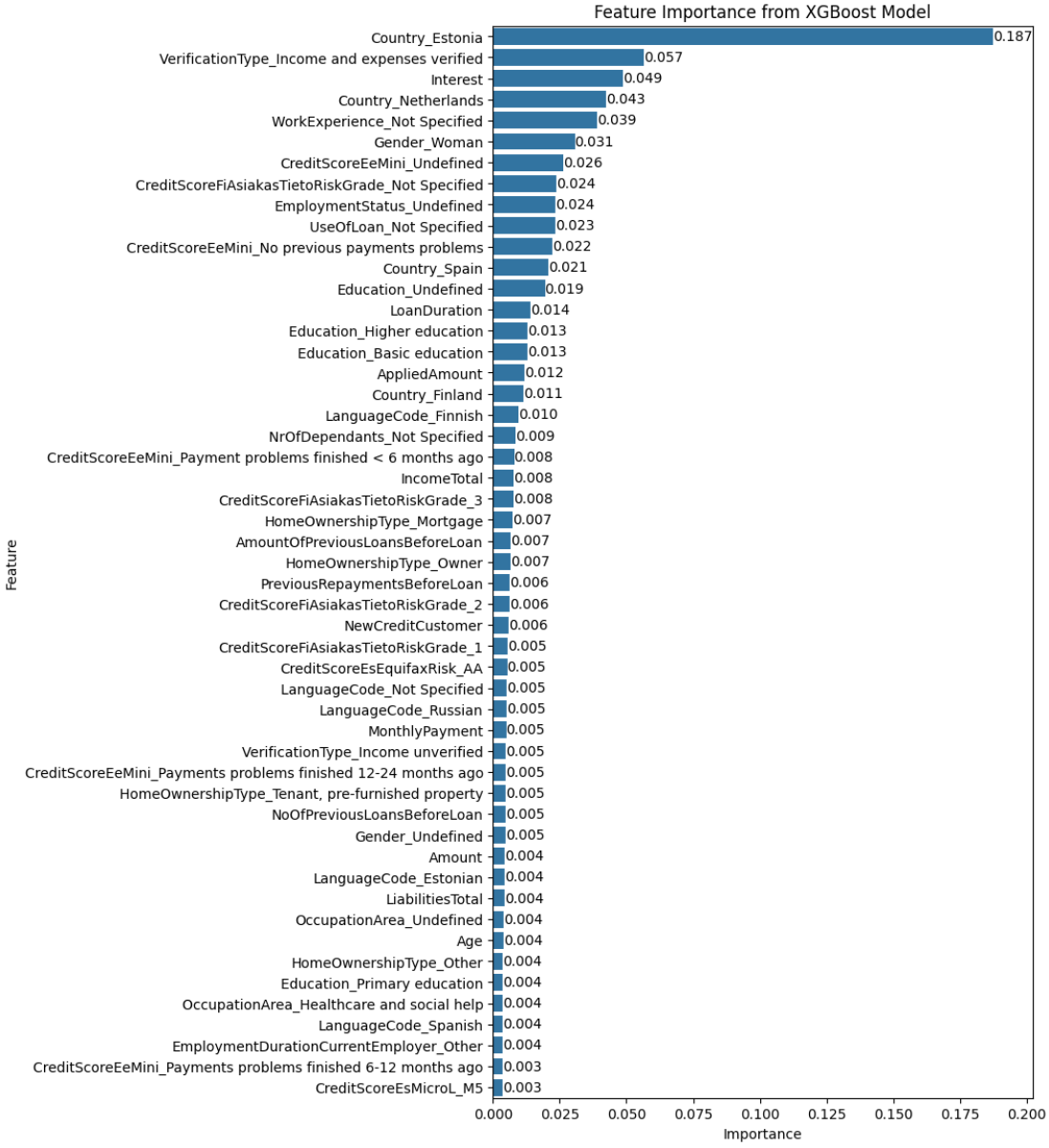
จากรูปที่ 11 พบว่าปัจจัยที่แบบจำลอง LightGBM ให้ความสำคัญสูงที่สุด คือ อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest) นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่แบบจำลองให้ความสำคัญ ได้แก่ จำนวนเงินที่ผู้ขอสินเชื่อได้ชำระก่อนหน้านี้ (PreviousRepaymentsBeforeLoan), จำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือน (MonthlyPayment), รายได้รวม (IncomeTotal) และระยะเวลาของสินเชื่อ (LoanDuration)



รูปที่ 12 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง Random Forest

จากรูปที่ 12 พบว่าปัจจัยที่แบบจำลอง Random Forest ให้ความสำคัญสูงที่สุด คือ อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest) นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่แบบจำลองให้ความสำคัญ ได้แก่ รายได้รวม (IncomeTotal), จำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือน (MonthlyPayment), อายุของผู้ขอสินเชื่อ (Age) และหนี้สินรวมต่อเดือน (LiabilitiesTotal) ตามลำดับ

พหุน ปณ ฑิโต ชีเว



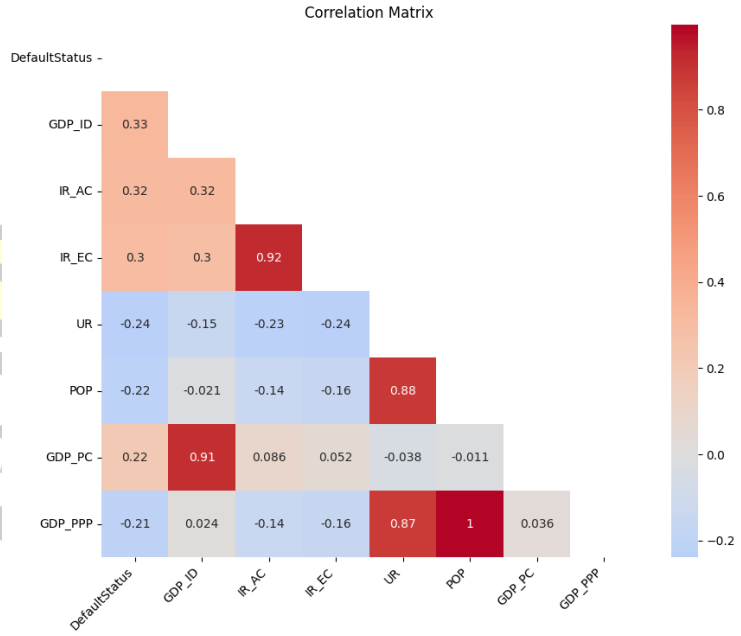
รูปที่ 13 ปัจจัยพื้นฐานที่นำไปสู่การผิดนัดชำระหนี้โดยการหาความสำคัญของคุณลักษณะจากแบบจำลอง XGBoost

จากรูปที่ 13 พบว่าปัจจัยที่แบบจำลอง XGBoost ให้ความสำคัญสูงที่สุด คือ ผู้ขอสินเชื่ออาศัยอยู่ในประเทศเอสโตเนีย (Country_Estonia) นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่แบบจำลองให้ความสำคัญ ได้แก่ ผู้ขอสินเชื่อให้ข้อมูลรายได้และค่าใช้จ่ายในการตรวจสอบคำขอสินเชื่อ (VerificationType_Income and expenses verified), อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest), ผู้ขอสินเชื่ออาศัยอยู่ในประเทศเนเธอร์แลนด์ (Country_Netherlands) และผู้ขอสินเชื่อไม่ได้ระบุประสบการณ์การทำงาน(WorkExperience_Not_Specified) ตามลำดับ

จากรูปที่ 11 - รูปที่ 13 แสดงค่าความสำคัญของคุณลักษณะที่ถูกปรับให้เป็นค่ามาตรฐาน (Normalized Feature Importance Score) โดยมีเกณฑ์ที่ผลรวมสะสม ≥ 0.8 จากผลลัพธ์พบว่าจำนวนตัวแปรที่เหมาะสมคือ 21, 26 และ 51 สำหรับแบบจำลอง LightGBM, Random Forest และ XGBoost ตามลำดับ

4.2 ปัจจัยทางเศรษฐศาสตร์มหภาคที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลจากการการวิเคราะห์ความสัมพันธ์ (Correlation) ของตัวแปรด้านเศรษฐศาสตร์มหภาคและตัวแปรตาม

ในการวิเคราะห์นี้ ผู้วิจัยได้ทำการคัดกรองตัวแปรด้านเศรษฐศาสตร์มหภาคที่มีค่า |Correlation| ≥ 0.2 กับตัวแปรตาม ซึ่งเป็นเกณฑ์ที่ใช้เพื่อระบุตัวแปรที่มีความสัมพันธ์ทางสถิติเพียงพอที่จะนำไปใช้ในการฝึกอบรมแบบจำลอง ผลลัพธ์แสดงให้เห็นถึงปัจจัยทางเศรษฐศาสตร์มหภาคที่มีผลต่อการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ได้แก่ ผลิตภัณฑ์มวลรวมในประเทศต่อหัวเทียบกับความเท่าเทียมกันของกำลังซื้อดอลลาร์ระหว่างประเทศ (GDP_ID), อัตราเงินเฟ้อเทียบกับดัชนีราคาผู้บริโภคโดยเฉลี่ย (IR_AC), อัตราเงินเฟ้อเทียบกับมูลค่าสิ้นงวดของดัชนีราคาผู้บริโภค (IR _EC), อัตราการว่างงาน (UR), จำนวนประชากร (POP), ผลิตภัณฑ์มวลรวมในประเทศต่อหัว (GDP_PC) และผลิตภัณฑ์มวลรวมภายในประเทศตามความเท่าเทียมกันของกำลังซื้อ (GDP_PPP) ตามลำดับ แสดงในรูปที่ 14



รูปที่ 14 ตัวแปรด้านเศรษฐศาสตร์มหภาคที่มีค่า |Correlation| ≥ 0.2 กับตัวแปรตาม

4.3 การเปรียบเทียบประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยชุดของตัวแปรที่แตกต่างกัน 4 รูปแบบ

ในส่วนนี้จะนำเสนอประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยชุดของตัวแปรที่แตกต่างกันในรูปแบบของ Classification Report ซึ่งวัดประสิทธิภาพของแบบจำลองด้วย Accuracy, Precision, Recall และ F1-Score ดังต่อไปนี้

ตารางที่ 3 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยตัวแปรทั้งหมด

อัลกอริทึม	จำนวนตัวแปร	Class	Precision	Recall	F1-Score	Accuracy
LightGBM	ทั้งหมด	Default	0.69	0.57	0.62	0.78
		No Default	0.81	0.88	0.84	
Random Forest	ทั้งหมด	Default	0.72	0.59	0.65	0.79
		No Default	0.82	0.89	0.85	
XGBoost	ทั้งหมด	Default	0.7	0.58	0.63	0.78
		No Default	0.81	0.88	0.84	

จากตารางที่ 3 แสดงผลลัพธ์ของการทดสอบแบบจำลองทั้ง 3 อัลกอริทึมที่ฝึกอบรมด้วยชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมด และข้อมูลไม่สมดุล ชุดข้อมูลทดสอบ (Testing Dataset) ประกอบด้วย คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) จำนวน 23,440 ราย และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) จำนวน 48,179 ราย ผลลัพธ์แสดงให้เห็นว่าแบบจำลองมีปัญหาในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) โดยมีค่า Precision, Recall และ F1-Score ต่ำกว่าคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) เป็นอย่างมาก ซึ่งเกิดจากปัญหาข้อมูลไม่สมดุล ทำให้แบบจำลองมีแนวโน้มที่จะพยากรณ์คลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ได้ถูกต้องมากกว่า

พหุน ปณฺฑิโต ชีเว

ตารางที่ 4 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยตัวแปรทั้งหมด และจัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling)

อัลกอริทึม	จำนวนตัวแปร	Class	Precision	Recall	F1-Score	Accuracy
LightGBM	ทั้งหมด	Default	0.75	0.77	0.76	0.76
		No Default	0.76	0.74	0.75	
Random Forest	ทั้งหมด	Default	0.76	0.79	0.77	0.77
		No Default	0.78	0.75	0.77	
XGBoost	ทั้งหมด	Default	0.75	0.77	0.76	0.76
		No Default	0.77	0.75	0.76	

จากตารางที่ 4 แสดงผลลัพธ์ของการทดสอบแบบจำลองทั้ง 3 อัลกอริทึม ที่ฝึกอบรมด้วยชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมด โดยมีการจัดการกับปัญหาข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) ชุดข้อมูลทดสอบ (Testing Dataset) ประกอบด้วย คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) จำนวน 23,204 ราย และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) จำนวน 23,260 ราย ผลลัพธ์แสดงให้เห็นว่า Accuracy ของแบบจำลองทั้ง 3 มีค่าลดลงเล็กน้อย แต่ยังอยู่ในระดับที่ยอมรับได้ ในแง่ของการพยากรณ์คลาสทั้ง 2 มีความสมดุลมากขึ้น จากค่า Precision, Recall และ F1-Score ที่มีความใกล้เคียงกันมากขึ้น แสดงให้เห็นว่าแบบจำลองมีความสมดุลมากขึ้นในการพยากรณ์คลาสทั้งสอง แสดงให้เห็นว่าการจัดการกับข้อมูลไม่สมดุลด้วยวิธีสุ่มลด (Undersampling) จึงช่วยให้แบบจำลองสามารถพยากรณ์ทั้งสองคลาสได้อย่างสมดุลมากขึ้น ซึ่งเป็นสิ่งที่สำคัญในบริบทของการวิจัยนี้ เนื่องจากการพยากรณ์การผิดนัดชำระหนี้มีความสำคัญต่อการบริหารความเสี่ยงทางการเงิน

ตารางที่ 5 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) และคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection)

อัลกอริทึม	จำนวนตัวแปร	Class	Precision	Recall	F1-Score	Accuracy
LightGBM	21	Default	0.74	0.77	0.76	0.75
		No Default	0.76	0.73	0.75	
Random Forest	26	Default	0.75	0.78	0.77	0.76
		No Default	0.77	0.74	0.76	
XGBoost	51	Default	0.75	0.78	0.76	0.76
		No Default	0.77	0.74	0.75	

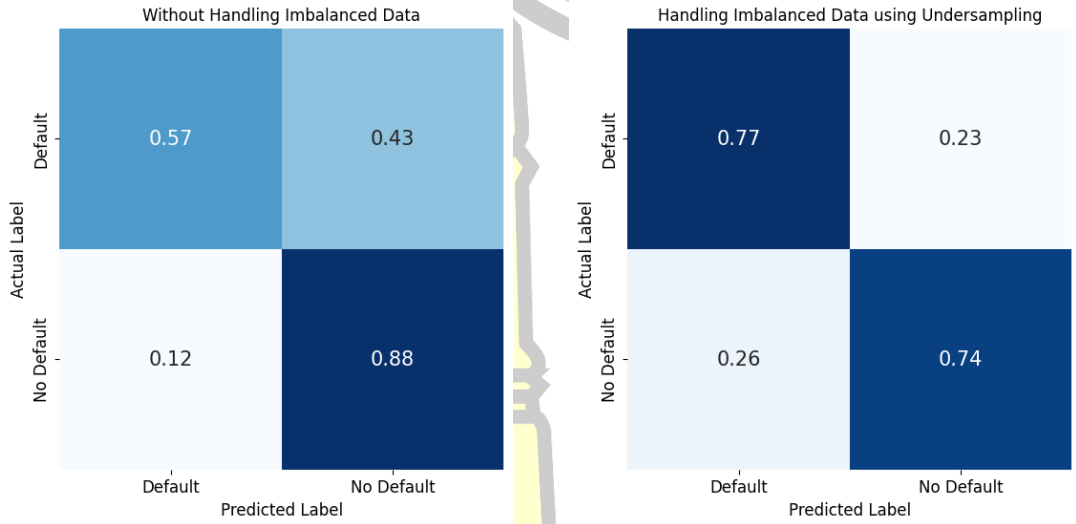
จากตารางที่ 5 แสดงผลลัพธ์ของการทดสอบแบบจำลองทั้ง 3 อัลกอริทึม ที่ฝึกอบรมด้วยชุดข้อมูลที่มีการจัดการกับปัญหาข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) และทำการคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) ตามหัวข้อที่ 4.1 ชุดข้อมูลทดสอบ (Testing Dataset) ประกอบด้วย คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) จำนวน 23,204 ราย และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) จำนวน 23,260 ราย ผลลัพธ์พบว่า Accuracy, Precision, Recall และ F1-Score ของแบบจำลองทั้ง 3 บางส่วนมีค่าลดลงเพียงเล็กน้อย บางส่วนมีค่าเท่าเดิม แสดงให้เห็นว่าการคัดเลือกคุณลักษณะช่วยลดจำนวนตัวแปรที่ใช้ในการฝึกอบรมแบบจำลองได้อย่างมีประสิทธิภาพ โดยไม่ทำให้ประสิทธิภาพในการพยากรณ์ของแบบจำลองลดลงมากนัก

ตารางที่ 6 Classification Report ของแบบจำลองที่ฝึกอบรมด้วยข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) ทำการคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) และรวมตัวแปรด้านเศรษฐศาสตร์มหภาค (Macroeconomic Dataset)

อัลกอริทึม	จำนวนตัวแปร	Class	Precision	Recall	F1-Score	Accuracy
LightGBM	21 + Macro Data	Default	0.75	0.79	0.77	0.76
		No Default	0.78	0.74	0.76	
Random Forest	26 + Macro Data	Default	0.76	0.8	0.78	0.77
		No Default	0.79	0.74	0.76	
XGBoost	51 + Macro Data	Default	0.75	0.79	0.77	0.77
		No Default	0.78	0.74	0.76	

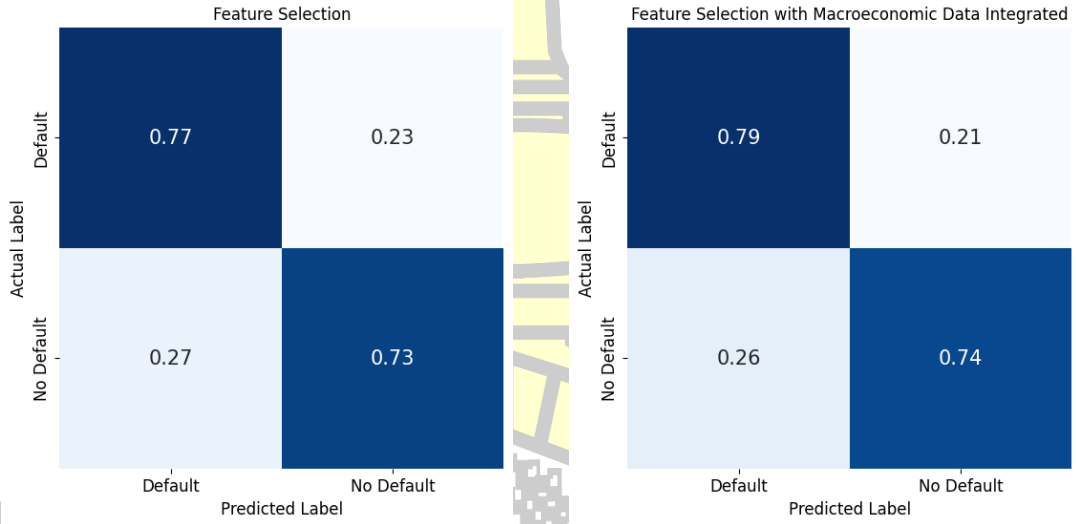
จากตารางที่ 6 แสดงผลลัพธ์ของการทดสอบแบบจำลองทั้ง 3 อัลกอริทึม ที่ฝึกอบรมด้วยชุดข้อมูลที่มีการจัดการกับปัญหาข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) มีการคัดกรองตัวแปรโดยการคัดเลือกคุณลักษณะ (Feature Selection) ตามหัวข้อที่ 4.1 จากนั้นรวมตัวแปรด้านเศรษฐศาสตร์มหภาคที่ได้รับคัดเลือกตามหัวข้อที่ 4.2 ชุดข้อมูลทดสอบ (Testing Dataset) ประกอบด้วย คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) จำนวน 23,204 ราย และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) จำนวน 23,260 ราย ผลลัพธ์พบว่า Accuracy, Precision, Recall และ F1-Score ของแบบจำลองทั้ง 3 บางส่วนมีค่าสูงกว่าระดับการพยากรณ์ที่ใช้ตัวแปรทั้งหมด (จากตารางที่ 4) แสดงให้เห็นว่าการคัดเลือกคุณลักษณะและรวมตัวแปรด้านเศรษฐศาสตร์มหภาค ทำให้การฝึกอบรมแบบจำลองมีประสิทธิภาพในการพยากรณ์ลูกค้าที่มีการผิดนัดชำระหนี้

และลูกค้าที่ไม่มีการผิดนัดชำระหนี้เพิ่มขึ้น มากกว่าการใช้ตัวแปรพื้นฐานด้านสินเชื่อทั้งหมดเพียงอย่างเดียว



15(ก) ข้อมูลไม่สมดุล

15(ข) ข้อมูลสมดุล



15(ค) การคัดเลือกคุณลักษณะ

15(ง) รวมตัวแปรเศรษฐศาสตร์มหภาค

รูปที่ 15 Normalized Confusion Matrix แบบจำลองที่ผึกอบรมโดยอัลกอริทึม LightGBM

ประสิทธิภาพของแบบจำลองที่ผึกอบรมโดยอัลกอริทึม LightGBM จากรูปที่ 15 พบว่าประสิทธิภาพของแบบจำลองที่ผึกอบรมด้วยข้อมูลไม่สมดุล (รูปที่ 15(ก)) ทำให้แบบจำลองมีปัญหาในการพยากรณ์คลาส Default ที่มีค่าพยากรณ์ได้ถูกต้องเพียง 57% เมื่อจัดการข้อมูลไม่สมดุลด้วยวิธีสุ่มลด (รูปที่ 15(ข)) แบบจำลองสามาถพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) ได้เพิ่มขึ้นเป็น 77% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) ลดลงเป็น 74% อยู่ใน

เกณฑ์ที่ยอมรับได้ แสดงถึงความสมดุลของแบบจำลองที่ดีขึ้น เมื่อทำการคัดเลือกคุณลักษณะ (รูปที่ 15(ค)) ความสามารถในการพยากรณ์คลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ลดลงเล็กน้อยเป็น 73% ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ การรวมตัวแปรเศรษฐศาสตร์มหภาค (รูปที่ 15(ง)) ทำให้ความสามารถในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้เพิ่มขึ้นเป็น 79% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้เทียบเท่ากับแบบจำลองที่ใช้ตัวแปรทั้งหมดจากข้อมูลที่มีการจัดการข้อมูลไม่สมดุล แสดงให้เห็นว่าการพยากรณ์มีความแม่นยำเทียบเท่ากับการใช้ทุกตัวแปรด้านสินเชื่อในการฝึกอบรมแบบจำลอง

ตัวอย่างการคำนวน Accuracy, Precision, Recall และ F1-Score จาก Normalized Confusion Matrix แบบจำลอง LightGBM ที่ฝึกอบรมด้วยตัวแปรทั้งหมด และข้อมูลไม่สมดุล จากข้อมูลทดสอบ (Testing Dataset) สำหรับชุดของตัวแปรทั้งหมด และข้อมูลไม่สมดุล ประกอบด้วยคลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) จำนวน 23,440 ราย และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) จำนวน 48,179 ราย จะได้ว่า

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{(0.57\cdot 23440)+(0.88\cdot 48179)}{(0.57\cdot 23440)+(0.43\cdot 23440)+(0.12\cdot 48179)+(0.88\cdot 48179)} = 0.7785$$

สำหรับคลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default)

$$Precision = \frac{TP}{TP+FP} = \frac{(0.57\cdot 23440)}{(0.57\cdot 23440)+(0.12\cdot 48179)} = 0.698$$

$$Recall = \frac{TP}{TP+FN} = \frac{(0.57\cdot 23440)}{(0.57\cdot 23440)+(0.43\cdot 23440)} = 0.57$$

$$F1-Score = \frac{2\cdot Precision\cdot Recall}{Precision+Recall} = \frac{2\cdot 0.698\cdot 0.57}{0.698+0.57} = 0.6275$$

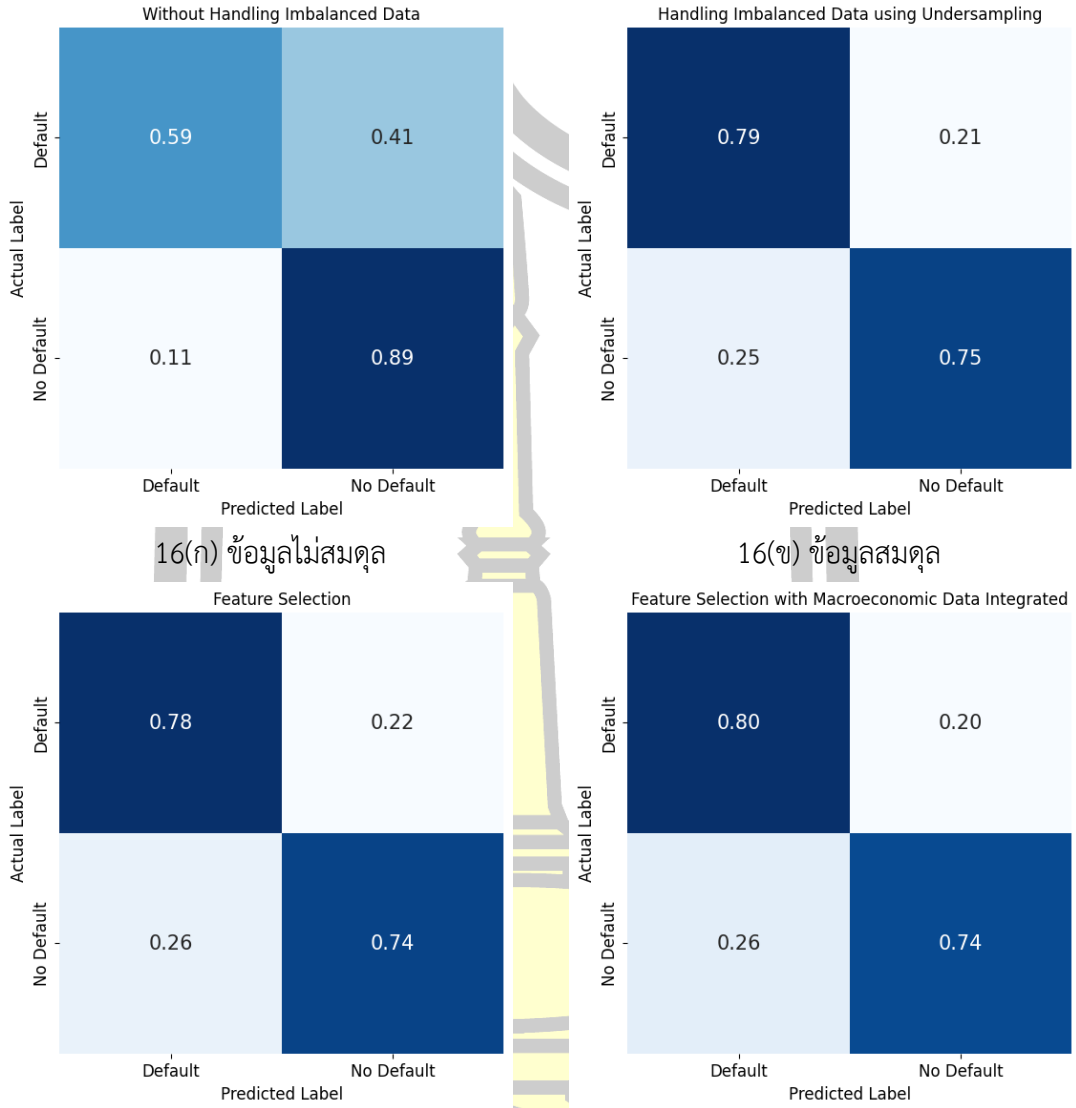
สำหรับคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default)

$$Precision = \frac{TN}{TN+FN} = \frac{(0.88\cdot 48179)}{(0.88\cdot 48179)+(0.43\cdot 23440)} = 0.8079$$

$$Recall = \frac{TN}{TN+FP} = \frac{(0.88\cdot 48179)}{(0.88\cdot 48179)+(0.12\cdot 48179)} = 0.88$$

$$F1-Score = \frac{2\cdot Precision\cdot Recall}{Precision+Recall} = \frac{2\cdot 0.8079\cdot 0.88}{0.8079+0.88} = 0.8424$$

ซึ่งสอดคล้องกับตารางที่ 3



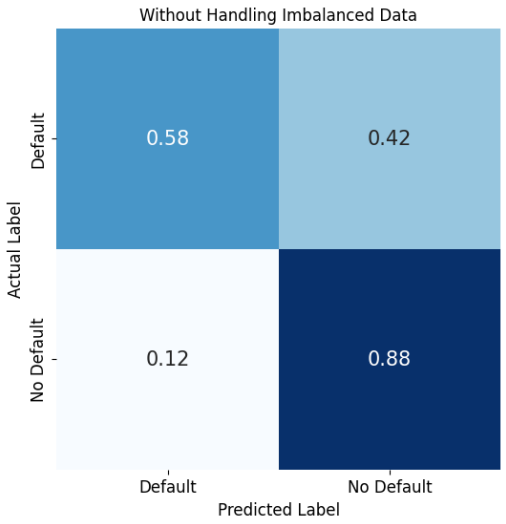
16(ก) ข้อมูลไม่สมดุล 16(ข) ข้อมูลสมดุล

16(ค) การคัดเลือกคุณลักษณะ 16(ง) รวมตัวแปรเศรษฐศาสตร์มหภาค

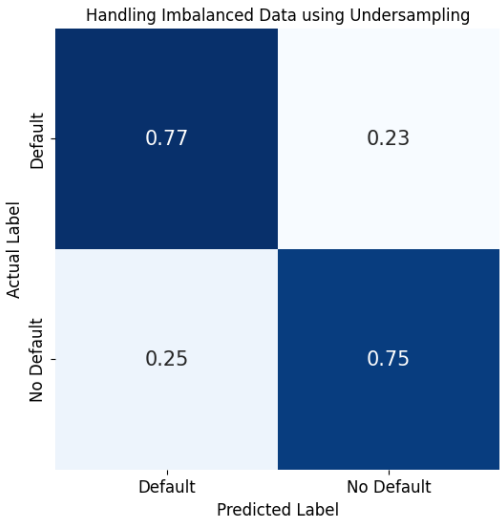
รูปที่ 16 Normalized Confusion Matrix แบบจำลองที่ผึกอบรมโดยอัลกอริทึม Random Forest

ประสิทธิภาพของแบบจำลองที่ฝึกอบรมโดยอัลกอริทึม Random Forest จากรูปที่ 16 พบว่าประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยข้อมูลไม่สมดุล (รูปที่ 16(ก)) ทำให้แบบจำลองมีปัญหาในการพยากรณ์คลาส Default ที่มีค่าพยากรณ์ได้ถูกต้องเพียง 59% เมื่อจัดการข้อมูลไม่สมดุลด้วยวิธีสุ่มลด (รูปที่ 16(ข)) แบบจำลองสามารถพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) ได้เพิ่มขึ้นเป็น 79% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) ลดลงเป็น 75% อยู่ในเกณฑ์ที่ยอมรับได้ แสดงถึงความสมดุลของแบบจำลองที่ดีขึ้น เมื่อทำการคัดเลือกคุณลักษณะ (รูปที่ 16(ค)) ความสามารถในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้และลูกค้าไม่มีการผิดนัดชำระหนี้ลดลงเล็กน้อยเป็น 78% และ 74% ตามลำดับ ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ การรวมตัวแปร

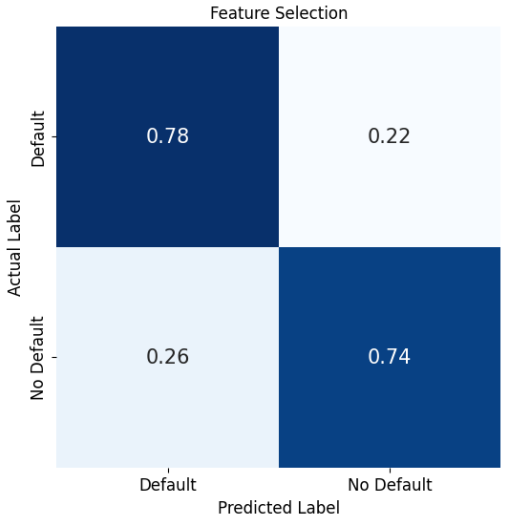
เศรษฐศาสตร์มหภาค (รูปที่ 16(ง)) ทำให้ความสามารถในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้เพิ่มขึ้นเป็น 80% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ลดลงเล็กน้อยเป็น 74% เทียบเท่ากับแบบจำลองที่ใช้ตัวแปรทั้งหมดจากข้อมูลที่มีการจัดการข้อมูลไม่สมดุล แสดงให้เห็นว่าการพยากรณ์มีความแม่นยำและสมดุลเทียบเท่ากับการใช้ทุกตัวแปรด้านสินเชื่อในการฝึกอบรมแบบจำลอง



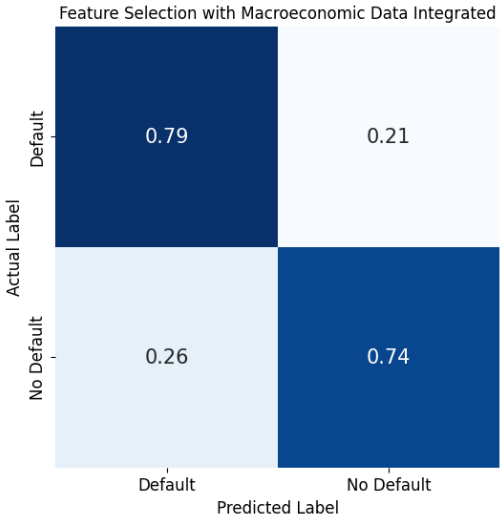
17(ก) ข้อมูลไม่สมดุล



17(ข) ข้อมูลสมดุล



17(ค) การคัดเลือกคุณลักษณะ



17(ง) รวมตัวแปรเศรษฐศาสตร์มหภาค

รูปที่ 17 Normalized Confusion Matrix แบบจำลองที่ฝึกอบรมโดยอัลกอริทึม XGBoost

ประสิทธิภาพของแบบจำลองที่ฝึกอบรมโดยอัลกอริทึม XGBoost จากรูปที่ 17 พบว่าประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยข้อมูลไม่สมดุล (รูปที่ 17(ก)) ทำให้แบบจำลองมีปัญหาในการพยากรณ์คลาส Default ที่มีค่าพยากรณ์ได้ถูกต้องเพียง 58% เมื่อจัดการข้อมูลไม่สมดุลด้วยวิธี

สุ่มลด (รูปที่ 17(ข)) แบบจำลองสามาถพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) ได้เพิ่มขึ้นเป็น 77% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) ลดลงเป็น 75% อยู่ในเกณฑ์ที่ยอมรับได้ แสดงถึงความสมดุลของแบบจำลองที่ดีขึ้น เมื่อทำการคัดเลือกคุณลักษณะ (รูปที่ 17(ค)) ความสามารถในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้เพิ่มขึ้นเป็น 78% และลูกค้าไม่มีการผิดนัดชำระหนี้ลดลงเล็กน้อยเป็น 74% ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ การรวมตัวแปรเศรษฐศาสตร์มหภาค (รูปที่ 17(ง)) ทำให้ความสามารถในการพยากรณ์คลาสที่ลูกค้ามีการผิดนัดชำระหนี้เพิ่มขึ้นเป็น 79% และคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ลดลงเล็กน้อยเป็น 74% เทียบเท่ากับแบบจำลองที่ใช้ตัวแปรทั้งหมดจากข้อมูลที่มีการจัดการข้อมูลไม่สมดุล แสดงให้เห็นว่าการพยากรณ์มีความแม่นยำและสมดุลเทียบเท่ากับการใช้ทุกตัวแปรด้านสินเชื่อในการฝึกอบรมแบบจำลอง

พหุน ปณฺฑิโต ชีเว

บทที่ 5
สรุปผล อภิปรายผล และข้อเสนอแนะ

ในการศึกษาวิจัยการพัฒนาแบบจำลองการเรียนรู้ของเครื่องเพื่อพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล โดยใช้อัลกอริทึม LightGBM, Random Forest และ XGBoost ด้วยชุดข้อมูลสินเชื่อของ Bondora ซึ่งเป็นแพลตฟอร์มการให้กู้ยืมแบบ P2P ชั้นนำของยุโรป และชุดข้อมูลด้านเศรษฐศาสตร์มหภาคของ World Economic Outlook (WEO) ที่รวบรวมโดยเจ้าหน้าที่ของ IMF ผู้วิจัยสามารถสรุปผลการศึกษา อภิปรายผล และข้อเสนอแนะได้ดังนี้

5.1 สรุปผล

การวิเคราะห์ความสำคัญของคุณลักษณะ (Feature Importance) จากแบบจำลองที่ได้รับการฝึกอบรมด้วยชุดข้อมูลที่จัดการกับข้อมูลไม่สมดุลโดยใช้วิธีสุ่มลด (Undersampling) พบว่าปัจจัย 5 อันดับแรกที่แต่ละแบบจำลองให้ความสำคัญเป็นดังนี้ 1. แบบจำลองจากอัลกอริทึม LightGBM ได้แก่ อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest), จำนวนเงินที่ผู้ขอสินเชื่อได้ชำระก่อนหน้านี้ (PreviousRepaymentsBeforeLoan), จำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือน (MonthlyPayment), รายได้รวม (IncomeTotal) และระยะเวลาของสินเชื่อ (LoanDuration) ตามลำดับ 2. แบบจำลองจากอัลกอริทึม Random Forest ได้แก่ อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest), รายได้รวม (IncomeTotal), จำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือน (MonthlyPayment), อายุของผู้ขอสินเชื่อ (Age) และหนี้สินรวมต่อเดือน (LiabilitiesTotal) ตามลำดับ 3. แบบจำลองจากอัลกอริทึม XGBoost ได้แก่ ผู้ขอสินเชื่ออาศัยอยู่ในประเทศเอสโตเนีย (Country_Estonia), ผู้ขอสินเชื่อให้ข้อมูลรายได้และค่าใช้จ่ายในการตรวจสอบคำขอสินเชื่อ (VerificationType_Income and expenses verified), อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อ (Interest), ผู้ขอสินเชื่ออาศัยอยู่ในประเทศเนเธอร์แลนด์ (Country_Netherlands) และผู้ขอสินเชื่อไม่ได้ระบุประสบการณ์การทำงาน (WorkExperience_Not_Specified) ตามลำดับ จากการวิเคราะห์ความสำคัญของคุณลักษณะของทั้ง 3 แบบจำลอง พบว่า อัตราดอกเบี้ยสูงสุดที่ยอมรับในการขอสินเชื่อเป็นปัจจัยที่สำคัญที่สุดที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล โดยผู้ขอสินเชื่อที่ยอมรับอัตราดอกเบี้ยสูงมีแนวโน้มที่จะผิดนัดชำระหนี้มากกว่าผู้ขอสินเชื่อที่ยอมรับอัตราดอกเบี้ยต่ำ นอกจากนี้ปัจจัยอื่น ๆ ที่มีความสำคัญสูง ได้แก่ รายได้รวมของผู้ขอสินเชื่อ ซึ่งผู้ขอสินเชื่อที่มีรายได้สูง

มีโอกาสที่จะผิดนัดชำระหนี้น้อยกว่าผู้ที่มีรายได้ต่ำ และจำนวนเงินโดยประมาณที่ผู้ขอสินเชื่อต้องชำระทุกเดือนที่สูงยังเพิ่มความเสี่ยงในการผิดนัดชำระหนี้

การวิเคราะห์ความสัมพันธ์ (Correlation) ของตัวแปรด้านเศรษฐศาสตร์มหภาคและตัวแปรตาม ซึ่งตัวแปรตาม (DefaultStatus) ถูกกำหนดให้ค่า 0 หมายถึง ลูกค้ามีการผิดนัดชำระหนี้ (Default) และค่า 1 หมายถึง ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) โดยใช้เกณฑ์ที่ค่า Correlation ≥ |0.2| พบว่า ผลิตภัณฑ์มวลรวมในประเทศต่อหัวเทียบกับความเท่าเทียมกันของกำลังซื้อดอลลาร์ระหว่างประเทศ (GDP_ID) เป็นปัจจัยที่มีความสำคัญสูงที่สุดและมีความสัมพันธ์เชิงบวกสูงที่สุดในบรรดาปัจจัยทั้งหมด ชี้ให้เห็นว่าประเทศของผู้ขอสินเชื่อที่มีผลิตภัณฑ์มวลรวมในประเทศต่อหัวเทียบกับความเท่าเทียมกันของกำลังซื้อดอลลาร์ระหว่างประเทศสูงมีโอกาสน้อยที่จะเกิดการผิดนัดชำระหนี้ นอกจากนี้ปัจจัยอื่นที่มีความสัมพันธ์เชิงบวกกับตัวแปรตาม ได้แก่ อัตราเงินเฟ้อเทียบกับดัชนีราคาผู้บริโภคโดยเฉลี่ย (IR_AC), อัตราเงินเฟ้อเทียบกับมูลค่าสิ้นงวดของดัชนีราคาผู้บริโภค (IR_EC) และผลิตภัณฑ์มวลรวมในประเทศต่อหัว (GDP_PC) ซึ่งแสดงให้เห็นว่าเมื่อปัจจัยเหล่านี้เพิ่มสูงขึ้น ความเสี่ยงในการผิดนัดชำระหนี้กลับมีแนวโน้มลดลง ในทางกลับกัน ปัจจัยที่มีความสัมพันธ์เชิงลบกับตัวแปรตาม แสดงให้เห็นว่าประเทศของผู้ขอสินเชื่อที่มีอัตราการว่างงาน (UR) จำนวนประชากร (POP) และผลิตภัณฑ์มวลรวมภายในประเทศตามความเท่าเทียมกันของกำลังซื้อ (GDP_PPP) ที่สูงจะส่งผลให้ความเสี่ยงในการผิดนัดชำระหนี้เพิ่มขึ้น

การพัฒนาแบบจำลองการเรียนรู้ของเครื่องโดยใช้อัลกอริทึม LightGBM, Random Forest และ XGBoost ที่ผู้วิจัยได้ทำการแบ่งชุดข้อมูลที่มีตัวแปรแตกต่างกัน 4 รูปแบบ ได้แก่ 1. ชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมดและข้อมูลไม่สมดุล 2. ชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมดและข้อมูลสมดุล 3. ชุดข้อมูลที่ข้อมูลสมดุลและมีการคัดเลือกคุณลักษณะ และ 4. ชุดข้อมูลที่ข้อมูลสมดุลมีการคัดเลือกคุณลักษณะและรวมตัวแปรเศรษฐศาสตร์มหภาค เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในแง่ของความสามารถในการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ผลการวัดประสิทธิภาพของแบบจำลองแสดงในตารางที่ 7

ตารางที่ 7 ผลการวัดประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องที่ฝึกอบรมด้วยชุดข้อมูลที่มีตัวแปรที่แตกต่างกัน 4 รูปแบบ

ชุดข้อมูล	Accuracy	Precision		Recall		F1-Score	
		Default	No Default	Default	No Default	Default	No Default
1. ชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมดและข้อมูลไม่สมดุล	0.78 – 0.79	0.69 – 0.72	0.81 – 0.82	0.57 – 0.59	0.88 – 0.89	0.62 – 0.65	0.84 – 0.85
2. ชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมดและข้อมูลสมดุล	0.76 – 0.77	0.75 – 0.76	0.76 – 0.78	0.77 – 0.79	0.74 – 0.75	0.76– 0.77	0.75 – 0.77
3. ชุดข้อมูลที่ข้อมูลสมดุลและมีการคัดเลือกคุณลักษณะ	0.75 – 0.76	0.74 – 0.75	0.76 – 0.77	0.77 – 0.78	0.73 – 0.74	0.76– 0.77	0.75 – 0.76
4. ชุดข้อมูลที่ข้อมูลสมดุลมีการคัดเลือกคุณลักษณะและรวมตัวแปรเศรษฐศาสตร์มหภาค	0.76 – 0.77	0.75 – 0.76	0.78 – 0.79	0.79 – 0.78	0.74	0.77 – 0.78	0.76

จากตารางที่ 7 ประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยชุดข้อมูลที่ประกอบด้วยตัวแปรทั้งหมดและข้อมูลไม่สมดุล พบว่าค่า Precision, Recall และ F1-Score ระหว่างคลาส Default และ No Default มีความไม่สมดุลกันอย่างชัดเจน โดยคลาส No Default มีค่า Precision และ Recall สูงกว่าอย่างมาก ในขณะที่คลาส Default มีค่าประสิทธิภาพต่ำ เช่น Recall เพียง 0.57 – 0.59 และ F1-Score เพียง 0.62 – 0.65 ทำให้แบบจำลองไม่เหมาะสมต่อการใช้งานจริง

เมื่อแก้ปัญหาข้อมูลไม่สมดุลด้วยการสุ่มลดข้อมูล (Undersampling) พบว่าค่าประสิทธิภาพของทั้งสองคลาสมีความสมดุลกันมากขึ้น คลาส Default มี Recall เพิ่มขึ้นเป็น 0.77 – 0.79 และ F1-Score เป็น 0.76 – 0.77 แสดงให้เห็นถึงความเหมาะสมของการจัดการข้อมูลไม่สมดุล

หลังจากนั้น ผู้วิจัยได้ทำการคัดเลือกคุณลักษณะ (Feature Selection) เพื่อลดจำนวนตัวแปรที่ใช้ โดยยังคงค่าประสิทธิภาพในระดับที่ยอมรับได้ แม้ว่าค่า Accuracy จะลดลงเล็กน้อย แต่แบบจำลองยังคงมีประสิทธิภาพที่สมดุลระหว่างคลาสทั้งสอง

สุดท้ายเมื่อรวมตัวแปรด้านเศรษฐศาสตร์มหภาคที่ผ่านการคัดเลือกตามค่าความสัมพันธ์ (Correlation) เข้าไปในชุดข้อมูล พบว่าแบบจำลองมีประสิทธิภาพเพิ่มขึ้นอย่างเห็นได้ชัด โดย Recall ของคลาส Default เพิ่มเป็น 0.79 – 0.8 และ F1-Score เพิ่มเป็น 0.77 – 0.78 แสดงถึงความสำคัญของการนำตัวแปรด้านเศรษฐศาสตร์มหภาคเข้ามาใช้ร่วมกับตัวแปรด้านสินเชื่อ ซึ่งช่วยให้แบบจำลองมีความแม่นยำและครอบคลุมมากขึ้น

5.2 อภิปรายผล

จากผลการวิจัยการวิเคราะห์ความสำคัญของคุณลักษณะ (Feature Importance) สามารถระบุปัจจัยที่สำคัญที่นำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลได้อย่างมีประสิทธิภาพ ซึ่งช่วยลดการใช้ตัวแปรที่มากเกินไปในการพัฒนาแบบจำลอง ลดภาระงานในการเก็บรวบรวมข้อมูล (Data Collecting) และลดพื้นที่ในการจัดเก็บข้อมูล (Data Storage) โดยที่ยังสามารถพยากรณ์การผิดนัดชำระหนี้เงินกู้ได้อย่างมีประสิทธิภาพ

การศึกษาข้อมูลสินเชื่อร่วมกับข้อมูลด้านเศรษฐศาสตร์มหภาคแสดงให้เห็นว่าพฤติกรรมทางเศรษฐกิจในระดับกว้างส่งผลต่อการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล และยังเป็นข้อมูลสาธารณะที่สามารถเข้าถึงได้ง่าย สามารถเพิ่มประสิทธิของแบบจำลองในการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลได้อย่างมีประสิทธิภาพ ซึ่งสอดคล้องกับงานวิจัยของ Nigmonov et al. (2022) ที่ศึกษาปัจจัยทางเศรษฐศาสตร์ภาคที่สำคัญที่มีอิทธิพลต่อการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคลโดยใช้การวิเคราะห์การถดถอย (Regression Analysis) ผลที่ได้พบว่าอัตราเงินเฟ้อที่สูงขึ้นจะเพิ่มอัตราการผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล ซึ่งเป็นปัจจัยเดียวกันที่พบในงานวิจัยนี้

การเปรียบเทียบประสิทธิภาพของแบบจำลองที่ฝึกอบรมด้วยอัลกอริทึม LightGBM, Random Forest และ XGBoost จากที่พบว่าข้อมูลไม่สมดุล (Imbalanced data) ซึ่งส่งผลให้แบบจำลองที่พัฒนาขึ้นไม่สามารถพยากรณ์คลาสทั้งสองแม่นยำ ผู้วิจัยได้แก้ไขปัญหาโดยการใช้วิธีสุ่มลด (Undersampling) เพื่อให้ข้อมูลสมดุล และสามารถนำไปพัฒนาแบบจำลองได้อย่างมีประสิทธิภาพ การวิเคราะห์ความสำคัญของคุณลักษณะ (Feature Importance) ช่วยให้สามารถกำหนดจำนวนตัวแปรที่เหมาะสมสำหรับแต่ละแบบจำลอง จากนั้นเมื่อทำการรวมตัวแปรด้านเศรษฐศาสตร์มหภาคในขั้นตอนการพัฒนาแบบจำลอง พบว่าแต่ละแบบจำลองมีประสิทธิภาพเทียบเท่ากับการใช้ตัวแปรด้านสินเชื่อทั้งหมดเพียงอย่างเดียวทั้ง 3 แบบจำลอง ในแง่ของการเลือกใช้จำนวนตัวแปรสำหรับการพัฒนาแบบจำลองให้น้อยที่สุดเพื่อลดภาระงานในการเก็บรวบรวมข้อมูล, พื้นที่ในการจัดเก็บข้อมูล, ระยะเวลา และการประมวลผลที่ซับซ้อน ในกรณีนี้แบบจำลองที่ฝึกอบรมด้วยอัลกอริทึม LightGBM ที่ใช้ตัวแปรน้อยที่สุด ได้แก่ ตัวแปรด้านสินเชื่อ 21 ตัวแปร และตัวแปรด้านเศรษฐศาสตร์มหภาค 7 ตัวแปร อาจเป็นทางเลือกที่ดี แต่ในแง่ของประสิทธิภาพในการพยากรณ์การผิดนัดชำระหนี้แบบจำลองที่ฝึกอบรมด้วยอัลกอริทึม Random Forest และ XGBoost มีค่า Accuracy สูงที่สุด สำหรับประสิทธิภาพในการพยากรณ์คลาสทั้งสอง อัลกอริทึม Random Forest มีประสิทธิภาพมากที่สุด โดยคลาสที่ลูกค้าไม่มีการผิดนัดชำระหนี้ (No Default) มีค่า Precision, Recall และ F1-Score เท่ากับ 0.79, 0.74 และ 0.76 ตามลำดับ คลาสที่ลูกค้ามีการผิดนัดชำระหนี้ (Default) มีค่า Precision, Recall และ F1-Score เท่ากับ 0.76, 0.8 และ 0.78 ตามลำดับ

อย่างไรก็ตามหากคำนึงถึงทั้งปัจจัยด้านการใช้ตัวแปรที่น้อย และแบบจำลองมีประสิทธิภาพในการพยากรณ์การผิดนัดชำระหนี้ แบบจำลองที่ฝึกอบรมด้วยอัลกอริทึม Random Forest ที่มีประสิทธิภาพในการพยากรณ์สูงที่สุด และใช้ตัวแปรด้านสินเชื่อ 26 ตัวแปร และตัวแปรด้านเศรษฐศาสตร์มหภาค 7 ตัวแปร ซึ่งมากกว่าอัลกอริทึม LightGBM เล็กน้อย อาจเป็นทางเลือกที่ดีที่สุดในการใช้งานจริง เพื่อประสิทธิภาพในการพยากรณ์ที่สูง อีกทั้งยังใช้ทรัพยากรในการเก็บรวบรวมข้อมูล, พื้นที่ในการจัดเก็บข้อมูล, ความซับซ้อน และระยะเวลาในการประมวลที่ต่ำ

5.3 ข้อเสนอแนะ

1. ในงานวิจัยนี้ได้แก้ปัญหาข้อมูลไม่สมดุล โดยใช้วิธีสุ่มลด (Undersampling) เป็นการลดจำนวนข้อมูลของกลุ่มส่วนมาก (Majority Class) ซึ่งส่งผลให้ข้อมูลที่นำมาศึกษาลดลง แม้จะช่วยปรับสมดุลข้อมูลได้ การเลือกใช้วิธีการจัดการกับข้อมูลไม่สมดุลที่เหมาะสม เช่น การสุ่มเพิ่ม (Upsampling) หรือการใช้เทคนิค Synthetic Minority Over-sampling Technique (SMOTE) อาจช่วยปรับปรุงประสิทธิภาพในการพยากรณ์ของแบบจำลองได้ดียิ่งขึ้น

2. การคัดเลือกคุณลักษณะ (Feature Selection) ที่ใช้ในงานวิจัยนี้ ผู้วิจัยเลือกใช้การวิเคราะห์ความสำคัญของคุณลักษณะ (Feature Importance) และการวิเคราะห์ความสัมพันธ์ (Correlation) เป็นเครื่องมือในการคัดเลือกคุณลักษณะ การศึกษาวิธีการคัดเลือกคุณลักษณะอื่น ๆ ที่เหมาะสม อาจนำไปสู่การคัดเลือกตัวแปรที่เหมาะสมมากกว่า และปรับปรุงความแม่นยำของแบบจำลองได้

3. ข้อมูลที่นำมาศึกษาในงานวิจัยนี้เป็นข้อมูลสินเชื่อของ Bondora ซึ่งเป็นแพลตฟอร์มการให้กู้ยืมแบบ P2P ชั้นนำของยุโรป การนำแบบจำลองที่ได้จากงานนี้วิจัยนี้ไปใช้กับข้อมูลสินเชื่อในพื้นที่อื่น ๆ อาจไม่มีประสิทธิภาพที่ดี เนื่องจากข้อมูลสินเชื่อมีความแตกต่างกันออกไปในแต่ละประเทศ ควรเลือกใช้ข้อมูลในการพัฒนาแบบจำลองให้ใกล้เคียงกับกลุ่มเป้าหมายจริง

4. ศึกษาตัวแปรด้านอื่น ๆ ที่อาจนำไปสู่การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล นอกเหนือจากตัวแปรด้านเศรษฐศาสตร์มหภาค เพื่อประสิทธิภาพของแบบจำลองที่ดียิ่งขึ้น

5. ศึกษาและทดสอบอัลกอริทึมการเรียนรู้ของเครื่องอื่น ๆ เพิ่มเติม เช่น Deep Learning หรือ Ensemble Learning ที่ซับซ้อนขึ้น เพื่อหาวิธีการที่เหมาะสมในการพัฒนาแบบจำลองการพยากรณ์การผิดนัดชำระหนี้บนแพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล

พหุน ปณฺ ฑิโต ชีเว

บรรณานุกรม

Azhari, M., Alaoui, A., Achraoui, Z., Ettaki, B., & Zerouaoui, J. (2019). Adaptation of the random forest method: solving the problem of pulsar search. The 4th International Conference on Smart City Applications,

Bloomenthal, A. (2022). *Macroeconomic Factor: Definition, Types, Examples, and Impact*. Investopedia. https://www.investopedia.com/terms/m/macroeconomic-factor.asp

Capital One. (2022). *Peer-to-Peer (P2P) Lending: How Does It Work?* https://www.capitalone.com/learn-grow/money-management/peer-to-peer-lending/

CFI Team. (2022). *Peer-to-Peer Lending*. Corporate Finance Institute. https://corporatefinanceinstitute.com/resources/wealth-management/peer-to-peer-lending/

Chang, A.-H., Yang, L.-K., Tsaih, R.-H., & Lin, S.-K. (2022). Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. *Quantitative Finance and Economics, 6*(2), 303-325.

Çığşar, B., & Ünal, D. (2019). Comparison of Data Mining Classification Algorithms Determining the Default Risk. *Scientific Programming, 2019*.

Dong, X. (2022). Loan Default Prediction based on Machine Learning (LightGBM Model). *BCP Business & Management, 25*, 457-468. https://doi.org/10.54691/bcpbm.v25i.1857

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232, 1144. https://doi.org/10.1214/aos/1013203451

Gangoda, C. A., & Wanninayake, L. (2023). *A Model for Diabetes Prediction Using Machine Learning* Applied Science, Business & Industrial Research Symposium,

Great Learning Team. (2022). *What is Machine Learning? Defination, Types, Applications, and more*. My Great Learning. https://www.mygreatlearning.com/blog/what-is-machine-learning/

Haldar, R., & Tavener, A. (2015). *Luxury P2P lending: How does it work*. Real Business. https://realbusiness.co.uk/luxury-p2p-lending-how-does-it-work

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *Journal of Data Mining & Knowledge Management Process, 5*(2), 01-11.

Intrend. (2019). เศรษฐศาสตร์ มหภาค หมายถึงอะไร *?* http://intrend.co/%E0%B8%97%E0%B8%B1%E0%B9%88%E0%B8%A7%E0%B9%84%E0%B8%9B/%E0%B9%80%E0%B8%A8%E0%B8%A3%E0%B8%A9%E0%B8%90%E0%B8%81%E0%B8%B4%E0%B8%88/%E0%B9%80%E0%B8%A8%E0%B8%A3%E0%B8%A9%E0%B8%90%E0%B8%A8%E0%B8%B2%E0%B8%AA%E0%B8%95%E0%B8%A3%E0%B9%8C-%E0%B8%A1%E0%B8%AB%E0%B8%A0%E0%B8%B2%E0%B8%84-%E0%B8%AB%E0%B8%A1%E0%B8%B2%E0%B8%A2%E0%B8%96%E0%B8%B6%E0%B8%87%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3/

Johnson, J. (2022). *What Is Peer-to-Peer (P2P) Lending?* Business News Daily. https://www.businessnewsdaily.com/16480-peer-to-peer-lending.html

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems,

Lake, R. (2021). *Personal Loan: Definition, Types, and How to Get One*. Investopedia. https://www.investopedia.com/personal-loan-5076027

LeBlanc, V., & Cox, M. A. A. (2017). Interpretation of the point-biserial correlation coefficient in the context of a school examination. *The Quantitative Methods for Psychology, 13*(1), 46-56. https://doi.org/https://doi.org/10.20982/tqmp.13.1.p046

Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. *IOP Conf. Ser.: Mater. Sci. Eng., 1022*.

Nemade, S. (2022). Loan Repayment Ability Using Machine Learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10*(7).

Nigmonov, A., Shams, S., & Alam, K. (2022). Macroeconomic determinants of loan defaults: Evidence from the U.S. peer-to-peer lending market. *Research in International Business and Finance, 59*.

P2P Lending Sites. (2022). *P2P lending history*. P2P Lending Sites. https://p2plendingsites.com/p2p-lending-history/

Profinit. (2022). *How can you move from reacting when customers default, to proactively anticipating loan defaults and solving the problem in advance?* Big Data For Banking. https://bigdataforbanking.com/solutions/credit-risk-monitoring/

ProjectPro. (2023). *Data Preparation for Machine Learning Projects: Know It All Here*. https://www.projectpro.io/article/data-preparation-for-machine-learning/595#mcetoc_1g42dp5ac6u

Sabilla, S. I., Sarno, R., & Triyana, K. (2019). Optimizing Threshold Using Pearson Correlation for Selecting Features of Electronic Nose Signals. *International Journal of Intelligent Engineering and Systems, 12*(6), 81-88. https://doi.org/10.22266/ijies2019.1231.08

Sampaio, F. (2023). *Oversampling and Undersampling Techniques in Fraud Prevention*. https://medium.com/@denizgunay/random-forest-af5bde5d7e1e

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of Default in P2PLending. *PLOS ONE, 10*(10).

Shadroch, G. (2022). *Pros & Cons of Peer-to-Peer Lending*. Sing Saver. https://www.singsaver.com.sg/blog/pros-and-cons-of-peer-to-peer-lending

Sharma, A. K., Li, L.-H., & Ahmad, R. (2023). Default Risk Prediction Using Random Forest and XGBoosting Classifier. 2021 International Conference on Security and Information Technologies with AI, Internet Computing and Big-data Applications, Cham.

Terko, A., Žunić, E., Donko, D., & Dželihodžić, A. (2019, 2019). *Credit Scoring Model Implementation in a Microfinance Context* 2019 International Conference on Applied Technologies (ICAT), Sarajevo, Bosnia and Herzegovina.

TIBCO. (2022). *What is a Random Forest?* https://www.tibco.com/reference-center/what-is-a-random-forest

Twin, A. (2021). *Repayment: Definition and How It Works With Different Loans*. Investopedia. https://www.investopedia.com/terms/r/repayment.asp

Wang, Y., Pan, Z., Zheng, J., Qian, L., & Li, M. (2019). A Hybrid Ensemble method for Pulsar Candidate Classification. *Astrophysics and Space Science, 364*(8).

Wong, K. J. (2022). *CatBoost vs. LightGBM vs. XGBoost*. Towards Data Science. https://towardsdatascience.com/catboost-vs-lightgbm-vs-xgboost-c80f40662924

Wu, W. (2022). Machine Learning Approaches to Predict Loan Default. *Intelligent Information Management, 14*, 157-164.

Xu, J., Lu, Z., & Xie, Y. (2021). Loan default prediction of Chinese P2P market: a machine learning methodology. *Sci Rep, 11*. https://doi.org/https://doi.org/10.1038/s41598-021-98361-6

Xu, J., Lu, Z., & Xie, Y. (2021, 2021/09/21). Loan default prediction of Chinese P2P market: a machine learning methodology. *Scientific Reports, 11*(1), 18759. https://doi.org/10.1038/s41598-021-98361-6

ประกาศธนาคารแห่งประเทศไทย สนส. 4/2562 เรื่อง การกำหนดหลักเกณฑ์ วิธีการ และเงื่อนไขการประกอบธุรกิจระบบหรือเครือข่ายอิเล็กทรอนิกส์สำหรับธุรกรรมสินเชื่อระหว่างบุคคลกับบุคคล (peer to peer lending platform), (2562).

ธนาคารแห่งประเทศไทย. (2563). *Peer-to-Peer Lending* สินเชื่อออนไลน์ระหว่างบุคคลรูปแบบใหม่.

https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_05Jun2019.aspx

รพีพัฒน์ อิงคสิทธิ์. (2563). *Peer-to-Peer Lending* กู้เงินยุคใหม่ไม่ต้องง้อธนาคาร. The Momentum. https://themomentum.co/the-role-of-peer-to-peer-lending/

หลักทรัพย์บัวหลวง. (2564). *Peer-to-Peer Lending (P2P) :* ทางเลือกการขอสินเชื่อและการลงทุนรูปแบบใหม่ที่สะดวก รวดเร็ว ไม่ต้องใช้สลิปเงินเดือน โดย *Share Loan by NestiFly*. https://knowledge.bualuang.co.th/knowledge-base/peer-peer-lending-p2p/

พหุน ปณฺ ฑิโต ชีเว

ประวัติผู้เขียน

ชื่อ	นายวฤทธิ วารีศรี
วันเกิด	1 เมษายน 2541
สถานที่เกิด	อำเภอเมือง จังหวัดยโสธร
สถานที่อยู่ปัจจุบัน	310 ม.1 ต.สำราญ อ.เมือง จ.ยโสธร 35000
ตำแหน่งหน้าที่การงาน	นักวิเคราะห์ข้อมูล
สถานที่ทำงานปัจจุบัน	บริษัท ไวร์เออ แอนด์ ไวร์เลส จำกัด, 18 อาคารทรูทาวเวอร์ ถนนรัชดาภิเษก แขวงห้วยขวาง เขตห้วยขวาง กรุงเทพมหานคร 10310
ประวัติการศึกษา	พ.ศ. 2563 วิทยาศาสตรบัณฑิต (วท.บ.) สาขาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม

พหูน ปณฺ ฑิโต ชีเว