



กรอบพัฒนาวิธีหากฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้
สินเชื่อออนไลน์ระหว่างบุคคล (Peer to Peer Lending) โดยอาศัยการเลือก
คุณลักษณะ และการแบ่งช่วงข้อมูล

โดย

ภาคิน เลิศวิพุธ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2565

AN ASSOCIATION RULES MINING FRAMEWORK FOR RISK
ANALYSIS ON PEER TO PEER LENDING BASED ON FEATURE
SELECTION AND DATA DISCRETIZATION

BY

PAKIN LOETWIPHUT

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF SCIENCE AND TECHNOLOGY
THAMMASAT UNIVERSITY
ACADEMIC YEAR 2022

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

สารนิพนธ์

ของ

ภาคิน เลิศวิพุธ

เรื่อง

กรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่าง
บุคคล (Peer to Peer Lending) โดยอาศัยการเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล

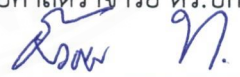
ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

เมื่อ วันที่ 19 มิถุนายน พ.ศ. 2566

ประธานกรรมการสอบสารนิพนธ์


(ผู้ช่วยศาสตราจารย์ ดร.ปกรณ์ ลีสุทธิพรชัย)

กรรมการและอาจารย์ที่ปรึกษาสารนิพนธ์


(ผู้ช่วยศาสตราจารย์ ดร.ฉันทนา ทะนานทอง)

กรรมการสอบสารนิพนธ์


(ผู้ช่วยศาสตราจารย์ ดร.ปป้อง ส่องเมือง)

กรรมการสอบสารนิพนธ์


(รองศาสตราจารย์ ดร.ทศนัย ชุ่มวัฒนะ)

คณบดี


(รองศาสตราจารย์ ดร.สุเพชร จิระจรกุล)

หัวข้อสารนิพนธ์	กรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการ วิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่าง บุคคล (Peer to Peer Lending) โดยอาศัยการเลือก คุณลักษณะ และการแบ่งช่วงข้อมูล
ชื่อผู้เขียน	ภาคิน เลิศวิพุธ
ชื่อปริญญา	วิทยาศาสตร์มหาบัณฑิต (วิทยาการคอมพิวเตอร์)
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาสารนิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.ธนธร ทะนันทอง
ปีการศึกษา	2565

บทคัดย่อ

สารนิพนธ์นี้นำเสนอกรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงด้านเครดิตในการให้สินเชื่อออนไลน์ระหว่างบุคคล (Peer to Peer Lending) เนื่องจากแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคลอยู่ในช่วงที่กำลังเติบโต และเป็นแพลตฟอร์มที่ช่วยให้บุคคลทั่วไปสามารถเข้าถึงแหล่งเงินทุนได้ง่ายขึ้นโดยไม่ต้องมีสถาบันการเงินเข้ามาเป็นตัวกลาง อย่างไรก็ตามความเสี่ยงในการผิดนัดชำระหนี้ยังคงอยู่ในเกณฑ์ที่สูง จึงเป็นความเสี่ยงของนักลงทุนที่เข้ามาปล่อยสินเชื่อผ่านแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล ทั้งนี้เพื่อเป็นตัวช่วยในการวิเคราะห์ความเสี่ยงของผู้ขอสินเชื่อ ทางผู้วิจัยได้ทำการพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงด้านเครดิตโดยใช้วิธีการคัดเลือกคุณสมบัติ (Feature Selection) ร่วมกับวิธีการแบ่งช่วงข้อมูล (Discretization) ที่เหมาะสม เพื่อหาความสัมพันธ์ของผู้ขอสินเชื่อต่อแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล โดยใช้อัลกอริทึมออพริโอรี ซึ่งหาความสัมพันธ์ที่ได้สามารถนำไปสนับสนุนการประเมินความเสี่ยงของผู้ขอสินเชื่อผ่านแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคลได้

คำสำคัญ : Peer to Peer Lending, Lending Club, การคัดเลือกคุณลักษณะ, การแบ่งช่วงข้อมูล, อัลกอริทึมออพริโอรี

Independent Study Title	AN ASSOCIATION RULES MINING FRAMEWORK FOR RISK ANALYSIS ON PEER-TO-PEER LENDING BASED ON FEATURE SELECTION AND DATA DISCRETIZATION
Author	Pakin Loetwiphut
Degree	Master of Science (Computer Science)
Department/Faculty/University	Computer Science Faculty of Science and Technology Thammasat University
Independent Study Advisor	Assistant Professor Tanatorn Tanantong, Ph.D.
Academic Year	2022

ABSTRACT

This independent study presents a framework for financial credit risk analysis in peer-to-peer (P2P) lending based on association rule mining. P2P lending is a growing sector that allows individuals to lend and borrow money without the involvement of traditional financial institutions. However, the high default risk in P2P lending poses a significant challenge to investors. To address this challenge, we develop a framework that combines feature selection and data discretization techniques to mine association rules from P2P lending data. The generated rules can be used to find patterns and relationships that show the creditworthiness and likelihood of payback of the borrower. These obtained rules can potentially support credit risk assessments in peer-to-peer lending."

Keywords: P2P Lending, Feature Selection, Data Discretization, Association Rule Mining, Apriori -Algorithm, Financial Risk- Analysis

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีจากความช่วยเหลือและความกรุณาจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ธนาธร ทะนันทอง ผู้ที่คอยให้คำปรึกษาตั้งแต่การเริ่มคิดหัวข้อสารนิพนธ์ตลอดจนวันสุดท้ายของการเรียน อีกทั้งยังสละเวลาคอยชี้แนะวิธีการเขียน และแนวทางต่างๆ ซึ่งเป็นประโยชน์ต่อการทำสารนิพนธ์ฉบับนี้ นอกจากนี้ผู้เขียนขอขอบคุณอาจารย์ทุกท่านที่เป็นผู้ถ่ายทอดวิชาความรู้และให้กำลังใจแก่ผู้เรียนตลอดระยะที่ศึกษาที่มหาวิทยาลัยธรรมศาสตร์ ที่สอนให้ผู้เรียนมีความรู้เพื่อนำไปประยุกต์ใช้ในการทำงานในอนาคต อีกทั้งยังให้ความสนับสนุนและกำลังใจให้กับผู้เรียนในเรื่องการเรียน การใช้ชีวิต และเป็นกำลังใจให้ผู้เรียนมีความสุขในชีวิตในวัยเรียนมหาวิทยาลัย

ขอกราบขอบคุณคณะกรรมการสอบเล่มสารนิพนธ์ทุกท่าน อันได้แก่ ผู้ช่วยศาสตราจารย์ ดร.ปรกรณ์ ลีสุทธิพรชัย ที่ท่านได้ให้ความกรุณาและให้เกียรติมาเป็นประธานกรรมการสอบในครั้งนี้ และ ผู้ช่วยศาสตราจารย์ ดร.ปกป้อง ส่องเมือง, รองศาสตราจารย์ ดร.ทศนัย ชุ่มวัฒนะ ที่ท่านได้ให้ความกรุณามาเป็นกรรมการสอบเล่มสารนิพนธ์ โดยคณะกรรมการทุกท่านได้มีเมตตาช่วยให้คำแนะนำ และแนวทางการแก้ไขในงานฉบับนี้จนทำให้ผู้วิจัยเข้าใจถึงปัญหาที่เกิดขึ้น และแนวทางแก้ไขที่สามารถใช้ได้จริง

นอกจากนี้ขอขอบคุณเจ้าหน้าที่สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์ที่คอยติดต่อและประสานงานให้ราบรื่น และอีกหลายท่านที่ยังไม่ได้กล่าวถึงในที่นี้สำหรับกำลังใจ แม้ว่า จะเป็นเพียงสิ่งเล็กน้อยแต่ก็สามารถทำให้ผู้ศึกษามีกำลังใจและจัดทำสารนิพนธ์ฉบับนี้สำเร็จขึ้นมา

หากผลการศึกษานี้มีข้อบกพร่องประการใด ผู้ศึกษาขอน้อมรับไว้เพื่อปรับปรุง แก้ไขในการศึกษาครั้งต่อไปผู้ศึกษาขอกราบขอบพระคุณทุกท่านเป็นอย่างยิ่ง ณ โอกาสนี้

ภาคิน เลิศวิพุธ

สารบัญ

หน้า

บทคัดย่อภาษาไทย (1)

บทคัดย่อภาษาอังกฤษ (2)

กิตติกรรมประกาศ (3)

สารบัญ (4)

สารบัญตาราง (8)

สารบัญภาพ (9)

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ 3

1.3 ขอบเขตการวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับ 4

1.5 นิยามศัพท์ที่เกี่ยวข้อง 4

บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 6

2.1 แพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending Platform) 6

2.1.1 ความหมายของแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล 6

2.1.2 ความแตกต่างของการกู้ยืมเงินผ่านธนาคาร	7
และการกู้ยืมเงินผ่านแพลตฟอร์ม P2P Lending	
2.1.3 กระบวนการขอสินเชื่อผ่านแพลตฟอร์ม	8
การให้สินเชื่อออนไลน์ระหว่างบุคคล	
2.1.4 ความเสี่ยงของแพลตฟอร์ม P2P Lending	9
ที่มีต่อผู้ใช้บริการและเสถียรภาพของระบบการเงิน	
2.1.5 Lending Club แพลตฟอร์ม P2P Lending	9
2.2 การทำวิศวกรรมคุณลักษณะ (Feature Engineering)	10
2.2.1 การแปลงคุณลักษณะ (Feature Transformation)	10
2.2.2 การเลือกคุณลักษณะ (Feature Selection)	10
2.3 การแปลงข้อมูล (Data Transformation)	11
2.3.1 การแก้ปัญหาข้อมูลไม่สมดุล (Data Balancing)	11
2.3.2 การแบ่งช่วงข้อมูล (Discretization)	12
2.4 กฎความสัมพันธ์ (Association Rule)	13
2.4.1 นิยามคำศัพท์ที่เกี่ยวข้องกับกฎความสัมพันธ์	13
2.4.2 วิธีการค้นหาความสัมพันธ์มีขั้นตอนหลัก 2 ขั้นตอน	15
2.5 งานวิจัยที่เกี่ยวข้อง	15
2.5.1 งานวิจัยที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending	15
2.5.2 งานวิจัยที่เกี่ยวข้องกับการจัดเตรียมข้อมูล	17
2.5.3 งานวิจัยที่เกี่ยวข้องกับการหาความสัมพันธ์	20
บทที่ 3 วิธีการดำเนินงานวิจัย	24
3.1 การเตรียมข้อมูล (Data Pre-Processing)	25
3.1.1 การลบคุณลักษณะ (Attribute Removal)	26
3.1.2 การกำจัดข้อมูลสูญหาย (Missing Value Removal)	26
3.2 การทำวิศวกรรมคุณลักษณะ (Feature Engineering)	26
3.2.1 การแปลงคุณลักษณะ (Feature Transformation)	27
3.2.2 การเลือกคุณลักษณะ (Feature Selection)	27

3.3 การแปลงข้อมูล (Data Transformation)	31
3.3.1. การแก้ปัญหาข้อมูลไม่สมดุล (Data Balancing)	31
3.3.2 การแบ่งช่วงข้อมูล (Discretization)	31
3.4 การสร้างกฎความสัมพันธ์ (Association rules mining)	35
3.5 การสร้างกฎสำหรับการวิเคราะห์ความเสี่ยง (Rule Construction For Risk Analysis)	36
3.6 การวิเคราะห์กฎความสัมพันธ์ด้วยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ	36
บทที่ 4 ผลการวิจัย	37
4.1 กฎความสัมพันธ์ที่ได้จากอัลกอริทึมอปรโอริ	37
4.2 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่ง (Strong Association)	38
4.2.1 ผลการวิจัย กฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 2 ไอเทมเซต	39
4.2.2 ผลการวิจัย กฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 3 ไอเทมเซต	40
4.2.3 ผลการวิจัย กฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 4 ไอเทมเซต	42
4.2.4 ผลการวิจัย กฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 5 ไอเทมเซต	44
4.3 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ	45
4.3.1 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ กรณีกำหนดไอเทมด้านขวาของกฎความสัมพันธ์เป็น สถานะปัจจุบันของสินเชื่อ	46
4.3.2 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ กรณีกำหนดไอเทมด้านขวาของกฎความสัมพันธ์เป็นคะแนนเครดิต FICO	50
4.4 การวิเคราะห์กฎความสัมพันธ์ โดยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ	52
4.4.1 การวิเคราะห์กฎความสัมพันธ์ที่แข็งแกร่ง (Strong Association)	52
4.4.2 การวิเคราะห์กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยง โดยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ	54
4.4.2.1 การวิเคราะห์ความเสี่ยงจากกฎความสัมพันธ์ที่ไอเทม ด้านขวาเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวน ตามเงื่อนไข	54
4.4.2.2 การวิเคราะห์ความเสี่ยงจากกฎความสัมพันธ์ที่ไอเทม ด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้	56

ตามเงื่อนไขเป็นระยะเวลานานกว่า 120	
4.4.2.3 การวิเคราะห์คุณลักษณะที่มีผลต่อคะแนนเครดิต FICO	57
เมื่อกำหนดไอเทมด้านขวเป็นคะแนนเครดิต FICO	
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	59
5.1 สรุปผลการวิจัย	59
5.2 ข้อเสนอแนะ	61
รายการอ้างอิง	62
ประวัติผู้เขียน	68

สารบัญตาราง

ตารางที่	หน้า
1.1 รายชื่อผู้ให้บริการระบบ Peer-to-Peer Lending Platform ในประเทศไทยที่ทดสอบภายใต้ Regulatory Sandbox ข้อมูล ณ วันที่ 15 กรกฎาคม	2
3.1 ตารางตัวอย่างข้อมูลสินเชื่อบนไลน์ระหว่างบุคคลจากแพลตฟอร์ม LendingClub (Yash, 2020)	26
3.2 ตารางแสดงรายการคุณลักษณะข้อมูลที่ใช้ในงานวิจัย	30
3.3 ตารางแสดงสัดส่วนข้อมูลสถานะเงินกู้ ก่อนและหลังการแปลงข้อมูลด้วยวิธีการ SMOTE	32
3.4 ตารางแสดงข้อมูลเชิงตัวเลขหลังจากทำการแบ่งช่วงข้อมูล	33
4.1 ตารางแสดงจำนวนกฎความสัมพันธ์ ในแต่ละช่วงของค่าสนับสนุน และค่าความเชื่อมั่น	40
4.2 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 2 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด	41
4.3 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 3 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด	43
4.4 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 4 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด	45
4.5 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 5 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด	47
4.6 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ	49
4.7 ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นคะแนนเครดิต FICO	52

สารบัญภาพ

ภาพที่	หน้า
2.1 มูลค่าของสินเชื่อ P2P Lending ในประเทศสหรัฐอเมริกาของ แพลตฟอร์ม Prosper และ Lending Club ระหว่างปี 2006 – 2016	7
2.2 ความแตกต่างของการกู้ยืมเงินแบบดั้งเดิมผ่านธนาคาร และ การกู้ยืมเงินผ่านแพลตฟอร์ม P2P Lending	8
3.1 กรอบพัฒนาวิธีหาคุณสมบัติความสัมพันธ์โดยอาศัยการเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล	25
3.2 แสดงจำนวนคุณลักษณะต่อความแม่นยำในการทำนายสถานะเงินกู้ ด้วยแบบจำลอง LightGBM ด้วยวิธีการ Recursive Feature Elimination with 5-Cross-Validation	29
3.3 ค่าความสำคัญของคุณลักษณะที่ส่งผลต่อการทำนายสถานะเงินกู้ ผ่านแบบจำลอง LightGBM	30

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ธุรกรรมทางการเงินมีส่วนเกี่ยวข้องกับการใช้ชีวิตประจำวันของมนุษย์มาอย่างยาวนาน ไม่ที่จะเป็นการฝากเงิน/ออมเงิน การกู้ยืม การโอนเงิน และการชำระเงินต่างๆ เป็นต้น ซึ่งธุรกรรมส่วนใหญ่จำเป็นต้องมีตัวกลางอย่างธนาคารหรือสถาบันการเงินในการเชื่อมต่อธุรกรรมให้เกิดขึ้นได้ การที่ผู้คนบางกลุ่มไม่สามารถเข้าถึงบริการด้านการเงินได้เป็นเหตุผลหนึ่งที่ทำให้เกิดความเหลื่อมล้ำทางด้านรายได้ จากข้อมูลรายงานผลการสำรวจการเข้าถึงบริการทางการเงินภาคครัวเรือนปี 2563 (ธนาคารแห่งประเทศไทย, 2563) แสดงให้เห็นว่าภาคครัวเรือนไม่สามารถเข้าถึงบริการด้านสินเชื่อได้คิดเป็น 5.0% โดยมีสาเหตุสำคัญ 3 ประการ มาจาก 1.ฐานะการเงิน/รายได้ไม่เพียงพอซึ่งที่สัดส่วน 35.6% 2.ไม่รู้จัก/ไม่มีความรู้ความเข้าใจ 18.6% และ 3.ไม่กล้าไปติดต่อ/คิดว่าจะถูกปฏิเสธ 13.5% เป็นเหตุให้ผู้คนบางส่วนต้องใช้บริการเงินกู้ยืมแบบซึ่งจะส่งผลให้มีปัญหาเกิดขึ้นในอนาคต การให้สินเชื่อออนไลน์ระหว่างบุคคลเป็นทางเลือกอีกรูปแบบหนึ่งของการให้กู้ยืม โดยให้บุคคลสามารถกู้และให้กู้ระหว่างกันได้โดยตรง โดยสถาบันการเงินไม่ได้มีส่วนเกี่ยวข้อง โดยทั้งผู้กู้และผู้ให้กู้มีปฏิสัมพันธ์กันผ่านแพลตฟอร์มออนไลน์ที่เรียกว่า แพลตฟอร์ม P2P Lending ซึ่งทำหน้าที่เป็นตัวกลาง ในการรวบรวมเงินทุน และกระจายเงินทุนไปยังผู้ให้กู้ จากข้อมูลของ (Acumen Research and Consulting ,2023) แสดงให้เห็นว่าการให้สินเชื่อออนไลน์ระหว่างบุคคลเติบโตอย่างรวดเร็วในช่วงไม่กี่ปีที่ผ่านมา โดยในปี พ.ศ.2564 มีมูลค่าตลาดรวม 82.3 พันล้านเหรียญสหรัฐ และคาดว่าจะสูงถึง 804.2 พันล้านเหรียญสหรัฐภายในปี พ.ศ.2573 ซึ่งเหตุผลสนับสนุนการคาดการณ์การเติบโตของแพลตฟอร์ม P2P Lending นั้น ส่วนหนึ่งมาจากเทคโนโลยี และอินเทอร์เน็ตที่ทำให้ผู้กู้และผู้ให้กู้สามารถเชื่อมต่อและทำธุรกรรมระหว่างกันได้ง่ายขึ้น

สำหรับสถานการณ์แพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคลในประเทศไทยนั้น ในปี พ.ศ. 2563 ธนาคารแห่งประเทศไทยได้กำหนดให้ผู้ประกอบธุรกิจ P2P Lending platform เข้าทดสอบการให้บริการในวงจำกัดภายใต้ Regulatory Sandbox ซึ่งเป็นแนวทางการทดสอบและพัฒนานวัตกรรมที่นำเทคโนโลยีใหม่มาสนับสนุนการให้บริการทางการเงิน โดยเปิดให้ผู้ให้บริการทางการเงินทดสอบการให้บริการด้วยเทคโนโลยีใหม่ในวงจำกัด เพื่อส่งเสริมให้เกิดนวัตกรรมทางการเงิน ขณะที่มีความเสี่ยงในการดูแลความเสี่ยง และคุ้มครองผู้ใช้บริการที่เหมาะสม จนประสบความสำเร็จก่อน เพื่อให้มั่นใจว่ามีการบริหารความเสี่ยงและการดูแลผู้ใช้บริการอย่างเหมาะสม จึงจะ

สามารถยื่นขอใบอนุญาตจากกระทรวงการคลังเพื่อประกอบธุรกิจในวงกว้าง (ธนาคารแห่งประเทศไทย, 2565) ตารางที่ 1.1 แสดงรายชื่อผู้ให้บริการระบบ Peer-to-Peer Lending ในประเทศไทย ที่ทดสอบภายใต้ Regulatory Sandbox

ตารางที่ 1.1

รายชื่อผู้ให้บริการระบบ Peer-to-Peer Lending Platform ในประเทศไทยที่ทดสอบภายใต้ Regulatory Sandbox ข้อมูล ณ วันที่ 15 กรกฎาคม 2565

ชื่อบริษัท	วันที่ได้รับอนุมัติ เข้าทดสอบ	สถานะของการทดสอบ
บริษัท ดีพสพาร์กส์ เพียร์ เลนดิง จำกัด	4 มิ.ย. 2563	อยู่ระหว่างทดสอบ
บริษัท เนสท์ดีฟลาย จำกัด	28 ก.ย. 2563	ออกจาก Regulatory Sandbox ได้รับใบอนุญาตจาก กระทรวงการคลัง เมื่อ 22 เม.ย. 65
บริษัท เพียร์ พาวเวอร์ แพลตฟอร์ม จำกัด	28 ก.ย. 2563	ออกจาก Regulatory Sandbox ยุติการทดสอบ เมื่อ 14 ก.ค. 65
บริษัท ได้เงิน ดอทคอม จำกัด	11 ก.พ. 2565	อยู่ระหว่างทดสอบ

อย่างไรก็ตามการทำธุรกรรมสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีความเสี่ยงในหลายด้าน อาทิเช่น ความเสี่ยงที่ผู้กู้จะก่อหนี้เกินความสามารถในการชำระหนี้ของตน หรือความเสี่ยงที่ผู้ให้กู้อาจไม่ได้รับชำระหนี้คืนตามสัญญา เนื่องจากการให้สินเชื่อไม่ใช่การฝากเงิน ผู้ให้กู้จึงไม่ได้รับการคุ้มครอง หากผู้กู้เกิดการผิดนัดชำระ รวมไปถึงความเสี่ยงที่ผู้ให้กู้จะขาดสภาพคล่องเนื่องจากไม่สามารถยกเลิกการให้สินเชื่อหรือเรียกให้ผู้กู้ชำระหนี้ก่อนครบกำหนดสัญญาได้ และที่สำคัญแพลตฟอร์ม P2P Lending เป็นธุรกรรมการให้สินเชื่อผ่านแพลตฟอร์มออนไลน์ที่สามารถเข้าถึงผู้ใช้บริการในวงกว้าง สัญญาสินเชื่อสามารถเกิดขึ้นได้อย่างรวดเร็วโดยที่ผู้กู้และผู้ให้กู้ไม่จำเป็นต้องรู้จักกัน จึงมีโอกาสเกิดการหลอกลวงทั้งจากแพลตฟอร์มและผู้กู้ ดังเช่นกรณีศึกษาในต่างประเทศที่แพลตฟอร์มมีการให้ข้อมูลที่บิดเบือนจากความเป็นจริง ผู้บริหารแพลตฟอร์มทุจริต แพลตฟอร์มหรือผู้กู้สร้างโครงการขอกู้ปลอม เป็นต้น (ธนาคารแห่งประเทศไทย, 2565)

ความเสี่ยงทางเครดิตเป็นหนึ่งในความเสี่ยงหลักที่เกี่ยวข้องกับการให้กู้เงินแบบ P2P ซึ่งหมายถึงความเสี่ยงที่ผู้กู้จะไม่สามารถชำระเงินกู้หรือไม่สามารถชำระตามข้อกำหนดที่ตกลงไว้ได้ ซึ่งโดยทั่วไปอัตราผู้ไม่ชำระหนี้ที่เกิดขึ้นในการให้กู้เงินแบบ P2P จะสูงกว่าอัตราผู้ไม่ชำระหนี้ของสินเชื่อ

ธนาคาร ซึ่งเป็นเพราะแพลตฟอร์มการให้กู้เงินแบบ P2P มักมีเกณฑ์การให้กู้ที่ไม่เข้มงวดเท่ากับสถาบันการเงิน ซึ่งส่งผลให้เกิดความเสี่ยงที่สูงขึ้นสำหรับผู้กู้ นอกจากนี้แล้วแพลตฟอร์มการให้กู้เงินแบบ P2P อาจไม่มีความสามารถในการวิเคราะห์เครดิตและติดตามการใช้เงินที่เท่าเดิมกับสถาบันการเงินแบบดั้งเดิม ในข้อมูลสาธารณะจากแพลตฟอร์มเลนดิงคลับ (Lending Club Platform) ซึ่งเป็นแพลตฟอร์ม P2P Lending ที่ใหญ่ที่สุดในสหรัฐอเมริกา เมื่อดูข้อมูลจากไตรมาสที่ 3 ปี 2018 ถึงไตรมาสที่ 2 ปี 2019 อัตราผู้ไม่ชำระหนี้ยังคงสูงกว่า 14% (Ko, P.-C et al., 2022)

ในครั้งนี้นำเพื่อเป็นการวิเคราะห์ความเสี่ยงต่างๆที่อาจเกิดขึ้นจากการทำธุรกรรมทางการเงินผ่านแพลตฟอร์ม P2P Lending ทางผู้วิจัยจึงประยุกต์ใช้อัลกอริทึมออริโอริเพื่อหาความสัมพันธ์ของคุณลักษณะที่ส่งผลต่อการชำระหนี้ของผู้ขอสินเชื่อ พร้อมทั้งพัฒนาหาความสัมพันธ์ที่เหมาะสมด้วยวิธีการคัดเลือกคุณลักษณะแบบต่างๆ รวมไปถึงการแบ่งช่วงข้อมูลที่เหมาะสมโดยใช้ข้อมูลของแพลตฟอร์มเลนดิงคลับซึ่งเป็นแพลตฟอร์ม P2P Lending ขนาดใหญ่ในประเทศสหรัฐอเมริกา มาประยุกต์ใช้ในการวิเคราะห์ในครั้งนี้ ซึ่งผลที่ได้จากการศึกษาครั้งนี้จะสามารถเป็นแนวทางให้ผู้ประกอบการในธุรกิจทางการเงินสามารถวิเคราะห์ความเสี่ยงจากการให้สินเชื่อได้ หรือผู้ที่สนใจนำผลวิจัยไปปรับปรุงและพัฒนาได้ต่อไป

1.2 วัตถุประสงค์

1. เพื่อศึกษาและพัฒนากรอบวิธีการหาความสัมพันธ์เชื่อที่ได้จากอัลกอริทึมออริโอริด้วยวิธีการคัดเลือกคุณลักษณะ (Feature Selection) และการแบ่งช่วงข้อมูล (Discretization)
2. เพื่อค้นหาความสัมพันธ์ที่มีความน่าสนใจหรือหาความสัมพันธ์ที่มีแข็งแกร่ง (Strong Association) และนำผลลัพธ์ที่ได้ไปวิเคราะห์ความสัมพันธ์ของข้อมูลผู้ขอสินเชื่อ รวมไปถึงประยุกต์ใช้ในการวิเคราะห์ความเสี่ยงต่อการชำระหนี้ของผู้ขอสินเชื่อได้

1.3 ขอบเขตการวิจัย

ผู้วิจัยได้กำหนดขอบเขตการวิจัยไว้ ดังนี้

1. งานวิจัยนี้ศึกษาเรื่อง “กรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล (Peer to Peer Lending) โดยอาศัยการเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล” มีขอบเขตการศึกษาหาความสัมพันธ์ด้วยข้อมูลจากแพลตฟอร์ม P2P Lending ที่มีชื่อว่า “เลนดิงคลับ” ตั้งแต่ปี ค.ศ.2007-2020 ซึ่งมีจำนวนผู้ขอสินเชื่อทั้งหมด 2,925,493 คน และจำนวนคุณลักษณะผู้กู้ทั้งหมด 141 คุณลักษณะ

2. งานวิจัยนี้มุ่งเน้นการศึกษากลุ่มผู้ขอสินเชื่อจำนวน 2 กลุ่ม ซึ่งแบ่งแยกตามสถานะชำระหนี้ เพื่อใช้ในการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อดังต่อไปนี้

- ลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid)
- กลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน (Charge Off)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่ได้รับจากงานวิจัยนี้

1. ได้ศึกษาและพัฒนาวิธีการหาความสัมพันธ์ด้วยการประยุกต์ใช้วิธีการคัดเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล
2. ได้ค้นพบความสัมพันธ์ที่มีความน่าสนใจหรือความสัมพันธ์ที่มีแข็งแกร่ง และสามารถนำมาวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล เพื่อลดความผิดพลาดในการตัดสินใจเกี่ยวกับการให้สินเชื่อและการจัดการความเสี่ยงได้

1.5 นิยามศัพท์ที่เกี่ยวข้อง

1. สินเชื่อส่วนบุคคล (Personal Loan) หมายถึง การให้สินเชื่อหรือกู้เงินให้แก่บุคคลเพื่อใช้ในการอำนวยความสะดวกในเรื่องรายจ่ายส่วนบุคคล โดยสินเชื่อส่วนบุคคลมักถูกใช้สำหรับการซื้อสินค้าหรือบริการที่ไม่มีความเกี่ยวข้องกับการลงทุน หรือการชำระหนี้ที่มีต้นทุนสูง เช่น การซื้อรถยนต์หรือบ้าน เงินทุนการศึกษา การจัดสรรงบประมาณสำหรับการเดินทางหรือเพื่อการสำรองเงินฉุกเฉิน สินเชื่อส่วนบุคคลมักมีอัตราดอกเบี้ยและเงื่อนไขการชำระเงินที่กำหนดไว้ล่วงหน้า และผู้ขอสินเชื่อจำเป็นต้องชำระหนี้ตามเงื่อนไขที่กำหนดไว้ในสัญญาสินเชื่อ

2. แพลตฟอร์มสินเชื่อออนไลน์ระหว่างบุคคล (Peer-to-Peer Lending Platform) หมายถึง แพลตฟอร์มที่ให้บริการการให้สินเชื่อระหว่างบุคคล โดยใช้ระบบออนไลน์เพื่อเชื่อมต่อผู้กู้และผู้ให้กู้โดยตรงโดยไม่ผ่านสถาบันการเงิน

3. ผู้ขอสินเชื่อหรือผู้กู้ (Borrower) หมายถึง บุคคลที่ต้องการขอสินเชื่อหรือกู้เงินจากผู้ให้สินเชื่อ เพื่อใช้ในการดำเนินกิจกรรมหรือการลงทุนที่ต้องการเงิน

4. ผู้ให้สินเชื่อหรือผู้ให้กู้ (Lender) หมายถึง บุคคลหรือองค์กรที่มีส่วนร่วมในการให้สินเชื่อหรือกู้เงินให้แก่ผู้ขอสินเชื่อ โดยผู้ให้สินเชื่อสามารถเป็นบุคคลทั่วไปหรือนักลงทุนที่สนใจในการลงทุนในสินเชื่อ

5. ความเสี่ยงจากการให้สินเชื่อ (Lending Risk) หมายถึง ความเป็นไปได้ที่ผู้กู้จะไม่สามารถชำระเงินตามเงื่อนไขที่กำหนดไว้ในสัญญาสินเชื่อ ซึ่งอาจเกิดจากปัจจัยต่างๆ เช่น สภาพการเงินของผู้กู้ ประวัติการชำระหนี้ก่อนหน้า หรือสถานการณ์เศรษฐกิจที่มีผลกระทบต่อผู้กู้ เป็นต้น

6. ผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ หมายถึง ผู้ที่มีประสบการณ์ในการประเมินความเสี่ยงและตัดสินใจการอนุมัติสินเชื่อให้กับลูกค้า



บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาการกรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล (Peer to Peer Lending) โดยอาศัยการเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล ผู้วิจัยได้ทำการสืบค้นข้อมูลจากแหล่งต่างๆ เพื่อนำมาใช้เป็นแนวทางในการศึกษา โดยมีสาระที่เกี่ยวข้อง ดังนี้

- 2.1 แพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending Platform)
- 2.2 การทำวิศวกรรมคุณลักษณะ (Feature Engineering)
- 2.3 การแปลงข้อมูล (Data Transformation)
- 2.4 กฎความสัมพันธ์ (Association Rule)
- 2.5 งานวิจัยที่เกี่ยวข้อง

2.1 แพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending Platform)

2.1.1 ความหมายของ Peer to Peer Lending Platform

P2P: Peer to Peer Lending Platform คือ ธุรกิจการกู้ยืมเงินระหว่างบุคคลผ่านระบบหรือเครือข่ายอิเล็กทรอนิกส์ โดยไม่จำเป็นต้องผ่านตัวกลางอย่างธนาคารหรือสถาบันการเงิน ซึ่งหมายความว่า บุคคลทั่วไปจะสามารถกู้ยืมเงินกันเองได้โดยตรง โดยมีแพลตฟอร์ม P2P Lending ทำหน้าที่เสมือนเป็นตลาดสินเชื่อออนไลน์ที่เชื่อมต่อระหว่างผู้กู้และผู้ให้กู้ให้มาเจอกัน ดังนั้นแนวคิดหลักของ Peer-to-Peer Lending ซึ่งคือ การทำให้บุคคลทั่วไปสามารถเข้าถึงแหล่งเงินกู้ได้สะดวกขึ้น ในขณะที่ผู้ลงทุนหรือผู้ปล่อยสินเชื่อก็มีทางเลือกในการลงทุนที่มากขึ้นกว่าเดิมเช่นกัน (ธนาคารแห่งประเทศไทย, 2565) แนวโน้มทางการเงิน P2P Lending เกิดขึ้นครั้งแรกในปี 2005 ในประเทศอังกฤษ โดยบริษัทที่มีชื่อว่า Zopa และได้ขยายไปกลุ่มประเทศอื่น ๆ เช่น ประเทศในสหภาพยุโรป สหรัฐอเมริกา และประเทศจีน เป็นต้น (พิมพ์ธัญญา, 2562) แม้ว่าในช่วงแรกแพลตฟอร์ม P2P Lending ยังไม่ได้รับความนิยมเท่าที่ควร จากข้อมูลมูลค่าสินเชื่อย่อยในประเทศสหรัฐอเมริกา ปี ค.ศ.2015 มูลค่าสินเชื่อจากแพลตฟอร์ม P2P Lending คิดเป็นเพียง 0.7% ของมูลค่าสินเชื่อย่อยทั้งหมด แต่ในช่วงเวลาต่อมาได้มีการเติบโตอย่างทวีคูณ ซึ่งเป็นผลมาจากเทคโนโลยีทางการเงิน (Financial Technology) ที่เติบโตขึ้นอย่างมาก ในปี 2559 มีจำนวนแพลตฟอร์ม P2P Lending กระจายตัวอยู่ทั่วโลกมากกว่า 3,000 แพลตฟอร์ม และมีมูลค่าของสินเชื่อในระบบสูงถึง 86,333 ล้านดอลลาร์ (ภาพที่ 2.1) (Havrylchyk et al., 2016)

ภาพที่ 2.1

มูลค่าของสินเชื่อ P2P Lending ในประเทศสหรัฐอเมริกาของแพลตฟอร์ม Prosper และ Lending Club ระหว่างปี 2006 – 2016



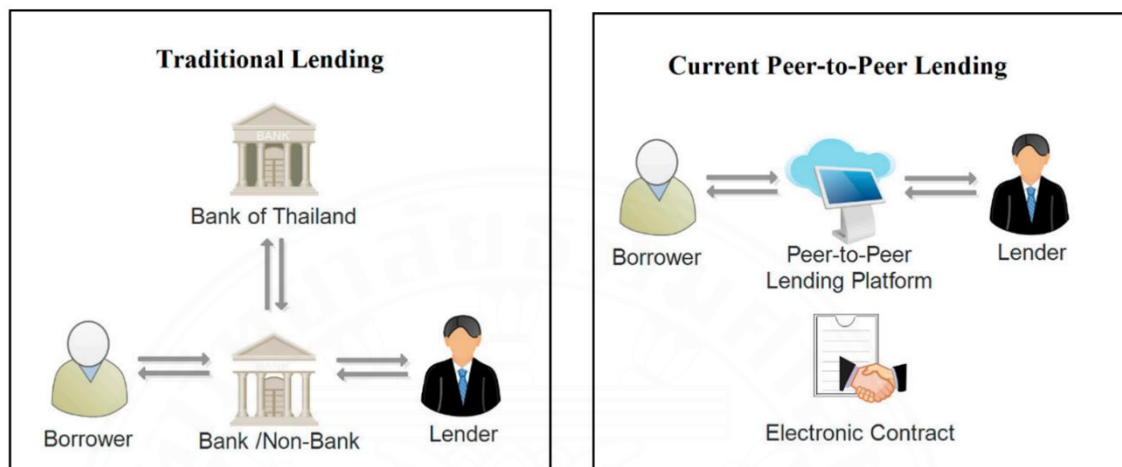
หมายเหตุ. จาก “What drives the expansion of the peer-to-peer lending?” (p. 2), by Havrylchyk et al., (2016)

2.1.2 ความแตกต่างของการกู้ยืมเงินผ่านธนาคาร และการกู้ยืมเงินผ่านแพลตฟอร์ม P2P Lending

การกู้ยืมเงินแบบดั้งเดิม เป็นรูปแบบที่เราต้องไปทำเรื่องขอสินเชื่อผ่านธนาคารหรือสถาบันการเงิน ซึ่งทำหน้าที่เป็นตัวกลางในการจัดหาแหล่งเงินทุน แน่นอนว่าต้องผ่านกระบวนการหลากหลายขั้นตอนและใช้เวลานาน ในขณะที่ P2P Lending จะแตกต่างจากการกู้ผ่านธนาคารเดิมตรงที่กระบวนการเหล่านั้นจะถูกยกไปไว้บนแพลตฟอร์มออนไลน์เกือบทั้งหมดและทำหน้าที่เสมือนตลาดสินเชื่อออนไลน์ที่ผู้กู้และผู้ให้กู้สามารถขอสินเชื่อกันได้โดยตรง (ภาพที่ 2.2) (พรชนก, 2562)

ภาพที่ 2.2

ความแตกต่างของการกู้ยืมเงินแบบดั้งเดิมผ่านธนาคาร (Traditional Bank) และ การกู้ยืมเงินผ่านแพลตฟอร์มการให้กู้ยืมระหว่างบุคคล (Peer-to-peer Lending)



หมายเหตุ. จาก “Success Factors Influencing Peer-to-Peer Lending to Support Financial Innovation.” (p. 2) by Chulawate, N. and Kiattisin, S. (2023)

2.1.3 กระบวนการขอสินเชื่อผ่านแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล

(Havrylchyk et al., 2016) ได้กล่าวถึงกระบวนการขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวิธีการดังนี้

ผู้กู้ และผู้ให้กู้ลงทะเบียนผ่านแพลตฟอร์ม P2P Lending โดยทำการกรอกข้อมูลส่วนตัวต่างๆ อาทิเช่น ชื่อ หมายเลขบัตรประจำตัวประชาชน ข้อมูลเงินฝากในสถาบันการเงินอื่น เป็นต้น โดยข้อมูลเหล่านี้จะถูกนำมาสร้างแบบจำลองคะแนนเครดิต (Credit Scoring)

1. ผู้กู้ระบุจำนวนวงเงินที่ต้องการกู้ และอัตราดอกเบี้ยสูงสุดที่ระบบเสนอให้ รวมไปถึงข้อมูลอื่น ๆ เช่น ระยะเวลาการชำระเงินกู้ , วัตถุประสงค์ของการกู้

2. ผู้ให้กู้ (ผู้ลงทุน) นำเงินลงทุนเข้าแพลตฟอร์ม และเลือกรูปแบบการลงทุน ซึ่งในปัจจุบันมีการลงทุน 2 รูปแบบ คือ 1.ผู้ให้กู้เลือกผู้กู้บนแพลตฟอร์มที่ต้องการและปล่อยเงินกู้ให้กับผู้กู้รายนั้น ๆ และอีกรูปแบบหนึ่งผู้ให้กู้นำเงินลงทุนให้กับกองทุนของบริษัทที่ทำธุรกิจ P2P Lending โดยแพลตฟอร์มจะทำหน้าที่จัดหาผู้กู้ที่เหมาะสมให้ ซึ่งในรูปแบบนี้ผู้ให้กู้จะไม่ทราบข้อมูลของผู้กู้

3. เมื่อรูปแบบการกู้ของผู้กู้และผู้ให้กู้ตรงกัน ข้อมูลจะถูกส่งไปยังตัวกลางเพื่อปล่อยสินเชื่อ ซึ่งในขั้นตอนนี้จะมีการตรวจสอบเอกสารเพิ่มเติมเพื่อตรวจสอบความน่าเชื่อถือของข้อมูล และเมื่อมีการตรวจสอบข้อมูลเรียบร้อยแล้วจะมีการโอนเงินเข้าสู่บัญชีของผู้กู้

2.1.4 ความเสี่ยงของแพลตฟอร์ม P2P Lending ที่มีต่อผู้ใช้บริการและเสถียรภาพของระบบการเงิน

ความเสี่ยงจากการดำเนินการของแพลตฟอร์ม Peer-to-Peer Lending สามารถจัดกลุ่มออกเป็น 5 ความเสี่ยงสำคัญ ซึ่งเป็นผลสำรวจในประเทศจีน สหรัฐอเมริกา แคนาดา สหราชอาณาจักร และสหภาพยุโรป (พรชนก ,2562) ประกอบด้วย

1. ภัยคุกคามทางไซเบอร์ (Cyber Risk) หรือความเสี่ยงที่จะถูกโจมตีทางอิเล็กทรอนิกส์
2. ธุรกิจแพลตฟอร์มล้ม (Platforms' Collapse) จากการขาดความน่าเชื่อถือซึ่งอาจเกิดได้จากหลายประการ อาทิ ความไม่โปร่งใสของแพลตฟอร์มจากการหลีกเลี่ยงการเปิดเผยข้อมูล ทำให้นักลงทุนถอนเงินลงทุนออกส่งผลให้แพลตฟอร์มขาดสภาพคล่องอย่างหนัก และนำไปสู่การปิดตัวของแพลตฟอร์มในที่สุด
3. การฉ้อโกงจากบุคคลที่สาม (Fraud) ด้วยวิธีการต่างๆ อาทิ กู้ยืมเงินผ่านช่องทางออนไลน์ด้วยตัวตนที่ไม่มีจริง เป็นต้น
4. การผิดนัดชำระหนี้ (Default) ซึ่งเป็นความเสี่ยงที่ผู้กู้ยืมสินเชื่อจะผิดนัดชำระหนี้เป็นจำนวนมากโดยเฉพาะในช่วงที่ประสบกับวิกฤตเศรษฐกิจ
5. ความเสี่ยงที่นักลงทุนรายใหญ่ จะเข้ามาแย่งการลงทุนของนักลงทุนรายย่อย (Crowding Out Of Retail Investor) ทำให้ผลตอบแทนโดยรวมลดลง

2.1.5 Lending Club แพลตฟอร์ม P2P Lending

Lending Club คือ แพลตฟอร์มธุรกรรมสินเชื่อระหว่างบุคคลกับบุคคลที่ใหญ่ที่สุดในโลกจากประเทศอเมริกา เปิดตัวในปี ค.ศ.2007 ซึ่งให้บริการด้านสินเชื่อหลายประเภท เช่น สินเชื่อส่วนบุคคล สินเชื่อบ้าน สินเชื่อรถยนต์ เป็นต้น อีกทั้งการกู้ยืมสินเชื่อผ่านแพลตฟอร์ม Lending Club ยังทำได้ทั้งกู้เดี่ยวและกู้รวมอีกด้วย ในส่วนของอัตราดอกเบี้ยต่อปีรวมค่าธรรมเนียม อยู่ระหว่างร้อยละ 6.46 - 30.99 (ข้อมูลอัตราดอกเบี้ย ณ เดือนกุมภาพันธ์ 2563) ค่าธรรมเนียมต่อปีอยู่ที่ ร้อยละ 1.0 - 6.0 วงเงินกู้สินเชื่อส่วนบุคคล 1,000 - 40,000 ดอลลาร์สหรัฐ โดยในปี ค.ศ.2020 สินเชื่อทั้งหมดที่ได้รับเงินผ่านแพลตฟอร์มของ Lending Club มีมูลค่าสูงถึง 53,722 ล้านดอลลาร์สหรัฐ ซึ่งคิดเป็นส่วนแบ่งตลาด ร้อยละ 71.9 ของประเทศอเมริกา (จิตติมา, 2563)

2.2 การทำวิศวกรรมคุณลักษณะ (Feature Engineering)

วิศวกรรมคุณลักษณะ (Feature Engineering) เป็นองค์ประกอบที่สำคัญอย่างหนึ่งของขั้นตอนการเตรียมข้อมูล โดยจุดมุ่งหมายของการทำวิศวกรรมคุณลักษณะนั้น คือการสร้างคุณลักษณะที่มีคุณภาพขึ้นมา เพื่อให้แบบจำลองมีประสิทธิภาพในการทำนายที่ดีขึ้น และมีข้อผิดพลาดน้อยลง อีกทั้งยังสามารถลดเวลาในประมวลผล และการสำรวจคุณลักษณะต่างๆ ได้อย่างมาก (Mohamad และคณะ, 2020) ประเภทของ Feature Engineering ประกอบด้วย 2 รูปแบบ ได้แก่ การแปลงคุณลักษณะ และการคัดเลือกคุณลักษณะ (Zheng and Casari, 2018)

2.2.1 การแปลงคุณลักษณะ (Feature Transformation)

เป็นกระบวนการสร้างคุณลักษณะใหม่โดยใช้จำนวนน้อยกว่าคุณลักษณะเดิม ซึ่งยังคงคุณลักษณะของข้อมูลที่เป็นประโยชน์ วิธีการสร้างคุณลักษณะใหม่อาจแตกต่างกัน เช่นการรวมเชิงเส้น (Linear Combination) หรือแมปปิงแบบไม่เชิงเส้นชนิดอื่น ๆ ในงานวิจัยฉบับนี้ใช้ค่าเฉลี่ยระหว่าง 2 คุณลักษณะ ซึ่งเป็นคุณลักษณะที่กำหนดค่าขอบบนของช่วงข้อมูล และขอบล่างของช่วงข้อมูล เพื่อเป็นการสร้างคุณลักษณะใหม่ที่เป็นค่ากลางของข้อมูลสองกลุ่มนี้ จึงได้ใช้ค่าเฉลี่ยเป็นตัวแทนของข้อมูล และกำหนดคุณลักษณะใหม่ขึ้น

2.2.2 การคัดเลือกคุณลักษณะ (Feature selection)

(Novakovic, 2016) กล่าวว่า การคัดเลือกคุณลักษณะเป็นการคัดกรองคุณลักษณะที่ไม่เกี่ยวข้องหรือซ้ำซ้อนในชุดข้อมูล โดยการเลือกคุณลักษณะเหล่านี้จะเน้นที่ลักษณะเด่นในชุดข้อมูลและละทิ้งคุณลักษณะที่ไม่สำคัญ รูปแบบของแต่ละคุณลักษณะจะไม่เปลี่ยนแปลง เป็นขั้นตอนที่สำคัญในการเตรียมข้อมูลก่อนการทำเหมืองข้อมูล เป็นเทคนิคการลดขนาดมิติ ลดจำนวนคุณลักษณะ และคัดเลือกคุณลักษณะที่มีความสำคัญต่อการพยากรณ์เหมืองข้อมูล และอาจกล่าวได้ว่าประสิทธิภาพของการจำแนกประเภทเหมืองข้อมูลขึ้นอยู่กับลักษณะ (Feature) ที่นำมาใช้ ซึ่งนอกจากจะช่วยลดเวลาในการสร้างตัวแบบพยากรณ์ให้เร็วขึ้นแล้ว ยังช่วยลดคุณลักษณะที่ไม่จำเป็นต่อการสร้างตัวแบบพยากรณ์ได้อีกด้วย (พัชรนิกันต์, 2561) ซึ่งสามารถแบ่งออกได้เป็น 3 วิธี ได้แก่ วิธีฟิลเตอร์ (Filter Method) ,วิธีแรปเปอร์ (Wrapper Method) และวิธีการฝังตัว (Embed Method) (อัศวิน สุรวัชโยธิน และ วรภัทร ไพรีเกรง, 2564)

1. วิธีการฝังตัว (Embedded methods) เป็นวิธีการคัดเลือกคุณลักษณะเป็นส่วนหนึ่งที่อยู่ในกระบวนการจำแนกประเภทด้วย เช่น กระบวนการสร้างต้นไม้การตัดสินใจ (Decision

Tree) โดยจะทำการคัดเลือกเซตย่อยที่มีคุณลักษณะที่เหมาะสมในแต่ละขั้นตอนวิธี (อัศวิน สุรวัช โยธิน และ วรภัทร ไพรีเกรง, 2564)

2. วิธีฟิลเตอร์ (Filter-based) เป็นวิธีที่ใช้การประเมินความมีประโยชน์ ของแต่ละลักษณะเฉพาะต่อการจำแนกข้อมูล โดยเรียกว่าค่าคะแนนลักษณะเฉพาะ (Feature Score) โดยขั้นตอนนี้จะเลือกลักษณะเฉพาะที่มีค่าคะแนนสูงมากกว่า ลักษณะเฉพาะที่มีค่าคะแนนต่ำ

3. วิธีแรปเปอร์ (Wrapper-Based) จะใช้อัลกอริทึมการค้นหา (Searching Algorithm) และอัลกอริทึมการเรียนรู้ (Learning Algorithm) ที่เตรียมไว้แล้วสำหรับกลุ่มของลักษณะเฉพาะที่เหมาะสมกับอัลกอริทึมเรียนรู้นั้น ทำให้วิธีนี้มีความแม่นยำในการจำแนกข้อมูลได้ดีกว่าวิธีฟิลเตอร์ (Pramokchon, 2017) ได้กล่าวว่า วิธีการคัดเลือกแบบแรปเปอร์ (Wrapper-Based) เป็นวิธีการคัดเลือกคุณลักษณะที่มีความแม่นยำในการจำแนกข้อมูลสำหรับปัญหาที่เฉพาะเจาะจง

ในงานวิจัยฉบับนี้ จะใช้วิธีการ Recursive Feature Elimination with 5-cross Validation (RFECV) โดย (Bahl et al., 2016) ได้อธิบายการคัดเลือกคุณลักษณะด้วยวิธีการ RFECV นั้นเป็นวิธีการคัดเลือกคุณลักษณะในกลุ่มวิธีการคัดเลือกแบบห่อหุ้ม (Wrapper-Based) โดยจะทำการลบคุณลักษณะที่ไม่จำเป็น หรือคุณลักษณะที่มีผลต่อการทำนายผลที่น้อยที่สุด ซึ่งการลบคุณลักษณะเหล่านี้ไม่ส่งผลกระทบต่อความผิดพลาดในการทำนายของโมเดล และเก็บคุณลักษณะที่เป็นอิสระกัน เพื่อให้ได้คุณลักษณะที่มีประสิทธิภาพสูงสุด กระบวนการทำงานของการคัดเลือกคุณลักษณะด้วยวิธีการ RFECV เป็นการคัดเลือกคุณลักษณะที่เริ่มจากการใช้คุณลักษณะทั้งหมดก่อนและตัดคุณลักษณะที่ไม่สำคัญทิ้งไปทีละคุณลักษณะ (Backward Feature Elimination) หลังจากนั้นจะลบคุณลักษณะที่มีค่าความสำคัญน้อยที่สุด และสร้างโมเดลและคำนวณความสำคัญของคุณลักษณะเพื่อทำนายความแม่นยำของโมเดลวนซ้ำไปเรื่อย ๆ เพื่อหาจำนวนคุณลักษณะที่เหมาะสม และคัดเลือกคุณลักษณะที่มีประสิทธิภาพมากที่สุด

2.3 การแปลงข้อมูล (Data Transformation)

2.3.1 การแก้ปัญหาข้อมูลไม่สมดุล (Data Balancing)

การแก้ปัญหาข้อมูลไม่สมดุล (Data Balancing) เกิดขึ้นเมื่อจำนวนข้อมูลของแต่ละคลาสแตกต่างกันมากทำให้ผลลัพธ์จากการจำแนกข้อมูลมีความโน้มเอียงไปทางข้อมูลที่มีจำนวนมาก ส่งผลให้ตัวแบบเรียนรู้มีความผิดพลาดได้ (อติวิญญ์ และคณะ, 2564) จึงต้องมีการแก้ปัญหาเพื่อจัดการความไม่สมดุลของข้อมูล ซึ่งประกอบด้วย เทคนิคการปรับเพิ่มข้อมูลด้วยวิธีการสุ่มโดยการสุ่มเพิ่มจำนวนของข้อมูลกลุ่มน้อยให้มีปริมาณใกล้เคียงกับข้อมูลกลุ่มมาก, เทคนิคการปรับลดข้อมูลด้วย

วิธีการสุ่มโดยการสุ่มลดจำนวนของข้อมูลกลุ่มมากให้มีปริมาณใกล้เคียงกับกลุ่มที่น้อย และเทคนิคการปรับเพิ่มและลดข้อมูลด้วยวิธีการสุ่ม (วิชญ์วิสิฐ, 2561) ดังต่อไปนี้

1. วิธีการสุ่มเกิน (Oversampling) เป็นการเพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อยให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมาก ซึ่งการเพิ่มข้อมูลนั้นจะเพิ่มโดยการสุ่มเลือกจากข้อมูลเดิม

2. วิธีสุ่มลด (Under sampling) เป็นการลดจำนวนข้อมูลที่อยู่ในกลุ่มส่วนมากให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลที่อยู่ในกลุ่มส่วนน้อย

3. วิธีผสมผสาน (Hybrid Methods) เป็นวิธีการที่นำเทคนิควิธีสุ่มเกิน และสุ่มลดมาทำงานร่วมกัน โดยพยายามหาค่ากลางในการชักตัวอย่างให้ได้ตามจำนวนที่อยู่ตรงกลางระหว่างข้อมูลในกลุ่มส่วนมากกับข้อมูลในกลุ่มส่วนน้อย

4. วิธีสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) เป็นเทคนิคการสุ่มตัวอย่างแบบพิเศษของการสุ่มเพิ่ม แทนที่จะสุ่มเพิ่มโดยใช้ข้อมูลเดิมแต่จะทำการสังเคราะห์ข้อมูลขึ้นมาใหม่จากข้อมูลเดิมที่มีอยู่ หลักการเพื่อนบ้านที่อยู่ใกล้ที่สุด (K-nearest neighbor)

2.3.2 การแบ่งช่วงข้อมูล (Discretization)

S. C. Tan. (2018) อธิบายการแบ่งช่วงข้อมูลคือการทำให้ข้อมูลเป็นแบบไม่ต่อเนื่อง เป็นการลดรูปข้อมูลเลขจำนวนจริงโดยแบ่งค่าข้อมูลออกเป็นช่วง แล้วแทนค่าแต่ละช่วงด้วยค่าช่วง (Interval Label) เนื่องจากอัลกอริทึมการทำเหมืองข้อมูลบางอย่างรับเฉพาะข้อมูลแบบกลุ่มเท่านั้น (Categorical) ซึ่งกระบวนการแบ่งช่วงข้อมูลแบบดั้งเดิมมีดังนี้

1. การกำหนดความถี่ของช่วงข้อมูลที่เท่ากัน (Equal-Frequency Binning) คือการกำหนดความถี่แต่ละช่วงให้มีขนาดเท่ากัน โดยไม่สนใจความกว้างของช่วง การแบ่งช่วงข้อมูลด้วยวิธีนี้จะได้ผลที่ดีเมื่อข้อมูลมีการแจกแจงแบบปกติ และสามารถจัดการได้ดีกับข้อมูลที่มีความโด่ง และขนาดข้อมูลที่แตกต่างกัน

2. การกำหนดช่วงความกว้างของข้อมูลที่เท่ากัน (Equal-Width Binning) คือการกำหนดช่วงความกว้างของข้อมูล โดยไม่สนใจความถี่ของข้อมูลในแต่ละช่วง การแบ่งช่วงข้อมูลด้วยวิธีนี้จะได้ผลที่ดีเมื่อข้อมูลมีการแจกแจงแบบปกติ และขนาดของข้อมูลมีความใกล้เคียงกัน แต่จะจัดการได้ไม่ดีหากข้อมูลนั้นมีความโด่งที่แตกต่างกันมาก

3. วิธีการแบ่งช่วงข้อมูลด้วยวิธี TwoStep Clustering โดยใช้โปรแกรม SPSS จะใช้อัลกอริทึม Bayesian Information Criterion (BIC) เพื่อกำหนดจำนวนกลุ่มโดยอัตโนมัติ ซึ่งแบ่งข้อมูลที่เป็นแบบต่อเนื่องเป็นชุดข้อมูลที่ไม่นับซ้อนกัน

เกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria : BIC) มีสูตรดังสมการต่อไปนี้

$$BIC = n \cdot \ln \left(\frac{SSE}{n} \right) + \frac{2(p+1)n\sigma^2}{SSE} - \frac{2n^2\sigma^4}{SSE^2} \quad (2.1)$$

เมื่อ	n	เป็นขนาดตัวอย่าง
	SSE	เป็นค่าความคลาดเคลื่อนกำลังสองของตัวแบบการถดถอย
	σ^2	เป็นค่าความคลาดเคลื่อนเฉลี่ยกำลังสองของตัวแบบการถดถอย
	p	เป็นจำนวนพารามิเตอร์ในตัวแบบการถดถอย
และ	ln	เป็นลอการิทึมฐานอี

2.4 กฎความสัมพันธ์ (Association Rule)

Association Rule เป็นเทคนิคที่ใช้ในงานวิจัยสำหรับการค้นหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยมีวัตถุประสงค์เพื่อค้นหาความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ (สุวิจักขณ์ ไศธนะกุล และคณะ, 2559) ตัวอย่างการนำกฎความสัมพันธ์นี้ไปใช้ ได้แก่ การวิเคราะห์การซื้อสินค้าของลูกค้า (Market Basket Analysis) ซึ่งวิธีนี้จะหาความสัมพันธ์ของสินค้าที่ลูกค้ามักจะซื้อพร้อมกัน ข้อมูลการซื้อสินค้าเหล่านี้สามารถนำมาช่วยในการวางแผนการตลาด เช่น การจัดวางสินค้าที่มักจะถูกซื้อด้วยกันไว้ใกล้ๆกัน หรือการจัดโปรโมชั่นให้กับสินค้า เป็นต้น (Bao et al., 2022)

ในงานวิจัยฉบับนี้ได้นำเสนอรูปแบบกฎความสัมพันธ์ (Association Rule) ซึ่งเป็นการแสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุที่เกิดขึ้น จะอยู่ในรูปแบบ $A \Rightarrow B$ การหากฎความสัมพันธ์มีเทคนิคอยู่หลายวิธีด้วยกัน วิธีที่ได้รับการยอมรับและนิยมอย่างหนึ่งคือวิธีอัลกอริทึมอปรไอริเป็นอัลกอริทึมในการสืบค้นกฎความสัมพันธ์ที่เกิดขึ้นในฐานข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.4.1 นิยามคำศัพท์ที่เกี่ยวข้องกับกฎความสัมพันธ์

1. ไอเทมเซต (Itemset) คือ ความสัมพันธ์ของข้อมูลที่ได้ในฐานข้อมูล โดยไอเทมเซตประกอบด้วย ไอเทมที่ k-itemsets เช่น กรณี 2-itemsets ได้แก่ {item1, item2} และ กรณี 3-itemsets ได้แก่ {item1, item2, item3} เป็นต้น (ภัทรารุณ แสงศิริ, ชนกรมณ สุขศรี, และ พรนภา เจมส์, 2557)

2. ฟรีควีนไอเทมเซต (Frequent Itemset) คือ ไอเทมเซตที่ผ่านค่าสนับสนุนขั้นต่ำ (Minimum Support) จะสร้างเป็นกฎความสัมพันธ์ โดยกฎความสัมพันธ์ทั่วไปเขียนได้ดังนี้ LHS => RHS (บุษราภรณ์ มหัทธชัย และคณะ, 2558) โดยที่ LHS (Left Hand Side) แสดงรูปแบบของไอเทมเซตด้านซ้ายของกฎความสัมพันธ์ และ RHS (Right Hand Side) แสดงรูปแบบของไอเทมเซตด้านขวาของกฎความสัมพันธ์ จากตัวอย่างกฎความสัมพันธ์ Item1 => Item2 LHS คือ Item1 และ RHS คือ Item2

3. ค่าสนับสนุน (Support) คือ ความน่าจะเป็นของจำนวนไอเทมเซตที่พบในฐานข้อมูลต่อจำนวนรายการทั้งหมด ซึ่งคำนวณหาค่าสนับสนุนของไอเทมเซต LHS ได้จาก

$$\text{Support (LHS)} = \frac{\text{Number of Transactions that Contain LHS}}{\text{Number of all Transactions}} \quad (2.2)$$

Support (LHS) คือ ค่าสนับสนุนของไอเทมเซต LHS

Number of Transactions that contain LHS คือ จำนวนรายการข้อมูลของไอเทมเซต LHS

Number of all Transactions คือ จำนวนรายการข้อมูลทั้งหมด

4. ค่าความเชื่อมั่น (Confidence) คือ การแสดงค่าความเชื่อมั่นของกฎความสัมพันธ์เมื่อรูปแบบ LHS ที่อยู่ทางด้านซ้ายของกฎเกิดขึ้นแล้วมีโอกาสเกิดรูปแบบ RHS ที่อยู่ทางด้านขวามากน้อยเท่าใด ซึ่งจะมีค่าอยู่ระหว่าง 0-1 ถ้าใกล้เคียง 1 หมายถึงมีความเชื่อมั่นในการหาความสัมพันธ์มาก สามารถคำนวณค่าที่จะเกิดขึ้นในรูปแบบเปอร์เซ็นต์ หรือ ร้อยละได้

$$\text{Confidence (LHS => RHS)} = \frac{\text{Support (LHS, RHS)}}{\text{Support (LHS)}} \quad (2.3)$$

Support (LHS, RHS) คือ ค่าสนับสนุนที่รูปแบบ LHS และ RHS ของกฎความสัมพันธ์เกิดขึ้นพร้อมๆกัน ในส่วนของ Support (LHS) คือค่าสนับสนุนรูปแบบที่อยู่ด้านซ้ายของกฎความสัมพันธ์

5. ค่าสหสัมพันธ์ หรือ ค่าลิฟต์ (Lift) คือ ค่าที่บ่งบอกว่าการเกิดรูปแบบ LHS และ RHS มีความสัมพันธ์กันมากหรือไม่ โดยถ้าค่าลิฟต์ มีค่าเท่ากับ 1 แสดงว่ารูปแบบ LHS และ RHS ไม่ขึ้นต่อกัน (Independent) แต่ถ้ามีค่ามากกว่า 1 มาก ๆ แสดงว่ากฎทั้งสองมีความสัมพันธ์กันมากด้วยเช่นกัน ค่าลิฟต์คำนวณได้จาก

$$\text{Lift (LHS} \Rightarrow \text{RHS)} = \frac{\text{Support (LHS, RHS)}}{\text{Support (LHS)} \times \text{Support (RHS)}} \quad (2.4)$$

Support (LHS, RHS) คือค่าสนับสนุนที่รูปแบบ LHS และ RHS ของกฎความสัมพันธ์ เกิดขึ้นพร้อม ๆ กัน Support (LHS) คือค่าสนับสนุนรูปแบบที่อยู่ด้านซ้ายของกฎความสัมพันธ์ และ Support (RHS) คือค่าสนับสนุนรูปแบบที่อยู่ด้านขวาของกฎความสัมพันธ์ (สุวิจักขณ์ โศรณะกุล และคณะ, 2559)

2.4.2 วิธีการค้นหาความสัมพันธ์มีขั้นตอนหลัก 2 ขั้นตอนดังนี้

1. การหา Frequent Itemset คือ กลุ่มของไอเทมที่มีค่าสนับสนุน (Support) มากกว่า หรือ เท่ากับค่าสนับสนุนขั้นต่ำ (Minimum Support) โดยซับเซตของ Frequent Itemset เช่น ถ้า {A,B} เป็น Frequent Itemset แล้ว {A} และ {B} ต้องเป็น Frequent Itemset ด้วยเช่นกัน
2. การสร้างกฎความสัมพันธ์จาก Frequent Itemset ที่ได้จากการดำเนินการในขั้นตอนที่ 1 และจะยอมรับกฎความสัมพันธ์ที่สร้างขึ้นมาก็ต่อเมื่อ กฎนี้มีค่าความเชื่อมั่น (Confidence) มากกว่าหรือเท่ากับ ค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) รูปแบบทั่วไปของกฎความสัมพันธ์คือ $X \rightarrow Y$ โดยที่ X เป็นเงื่อนไข และ Y เป็นผลลัพธ์ที่เกิดขึ้น หรือ อยู่ในรูปของถ้า...แล้ว (IF...THEN...) การประเมินค่าของกฎว่าจะสามารถนำไปใช้งานได้หรือไม่จะใช้ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence)

2.5 งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้างานวิจัยต่างๆ สามารถแบ่งออกเป็น 3 กลุ่ม ได้แก่ กลุ่มที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending กลุ่มที่เกี่ยวข้องกับการจัดเตรียมข้อมูล และกลุ่มที่เกี่ยวข้องกับการหาความสัมพันธ์

2.5.1 งานวิจัยที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending

Zixi Wang et al., (2019) ได้ศึกษาการป้องกันความเสี่ยงที่เกิดขึ้นจากข้อมูลออนไลน์ โดยใช้ข้อมูลจาก P2P lending เนื่องจากมองว่า P2P lending เกิดขึ้นมาจากการพัฒนาของอินเทอร์เน็ตที่ใช้ร่วมกับธุรกรรมทางการเงินซึ่งมีความเสี่ยงสูงที่จะเกิดการทุจริตได้ง่าย ในงานศึกษาจึงได้ทำการวิเคราะห์ข้อมูลในหลายมิติของโปรไฟล์ผู้กู้ยืม ซึ่งงานวิจัยนี้กล่าวถึงโปรไฟล์ของผู้กู้ยืมใน 4 มิติ ได้แก่ คุณลักษณะพื้นฐาน ,คุณลักษณะความสามารถ ,คุณลักษณะทางสังคม และคุณลักษณะทางจิตวิทยา เพื่อสร้างแบบจำลอง จากนั้นจะกำหนดวิธีการวิเคราะห์ข้อมูลโดยเฉพาะใน

แต่ละมิติ และคุณลักษณะที่แตกต่างกัน ในบทความนี้ได้นำข้อมูลของผู้กู้ที่ลงทะเบียนผ่านแพลตฟอร์มชั้นนำด้าน P2P lending ของ SINA และ Weibo เป็นแหล่งข้อมูล และใช้วิธี Bayesian Network , การเงินเชิงพฤติกรรม (Behavior Finance Theory) , และการทำเหมืองข้อความ (Text Mining Technology) ในการวิเคราะห์ ซึ่งได้ผลที่น่าเชื่อถือ

อติวิชญ์ ชนินทรโชติก และคณะ (2564) ได้ศึกษาอัตราการผิดนัดชำระหนี้ของผู้กู้จากชุดข้อมูล P2P Lending ของแพลตฟอร์ม Lending Club โดยใช้เทคนิคการสุ่มข้อมูลในการแก้ปัญหาความไม่สมดุลของข้อมูลและเทคนิค Feature Selection ที่ช่วยเพิ่มประสิทธิภาพของโมเดลให้มีประสิทธิภาพเพิ่มมากขึ้น ซึ่งผลลัพธ์ที่ทำให้โมเดลมีค่ามากที่สุดคือ การรวมกันระหว่างเทคนิค IHT Under Sampling และโมเดล Neural Network ได้ประสิทธิภาพ Recall 93.13% และค่า AUC 59.2% และปัจจัยที่ส่งผลในการทำนายว่าผู้กู้มีแนวโน้มในการผิดนัดชำระหนี้หรือไม่ อันดับแรกคือ อัตราดอกเบี้ย, คะแนน Credit Scoring, ระยะเวลาในการกู้, สถานะบ้านที่อยู่อาศัยของผู้กู้ และจำนวนเงินเฉลี่ยในบัญชีทั้งหมด

P. Mukherjee et al., (2022) ได้ศึกษาเทคนิคการเรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) เพื่อทำนายผู้ขอสินเชื่อที่ไม่สามารถชำระหนี้ได้ตามเงื่อนไข ประกอบด้วย 4 เทคนิค 1) Self-Organizing Map 2) Density based Spatial Clustering, 3) Elliptic Envelope และ 4) Auto-Encoders บนชุดข้อมูล Lending Club โดยใช้เทคนิคการคัดเลือกลักษณะ เพื่อเพิ่มประสิทธิภาพในการตรวจจับข้อผิดพลาดในรูปแบบของผู้กู้ที่ไม่ชำระเงินกู้ ผลการวิจัยแสดงให้เห็นว่า self-organizing map มีประสิทธิภาพที่ดีที่สุดในการทำนายผู้ขอสินเชื่อที่ไม่ชำระเงินกู้ได้ ด้วยความแม่นยำ (precision) 0.79 และค่าความถูกต้อง (recall) 0.816

K. Chengeta et al., (2023) ได้ศึกษาขั้นตอนการเรียนรู้ของเครื่องจักร (Machine Learning Algorithms) เพื่อทำนายความเป็นไปได้ในการไม่ชำระหนี้ของผู้ขอสินเชื่อ และความผิดปกติในของใบสมัครผ่านแพลตฟอร์ม P2P Lending โดยใช้ข้อมูลจากแพลตฟอร์ม Prosper และ Lending Club ด้วยอัลกอริทึม CNN, XGBoost, CatBoost และ LightGBM ซึ่งถูกปรับแต่งให้เหมาะสมเพื่อลดการโอเวอร์ฟิตติ้ง (Overfitting) ในการศึกษาพบว่า สถานะการจ้างงาน, อายุ และคะแนนเครดิตเป็นตัวทำนายการผิดนัดชำระหนี้ที่ดีที่สุด

Cheng, Y.-C et al., (2021) ได้เสนอระบบการทำนายความเสี่ยงด้านเครดิต (Credit risk prediction system- CRPS) เพื่อแก้ไขปัญหาความไม่สมดุลของข้อมูลและคัดเลือกคุณลักษณะ และทำนายความเสี่ยงจากการผิดนัดชำระหนี้ โดยใช้ข้อมูลการขอสินเชื่อออนไลน์ระหว่างบุคคล P2P Lending จากแพลตฟอร์ม LendingClub ผลลัพธ์ที่ได้พบว่า ความแม่นยำของ CRPS สามารถบรรลุได้ถึง 99% อัตราครอบครอง (recall) สูงถึง 0.95 และ F1-Score คือ 0.97

CRPS สามารถทำนายความเสี่ยงเครดิตอย่างแม่นยำด้วยคุณสมบัติน้อยกว่า 10 รายการและจัดการกับปัญหาความไม่สมดุลของข้อมูลได้เป็นอย่างดี

Chulawate and Kiattisin (2023) ได้เสนองานวิจัยเพื่อช่วยในการระบุปัจจัยที่มีต่อการประสบความสำเร็จในธุรกิจการให้สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending) ของประเทศไทย โดยรวบรวมข้อมูลจากผู้ตอบแบบสอบถามที่มีจำนวน 300 คน และใช้สมการโครงสร้าง (Structural Equation Modeling :SEM) เพื่อวิเคราะห์ข้อมูลผ่าน Mplus เวอร์ชัน 7 เพื่อให้เข้าใจข้อมูลโดยละเอียดยิ่งขึ้น และมุ่งเน้นและให้ความสำคัญกับความแตกต่างของธุรกิจการเงินในแต่ละประเทศเนื่องจากสถานการณ์และสภาพแวดล้อมที่แตกต่างกันซึ่งอาจเป็นปัจจัยที่ท้าทายในการสรุปข้อมูล และการสร้างความยั่งยืนในอุตสาหกรรมการเงิน โดยผลลัพธ์ที่ได้นั้นสามารถระบุปัจจัยออกเป็น 3 กลุ่ม ได้แก่ 1. ปัจจัยด้านความเสี่ยง อาทิเช่น อัตราดอกเบี้ยที่สูง, การเพิ่มขึ้นของอัตราเงินเฟ้อ เป็นต้น . 2. ปัจจัยด้านความเชื่อมั่น อาทิเช่น ปัจจัยด้านการเงิน ,สถานะเครดิต และ 3. ปัจจัยด้านข้อมูล อาทิเช่น นโยบายของบริษัท, การผิณัดชำระของผู้ขอสินเชื่อ เป็นต้น

2.5.2 งานวิจัยที่เกี่ยวข้องกับการจัดเตรียมข้อมูล

วิญญูวิสิฐ เกสรสิทธิ์ และคณะ (2561) ได้ศึกษาถึงการจำแนกโดยใช้ข้อมูลที่ไม่สมดุลเป็นปัญหาสำคัญในเทคนิคการจำแนก ซึ่งการจำแนกข้อมูลที่มีข้อมูลในกลุ่มมากและกลุ่มน้อยปะปนกันจะทำให้ข้อมูลในกลุ่มมากจะมีคุณสมบัติบางประการที่บดบังคุณลักษณะของกลุ่มน้อย ทำให้การจำแนกข้อมูลในกลุ่มน้อยไม่สามารถจำแนกได้อย่างมีประสิทธิภาพ ในงานวิจัยจึงได้แก้ปัญหาข้อมูลไม่สมดุลที่ระดับข้อมูลจำนวน 4 วิธี คือ วิธีสุ่มเกิน วิธีสุ่มลด วิธีผสมผสาน และวิธีสังเคราะห์ข้อมูลใหม่ (SMOTE) โดยใช้เทคนิคการจำแนก คือ วิธีการถดถอยโลจิสติกแบบมัลติโนเมียล และวิธีต้นไม้การตัดสินใจ ในการจำแนกผู้ป่วยโรคเบาหวาน ซึ่งจากการเปรียบเทียบประสิทธิภาพของสถิติและอัลกอริทึมในการจำแนก พบว่าข้อมูลที่แก้ปัญหาคือความไม่สมดุลด้วยวิธีสังเคราะห์ข้อมูลใหม่สามารถจำแนกผู้ป่วยโรคเบาหวานมีผลลัพธ์ที่ดีที่สุด ที่ความถูกต้อง 65.33% ในส่วนของวิธีการอื่นๆ วิธีการสุ่มเพิ่ม ,วิธีการสุ่มลด ,วิธีการผสมผสาน มีความถูกต้องที่ 64.00% ,64.40% และ 64.55% ตามลำดับ

Yan and Zhang (2015) ได้ศึกษาอัลกอริทึม SVM-RFE สำหรับการคัดเลือกคุณลักษณะที่มีประสิทธิภาพ อย่างไรก็ตามหาชุดข้อมูลมีคุณลักษณะที่มีความสัมพันธ์กันสูง เกณฑ์การจัดอันดับของ SVM-RFE จะเกิดความเอนเอียงทำให้ผลที่ได้ไม่น่าเชื่อถือ เพราะฉะนั้นหลังจากตรวจสอบความสัมพันธ์ของคุณลักษณะแล้วนั้นจะใช้วิธีการ Correlation Bias Reduction (CBR) เพื่อเข้าขั้นตอนลดคุณลักษณะ SVM-RFE + CBR การทดลองดำเนินการกับชุดข้อมูลวิเคราะห์ลม

หายใจ ออกแบบจุ่มอิเล็กทรอนิกส์ในอนาคตซึ่งผลสรุปที่ได้พิสูจน์ให้เห็นว่าวิธีการ SVM-RFE + CBR มีประสิทธิภาพที่สูงกว่า SVM-RFE แบบดั้งเดิม

S. C. Tan (2018) ได้ศึกษาถึงการแบ่งช่วงข้อมูล (Discretization) โดยที่ไม่คำนึงถึงการกระจายตัวของข้อมูล เพื่อประเมินประสิทธิภาพของอัลกอริทึมออพริโมริ ตามชุดข้อมูล 4 ชุด โดยเสนอวิธีการ TwoStep Clustering ในการแบ่งช่วงข้อมูลเปรียบเทียบการแบ่งช่วงข้อมูลแบบดั้งเดิม 2 วิธี ได้แก่ 1. การกำหนดช่วงความกว้างของข้อมูลที่เท่ากัน (Equal-Width Binning) 2. การกำหนดความถี่ของช่วงข้อมูลที่เท่ากัน (Equal-Frequency Binning) ซึ่งในงานวิจัยได้อธิบายข้อเสียที่เกิดจากการแบ่งช่วงข้อมูลแบบดั้งเดิมไว้ นั่นคือปัญหาของการกำหนดช่วงข้อมูล ซึ่งหากน้อยเกินไปจะทำให้สูญเสียข้อมูลที่สำคัญได้ ในทางตรงข้ามหากกำหนดช่วงข้อมูลが多เกินไปนั้นก็จะเป็นตัวแทนของข้อมูลที่ไม่เหมาะสม จึงเป็นเหตุของการเสนอวิธีการ TwoStep Clustering ในการแบ่งช่วงข้อมูล ในงานวิจัยนี้ได้แสดงให้เห็นว่าการแยกตามการจัดกลุ่มอัตโนมัติ TwoStep Clustering เป็นวิธีที่มีคุณภาพ ซึ่งสามารถใช้ได้กับข้อมูลที่มีการกระจายตัวได้ทุกประเภท ดังนั้นจึงเป็นทางเลือกที่ดีสำหรับการแบ่งช่วงข้อมูลก่อนการหาความสัมพันธ์

Chen and Suzhou (2021) ได้พัฒนาวิธีการทำนายความเสี่ยงในการผิดนัดชำระหนี้ใน P2P lending โดยใช้ XGBoost และวิธีการจัดการข้อมูล 3 วิธีการ คือ SMOTE, NearMiss, และการเลือกสุ่ม 1:1 แบบกำหนดเอง เพื่อแก้ไขปัญหาของข้อมูลที่ไม่สมดุลกัน ผลลัพธ์แสดงให้เห็นว่า XGBoost พร้อมวิธีการจัดตัวอย่างเหล่านี้เป็นวิธีที่มีประสิทธิภาพที่สุดในการทำนายความเสี่ยงในการผิดนัดชำระหนี้โดยไม่มีปัญหาความไม่สมดุลของข้อมูล นอกจากนี้ ข้อมูล P2P lending มักจะมีจำนวนคุณสมบัติที่มากมาย ซึ่งอาจทำให้ยากต่อการระบุคุณสมบัติที่เกี่ยวข้องและลำดับของลักษณะ

Suchada and Devasia (2022) ได้สร้างระบบทำนายความเสี่ยงเครดิต (CRPS) สำหรับข้อมูล Lending Club โดยใช้การคัดเลือกคุณลักษณะด้วย RFECV และสังเคราะห์ข้อมูลด้วย borderline-SMOTE เพื่อใช้กับแบบจำลองทำนายความเสี่ยง การทดลองแสดงให้เห็นว่า XGBoost กับ RFECV ได้รับความแม่นยำอยู่ที่ 99% ค่า Recall ที่ 0.95 และค่า F1-Score ที่ 0.97 โดยใช้เพียงแค่ 10 คุณลักษณะ นอกจากนี้ เทคนิคการเลือกคุณลักษณะเช่น RFECV สามารถลดจำนวนคุณลักษณะที่ใช้ในโมเดลได้อย่างมีนัยสำคัญ ซึ่งไม่เพียงแต่ช่วยประหยัดเวลาแต่ยังเพิ่มประสิทธิภาพในการโฟกัสไปที่ตัวทำนายที่สำคัญที่สุดของความเสี่ยงเครดิต ในท้ายที่สุดการแบ่งกลุ่มข้อมูลต่อเนื่องเป็นอีกหนึ่งความท้าทายที่ข้อมูลการให้กู้ยืมแบบ P2P พบอยู่

Mei and Jiang (2016) ได้เสนองานการคัดเลือกคุณลักษณะด้วยการใช้วิธีการหาความสัมพันธ์ ซึ่งประกอบด้วยขั้นตอนหลัก 3 ขั้นตอน ดังนี้ 1. กำหนดพารามิเตอร์สำหรับการหาความสัมพันธ์ด้วยอัลกอริทึมออพริโมริ 2. คัดเลือกความสัมพันธ์ ด้วยวิธีการที่เฉพาะเจาะจง 3. ขั้นสุดท้ายคือการสร้างเซตคุณลักษณะซึ่งอธิบายผลลัพธ์จากการเลือกคุณลักษณะ และแสดง

กระบวนการที่ใช้ประเมินคุณลักษณะโดยอัลกอริทึมกำหนดไว้ การทดลองแสดงให้เห็นว่าวิธีการที่นำเสนอทำงานได้ดีกว่าวิธีอื่นๆซึ่งมีอัลกอริทึมการเลือกคุณลักษณะต่างกัน กรอบพัฒนาที่นำเสนอไม่เพียงลดขนาดของคุณลักษณะ แต่ยังปรับปรุงและเพิ่มประสิทธิภาพการจำแนกประเภทอีกด้วย

Bahl et al., (2019) ได้เสนองานวิจัยเพื่อจัดกลุ่มนาโนวัสดุโดยอาศัยคุณสมบัติทางกายภาพ และคุณสมบัติด้านเคมี เพื่อจำแนกนาโนวัสดุที่เป็นอันตรายต่อระบบการย่อยอาหาร และระบบทางเดินหายใจของร่างกาย การวิเคราะห์ได้ใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ด้วยวิธีการ Principal Component Analysis (PCA) จากนั้นใช้วิธีการ K-nearest Neighbors ในการหาความคล้ายคลึงกันเพื่อแยกกลุ่มนาโนวัสดุ ต่อมาใช้วิธีการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้วิธี Random Forests (RFs) เพื่อหาค่าคล้ายคลึงของกลุ่มนาโนวัสดุ โดยให้ความสนใจเฉพาะกลุ่มสมบัติที่มีค่าสหสัมพันธ์ที่สูงกับสิ่งที่สนใจ ต่อจากนั้นได้ทำการปรับปรุงประสิทธิภาพ ด้วยวิธีการ Recursive Feature Elimination (RFE) เพื่อลบคุณสมบัติที่ไม่เกี่ยวข้องออก ผลการวิจัยแสดงให้เห็นว่า Random Forests model ที่คัดเลือกคุณลักษณะด้วยวิธีการ RFE ให้ผลลัพธ์ที่ดีที่สุด โดยได้รับความแม่นยำในการแยกกลุ่มที่สมดุลมากถึง 0.82 ในระดับความเป็นอันตรายที่นำเสนอ จาก 11 คุณสมบัติที่ศึกษา พบว่า Zeta Potential, Redox Potential และ Dissolution Rate เป็นตัวชี้วัดที่มีผลทำนายที่แข็งแกร่งที่สุดต่อกิจกรรมชีวภาพของนาโนวัสดุในชุดข้อมูลปัจจุบัน

Misra et al., (2020) ได้อธิบายการคัดเลือกคุณลักษณะ (Feature selection) ว่าเป็นสิ่งสำคัญและช่วยเพิ่มความแม่นยำของโมเดลอย่างมีนัยสำคัญ จึงได้เสนอวิธีการลดคุณลักษณะที่ไม่จำเป็น เพื่อเพิ่มความแม่นยำในการจำแนกด้วยวิธี Recursive Feature Elimination with Cross-Validation (RFECV) เพื่อทำนายการเกิดโรคเบาหวานชนิดที่ 2 (Type-II Diabetes) โดยอุปสรรคสำคัญของวิธีนี้คือการจัดการกับปัญหา Overfitting และเพิ่มความแม่นยำโดยไม่ต้องลบข้อมูลที่ไม่จำเป็นออกไป นอกจากนี้ยังเสนอวิธีการจัดเตรียมข้อมูลต่างๆ เพื่อนำไปใช้กับอัลกอริทึมที่แตกต่างกัน เช่น ส่วนการเรียนรู้เครื่องจักร (Machine Learning) , Naïve Bayes, Support Vector Machine และ Decision Tree (DT) ซึ่งผลการทำนายของโมเดลการถดถอยโลจิสติก (Logistic Regression) ให้ความแม่นยำสูงที่สุด (84%) ซึ่งใกล้เคียงกับโมเดลอื่น ๆ

อัศวิน สุรวัชโยธิน และ วรภัทร ไพรีเกรง (2564) ได้สร้างตัวแบบการทำนายในการเลือกศึกษาต่อระดับอุดมศึกษาตามทักษะของกลุ่มผู้เรียนในเชิงพหุปัญญาโดยใช้เทคนิคบูรณาการโดยอาศัยการแก้ปัญหาข้อมูลขาดหายด้วย Single Imputation การปรับข้อมูลสมดุลด้วย SMOTE การคัดเลือกคุณลักษณะด้วย Recursive Feature Elimination (RFE) โดยการใช้เทคนิค Stacking คือ ทักษะด้านเหตุผลและคณิตศาสตร์ มีค่าความสำคัญ 77% เมื่อเปรียบเทียบกับเทคนิคอื่น ซึ่ง

ทักษะนี้เป็นรากฐานสำคัญของวิทยาการหลายสาขาส่งผลให้ทักษะนี้มีความสำคัญในการศึกษาต่อระดับอุดมศึกษา

2.5.3 งานวิจัยที่เกี่ยวข้องกับการหาความสัมพันธ์

บุษราภรณ์ มัทธนชัย และคณะ (2559) ได้ศึกษาหาความสัมพันธ์ของรายวิชาที่มีผลต่อการพัฒนาทักษะนักศึกษา โดยใช้อัลกอริทึมออริโอริซึ่งเป็นเทคนิคในการทำเหมืองข้อมูล โดยการศึกษากลุ่มตัวอย่างจาก นักศึกษาของภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ที่เข้าศึกษาในปีการศึกษา 2553-2558 ซึ่งมีจำนวน 1,309 คน และทำการเตรียมและปรับข้อมูลเพื่อนำไปค้นหาความสัมพันธ์โดยใช้อัลกอริทึมออริโอริ และทดสอบความถูกต้องของกฎที่ได้ ผลการวิจัยพบว่ากฎ ความสัมพันธ์ของรายวิชาที่มีผลต่อการพัฒนาทักษะนักศึกษาชั้นปีที่ 1, 2, 3 มีจำนวนเท่ากับ 27, 15, 14 กฎ ตามลำดับ จากการวิจัยนี้สามารถนำความสัมพันธ์ ไปประยุกต์ใช้ในการพัฒนาระบบแนะนำการลดความเสี่ยงจากการพัฒนาทักษะนักศึกษา ด้วยปรับปรุงหลักสูตร กระบวนการเรียน การสอน การวางแผนในการลงทะเบียนของนักศึกษา และงานอื่น ๆ ที่เกี่ยวข้องเพื่อคุณภาพการศึกษาให้ดีขึ้น

K. R. Seeja et al., (2014) ได้ศึกษาและพัฒนาแบบจำลองในการตรวจจับการทุจริตบัตรเครดิตแบบอัตโนมัติ โดยใช้อัลกอริทึมออริโอริในการหาความสัมพันธ์ เพื่อสกัดคุณลักษณะของลักษณะการทุจริต และลักษณะของรายการที่เป็นปกติออกจากกัน และนำรูปแบบที่จำแนกมาสร้างเป็นระบบในการตรวจสอบการทุจริต ในด้านประเมินแบ่งเป็นการจำแนกใน 4 รายการ ประกอบด้วย อัตราการตรวจจับการฉ้อโกง อัตราการเตือนที่ผิดพลาด การจำแนกประเภทที่สมดุล อัตราและค่าสัมประสิทธิ์สหสัมพันธ์ของแมทริคซ์ และทำการเปรียบเทียบการจำแนกข้อมูลด้วยวิธีต่างๆ ได้แก่ support vector machine (SVM), K-near Neighbor classifier, Naive Bayes ซึ่งจากผลทดลองนั้นแสดงให้เห็นว่าวิธีการดังกล่าวมีประสิทธิภาพในการจำแนกลักษณะการทุจริตของรายการบัตรเครดิต และลักษณะของรายการที่ปกติได้

Hegazy (2016) ได้ศึกษาและพัฒนาโมเดลในการตรวจสอบการทุจริตในการใช้บัตรเครดิต การพัฒนาในงานวิจัยนี้ได้กล่าวถึงการทำเหมืองข้อมูลโดยใช้อัลกอริทึมออริโอริเพื่อสร้างหาความสัมพันธ์ของลูกค้าที่ทุจริตการใช้บัตรเครดิต และกลุ่มลูกค้าปกติ หลังจากนั้นจะใช้อัลกอริทึม Lingo เพื่อหารูปแบบลูกค้าที่มีคุณสมบัติของการเป็นลูกค้าที่ทุจริต และกลุ่มลูกค้าปกติ ในด้านของผลลัพธ์ที่ได้นำมาเปรียบเทียบกับผลการประเมินการทุจริตในรูปแบบเก่า ซึ่งได้แสดงให้เห็นว่าการปรับปรุงอัลกอริทึมในการตรวจสอบการทุจริตดีขึ้นกว่าการประเมินในรูปแบบก่อน แต่อย่างไรก็ตามการตรวจสอบการทุจริตรูปแบบใหม่นี้ยังมีข้อจำกัดที่อาจเกิดขึ้นได้เนื่องจากรูปแบบของลูกค้าที่ทุจริต และลูกค้าปกติ อาจจะมีลักษณะที่ซ้ำกันได้

Hsueh and Kuo (2017) ได้อธิบายถึงการหาความสัมพันธ์กับข้อมูลการให้กู้ยืมผ่านแพลตฟอร์ม P2P Lending ว่ามีความท้าทายหลายประการ เช่น ข้อมูลที่ไม่สมดุล การจัดกลุ่มของข้อมูลตัวเลข ซึ่งทำการวิเคราะห์ข้อมูลจากแพลตฟอร์ม Zopa ในสหราชอาณาจักร โดยใช้ช่วงควอไทล์เพื่อจัดอันดับข้อมูลตัวเลข จากนั้นทำการหาความสัมพันธ์โดยการตั้งค่าความสัมพันธ์ขั้นต่ำไว้ที่ 0.2 และความเชื่อมั่นที่ 0.9 อย่างไรก็ตาม อันหนึ่งในความท้าทายหลักของการใช้เครื่องมือชุดค่าไรอาชีพกับข้อมูลการให้กู้ยืม P2P คือความไม่สมดุลของข้อมูล ข้อมูลการให้กู้ยืม P2P มักมีจำนวนสินเชื่อที่ไม่ผิดนัดชำระเงินมากมายและจำนวนสินเชื่อที่ผิดนัดชำระเงินน้อยมาก ซึ่งอาจทำให้ยากต่อการระบุรูปแบบและความสัมพันธ์ในข้อมูล

Desai and Kaiwade (2018) ได้ศึกษาการหาความสัมพันธ์ของรายการธุรกรรมทางบัญชีธนาคารของลูกค้า กับพฤติกรรมทางการเงิน โดยใช้อัลกอริทึมออริโอริเป็นเครื่องมืองานวิจัยนี้ศึกษาเพื่อเป็นแนวทางการเพิ่มประสิทธิภาพของการหาข้อมูลลูกค้าเงินฝากเพื่อเพิ่มเงินทุนให้กับธนาคาร

Jesus Silva et al., (2019) ได้ศึกษาพฤติกรรมของลูกค้าในบริษัทต่างๆ เพื่อสร้างกลยุทธ์ทางการตลาดที่ส่งถึงลูกค้าประเภทนั้นๆ ได้อย่างตรงจุด โดยนำระเบียบวิธี CRISP-DM มาประยุกต์ใช้สำหรับกระบวนการทำเหมืองข้อมูล ฐานข้อมูลที่ใช้คือฐานข้อมูลที่สอดคล้องกับภาคธุรกิจของ SMEs ที่อ้างอิงถึงลูกค้าและยอดขายสินค้าแต่ละประเภท การวิเคราะห์ทำขึ้นตามแบบจำลอง PFM (Presence, Frequency, Monetary Value) และในแบบจำลองนี้มีการใช้อัลกอริทึมการจัดกลุ่มประเภทสินค้า ซึ่งประกอบด้วยเทคนิค K-Means K-Medoids และ Self-Organizing Maps (SOM) ในการตรวจสอบผลลัพธ์ของอัลกอริทึมการจัดกลุ่ม และเลือกกลุ่มที่ให้คุณภาพดีที่สุด และในท้ายที่สุดอัลกอริทึมออริโอริถูกใช้ในการหาความสัมพันธ์ระหว่างสินค้าและกลุ่มประเภทของลูกค้าเพื่อประเมินความถูกต้องของอัลกอริทึมที่ใช้ k-mean, k-medoids และ Self-Organizing Maps (SOM) กฎการจำแนกประเภทถูกสร้างขึ้นโดยพิจารณาจากกลุ่มที่สร้างโดยอัลกอริทึมที่กล่าวถึงในงานวิจัยนี้เป็นแอตทริบิวต์การตัดสินใจ นอกจากนี้ จากระดับการทำนาย ผลลัพธ์ยังชี้ให้เห็นว่าการจำแนกกลุ่มที่สร้างโดย CLARA ของอัลกอริทึม K-Medoids ให้ความแม่นยำที่สูง กลุ่มลูกค้าของบริษัทที่อยู่ในภาคธุรกิจ SME ที่ได้รับจากการใช้เทคนิค Data Mining เผยให้เห็นระดับความภักดี: สูงปานกลาง ต่ำ และต่ำมาก ผลลัพธ์เหล่านี้จะช่วยให้บริษัทสามารถพัฒนากลยุทธ์การรักษาลูกค้าต่อไปได้ การประยุกต์ใช้อัลกอริทึมเพื่อหาความสัมพันธ์ในชุดข้อมูลการซื้อสินค้าของลูกค้า สามารถอธิบายความสัมพันธ์ด้วยระดับความเชื่อมั่นที่สูง

Tanantong and Ramjan (2021) ได้เสนอกรอบพัฒนาการค้นหาความสัมพันธ์ที่เกี่ยวข้องกับการตลาดผ่านสื่อสังคมบนแพลตฟอร์ม Twitter และทำความสะอาดข้อมูล และจัดกลุ่มข้อมูลเป็น 5 กลุ่ม และทำการค้นหาความสัมพันธ์ด้วยอัลกอริทึม Apriori เพื่อกำหนดวัตถุประสงค์

ที่ถูกใช้งานบ่อยและสกัดกฎความสัมพันธ์อย่างเฉพาะเจาะจง อาทิเช่น จำนวนการรีทวีตและกดถูกใจ เพื่อเลือกเป็นกฎความสัมพันธ์ที่มีประสิทธิภาพสูง ผลการวิเคราะห์พบว่ากฎความสัมพันธ์นี้สามารถวิเคราะห์ต่อไปเพื่อออกแบบการปรับปรุงในการตลาดผ่านสื่อสังคมได้

Zhang (2021) ได้ทำการศึกษากฎความสัมพันธ์ระหว่างอัตราส่วนการเงินและความเสี่ยงในการล้มละลายของกิจการขนาดเล็กและกลางของกิจการ P2P Lending โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ด้วยอัลกอริทึม Apriori การศึกษาพบว่าตัวบ่งชี้บางตัว เช่น อัตราส่วนทุนหมุนเวียน (Current Ratio) อัตราส่วนทุนหมุนเวียนเร็ว (Quick Ratio) อัตราผลตอบแทนจากสินทรัพย์ (ROA) และอัตราส่วนเงินสด (Cash Ratio) มีความสัมพันธ์ที่สำคัญกับความเสี่ยงในการล้มละลายของกิจการ เมื่ออัตราส่วนเหล่านี้ต่ำกว่าระดับหนึ่ง นั้นแสดงถึงโอกาสที่สูงขึ้นของปัญหาการเงินสำหรับกิจการนั้น ผลการวิจัยนี้สามารถนำมาใช้ในการช่วยพัฒนาระบบเตือนภาวะความเสี่ยงทางการเงินล่วงหน้าสำหรับกิจการเหล่านี้ ในบริบทของการให้กู้เงินแบบ P2P (Peer-to-Peer) การทำกฎเชื่อมโยงสามารถให้ระบุรูปแบบในข้อมูลที่จะช่วยในการระบุความเสี่ยงที่เป็นไปได้

S. Datta and K. Mali (2021) เสนอกรอบพัฒนาการค้นหากฎความสัมพันธ์ที่มีความสำคัญและความเชื่อมโยงสูง (High associability) ด้วยวิธีการ SARMHA ซึ่งถูกพัฒนาขึ้นโดยใช้ค่าสนับสนุน (Support), ค่าความเชื่อมั่น (Confidence), และการไม่เชื่อมโยงที่ยืดหยุ่น (Flexible Dissociation) ผลลัพธ์แสดงให้เห็นว่ากรอบพัฒนาการค้นหากฎความสัมพันธ์ที่ใช้อัลกอริทึม SARHMA โดยอาศัยค่าการไม่เชื่อมโยงที่ยืดหยุ่น (Flexible Dissociation) มีประสิทธิภาพ ซึ่งพบว่ากฎความสัมพันธ์ที่ได้จากการหาด้วยวิธีการอื่นนั้น จะไม่ถูกคัดเลือกเมื่อใช้อัลกอริทึม SARMHA เนื่องจากกฎความสัมพันธ์นั้นมีค่าความเชื่อมโยงที่ต่ำ

Kareem et al., (2017) เสนอกรอบพัฒนาการค้นหากฎความสัมพันธ์สำหรับการใช้ในการตรวจสอบการทุจริตจากการक्रमประกันสุขภาพ ในการศึกษาจะนำวิธีการรวมกลุ่มที่กำลังพัฒนาอย่างรวดเร็ว (Evolving Clustering Method, ECM) เป็นเทคนิคที่ไม่ต้องมีการกำหนดกลุ่มล่วงหน้า ในส่วนของการทดลอง เพื่อจัดกลุ่มรูปแบบการक्रमประกันสุขภาพเพื่อนำไปใช้ในการหาความสัมพันธ์ นอกจากนี้เมื่อได้กฎความสัมพันธ์ที่น่าสนใจแล้วนั้น ได้ทำการวิเคราะห์ข้อมูลด้วยการอัลกอริทึม SVM (Support Vector Machine) ผลลัพธ์ที่ได้มีความน่าสนใจสำหรับการตรวจจับการฉ้อโกงของการทุจริตจากการक्रमประกันซึ่งยังไม่ได้รับการศึกษาอย่างแพร่หลายในด้านประกันสุขภาพ

Bao et al., (2022) มองว่าโครงสร้างเชิงดัชนีแบบเดิมในการประเมินกฎความสัมพันธ์ จะประกอบด้วยค่าสนับสนุน (Support) และ ค่าความเชื่อมั่น (Confidence) เป็นตัววัดความสำคัญและความแม่นยำของกฎความสัมพันธ์ แต่มีข้อเสียหลายประการส่งผลให้งานวิจัยส่วนใหญ่นำเสนอเทคนิควิธีการใหม่ๆ ซึ่งวิธีที่เป็นที่นิยมมากที่สุดคือ Lift, Improvement, Validity,

Conviction, Chi-square analysis เป็นต้น จึงได้ทำการวิเคราะห์ข้อดีและข้อเสียของตัวชี้วัดที่ใช้กันทั่วไปในการวิเคราะห์กฎความสัมพันธ์ พร้อมทั้งนำเสนอ 4 ตัวชี้วัดใหม่ (Bi-support, Bi-lift, Bi-improvement, และ Bi-confidence) ที่พัฒนาขึ้นจากการวิเคราะห์ดังกล่าว ในท้ายที่สุดงานวิจัยนี้เสนอโครงสร้างใหม่ของการวัดความน่าสนใจแบบสองทิศทางเพื่อปรับปรุงโครงสร้างดัชนีแบบเดิม ซึ่งผลลัพธ์ที่ได้สามารถสรุปว่า โครงสร้างการวัดความน่าสนใจแบบสองทิศทางมีประสิทธิภาพที่ดีกว่าวิธีการแบบดั้งเดิม



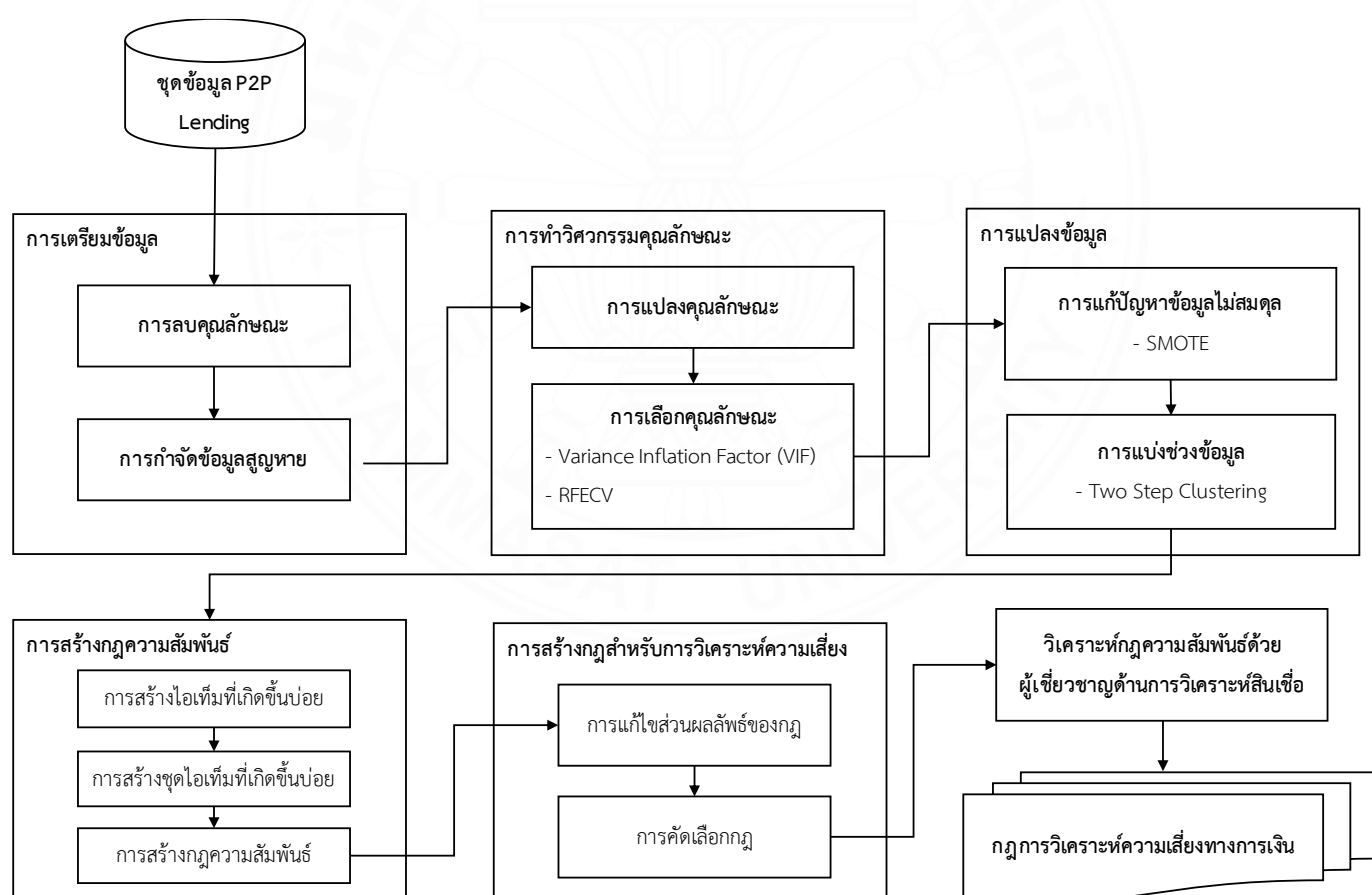
บทที่ 3

วิธีการดำเนินงานวิจัย

การดำเนินงานวิจัยประกอบไปด้วย 6 ขั้นตอน ประกอบด้วย การเตรียมข้อมูล การทำวิศวกรรมคุณลักษณะ การแปลงข้อมูล การสร้างกฎความสัมพันธ์ การสร้างกฎสำหรับการวิเคราะห์ความเสี่ยง วิเคราะห์กฎความสัมพันธ์ด้วยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ ซึ่งแสดงดังภาพที่ 3.1

ภาพที่ 3.1

กรอบพัฒนาวิธีหากฎความสัมพันธ์โดยอาศัยการเลือกคุณลักษณะ และการแบ่งช่วงข้อมูล



3.1 การเตรียมข้อมูล (Data Pre-Processing)

การเตรียมข้อมูล (Data pre-processing) เป็นขั้นตอนสำคัญสำหรับกรอบพัฒนาวิธีหาความสัมพัทธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล ในงานวิจัยนี้ได้ใช้ข้อมูลการให้สินเชื่อออนไลน์ระหว่างบุคคลจากแพลตฟอร์มการให้กู้ยืมแบบ P2P Lending ที่ใหญ่ที่สุดของสหรัฐอเมริกาที่มีชื่อว่า LendingClub (Yash, 2020) โดยใช้ข้อมูลในช่วงพ.ศ. 2550-2563 ซึ่งข้อมูลทั้งหมดมีจำนวนผู้กู้ทั้งหมด 2,925,493 คน และมีจำนวนคุณลักษณะทั้งหมด 141 คอลัมน์ ตัวอย่างของข้อมูลที่ใช้ในงานวิจัยนี้จะแสดงในตารางที่ 3.1

ตารางที่ 3.1

ตารางตัวอย่างข้อมูลสินเชื่อออนไลน์ระหว่างบุคคลจากแพลตฟอร์ม LendingClub (Yash, 2020)

<i>id</i>	<i>loan_amnt</i>	<i>deferral_term</i>	<i>revol_bal_joint</i>	<i>dti_joint</i>	<i>zip_code</i>	<i>addr_state</i>
000203532	21000	n/a	13343	n/a	606xx	IL
000663451	6400	n/a	n/a	n/a	365xx	AL
000443279	4000	n/a	35827	n/a	775xx	TX
000742186	11500	n/a	30324	n/a	238xx	VA
000597005	8000	n/a	n/a	n/a	070xx	NJ
000997918	17400	n/a	47995	n/a	628xx	IL
000648213	20000	n/a	0	17.67	021xx	MA
000792768	28000	2	n/a	n/a	117xx	NY
000762087	11200	n/a	10052	n/a	145xx	NY
000847527	19075	n/a	n/a	n/a	481xx	MI
000749019	3000	n/a	n/a	17.92	617xx	IL
000569379	7750	2	26220	19.02	430xx	OH
000406211	9000	n/a	n/a	n/a	774xx	TX
000494432	18000	n/a	15490	n/a	787xx	TX

ในขั้นตอนการเตรียมข้อมูลจะแบ่งเป็น 2 กระบวนการ ประกอบด้วย การลบคุณลักษณะ และการกำจัดข้อมูลสูญหาย ดังแสดงรายละเอียดดังต่อไปนี้

3.1.1 การลบคุณลักษณะ (Attribute Removal)

ในกระบวนการลบคุณลักษณะ ผู้วิจัยได้ทำการลบคุณลักษณะบางประการที่ไม่จำเป็นต่อการหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล อาทิเช่น เลขประจำตัว รหัสไปรษณีย์ และเมืองที่อยู่อาศัย เป็นต้น รวมไปถึงการกำจัดคุณลักษณะที่มีข้อมูลสูญหายมากกว่า 50% ซึ่งมีจำนวน 44 คุณลักษณะ เพื่อลดความซับซ้อน ลดการใช้เวลาในการวิเคราะห์ข้อมูล และเพื่อให้ง่ายต่อการตีความ

3.1.2 การกำจัดข้อมูลสูญหาย (Missing Value Removal)

ในกระบวนการนี้จะทำการกำจัดข้อมูลผู้ขอสินเชื่อที่มีข้อมูลสูญหายทั้งหมด โดยไม่ใช้วิธีการแทนที่ข้อมูลสูญหาย(Missing Value Replacement) เนื่องจากทางผู้วิจัยไม่ทราบถึงค่าข้อมูลสูญหายที่แท้จริง ทำให้การแทนที่ข้อมูลสูญหายอาจจะทำให้การวิเคราะห์ผิดพลาดได้ ในการเตรียมข้อมูลสำหรับใช้ในกระบวนการหาความสัมพันธ์ ผู้วิจัยได้คัดเลือกสถานะหนี้ (*loan_status*) ซึ่งเป็นตัวแปรหลักที่ใช้ในการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล จำนวน 2 ประเภท ได้แก่ กลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid) และ กลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน (Charge Off)

ข้อมูลในท้ายที่สุดหลังจากผ่านกระบวนการเตรียมข้อมูลมีจำนวนรวม 672,237 รายการ แบ่งเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid) จำนวน 538,917 คน และ กลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน (Charge Off) จำนวน 133,320 คน

3.2 การทำวิศวกรรมคุณลักษณะ (Feature Engineering)

การทำวิศวกรรมคุณลักษณะ (Feature Engineering) เป็นกระบวนการที่เกี่ยวข้องกับการสกัดหรือสร้างคุณลักษณะใหม่ขึ้นมา เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์ และช่วยให้ผลลัพธ์ที่ได้มีความหมายและสามารถอธิบายได้ดียิ่งขึ้น ซึ่งในงานวิจัยนี้การทำวิศวกรรมคุณลักษณะได้ประกอบด้วย 2 ขั้นตอนหลัก ได้แก่ การแปลงคุณลักษณะ (Feature Transformation) และ การเลือกคุณลักษณะ (Feature Selection) (Novakovic, 2016)

3.2.1 การแปลงคุณลักษณะ (Feature Transformation)

จากการตรวจสอบข้อมูล LendingClub (Yash, 2020) พบว่าคุณลักษณะบางประการหนึ่ง คุณลักษณะถูกแทนด้วยสองคุณลักษณะเพื่อเป็นตัวแทนของค่าต่ำสุดและสูงสุดของช่วงข้อมูล ได้แก่ `fico_range_low` และ `fico_range_high` ซึ่งเป็นตัวแทนของขอบล่าง และขอบบนของคะแนน FICO SCORE ตามลำดับ นอกจากนี้ยังมีคุณลักษณะอื่นๆ เช่น `last_fico_range_low` และ `last_fico_range_high` ซึ่งเป็นการกำหนดช่วงข้อมูลโดยใช้ 2 คุณลักษณะเช่นเดียวกัน ในครั้งนี้เพื่อเป็นการปรับปรุงคุณภาพในการวิเคราะห์ ลดความซ้ำซ้อนของข้อมูล และช่วงให้ผลลัพธ์ที่ได้สามารถอธิบายได้ดียิ่งขึ้น ในการศึกษาครั้งนี้ผู้วิจัยได้ใช้กระบวนการแปลงคุณลักษณะ โดยการสร้างคุณลักษณะใหม่โดยใช้ค่าเฉลี่ยของแต่ละช่วงข้อมูลระหว่างสองคุณลักษณะ และกำหนดคุณลักษณะใหม่ในชื่อ `fico_score` และ `last_fico_score` ตามลำดับ

3.2.2 การเลือกคุณลักษณะ (Feature Selection)

1. ในขั้นตอนแรกของการเลือกคุณลักษณะ คุณลักษณะจำนวน 21 คุณลักษณะที่มีค่าซ้ำกันมากกว่า 70% ถูกลบออก ซึ่งคุณลักษณะเหล่านี้บ่งชี้ถึงการกระจายตัวที่ไม่มีประสิทธิภาพ ทำให้ผลลัพธ์ที่ได้จากการหาความสัมพันธ์ที่เป้าหมายของงานวิจัยสามารถทำได้ ตัวอย่างคุณลักษณะ เช่น `pub_rec_bankruptcies`, `recovery`, `collection_recovery_fee` เป็นต้น การลบคุณลักษณะเหล่านี้ออก จะช่วยปรับปรุงและตีความผลลัพธ์จากการหาความสัมพันธ์ได้ดีมากยิ่งขึ้น

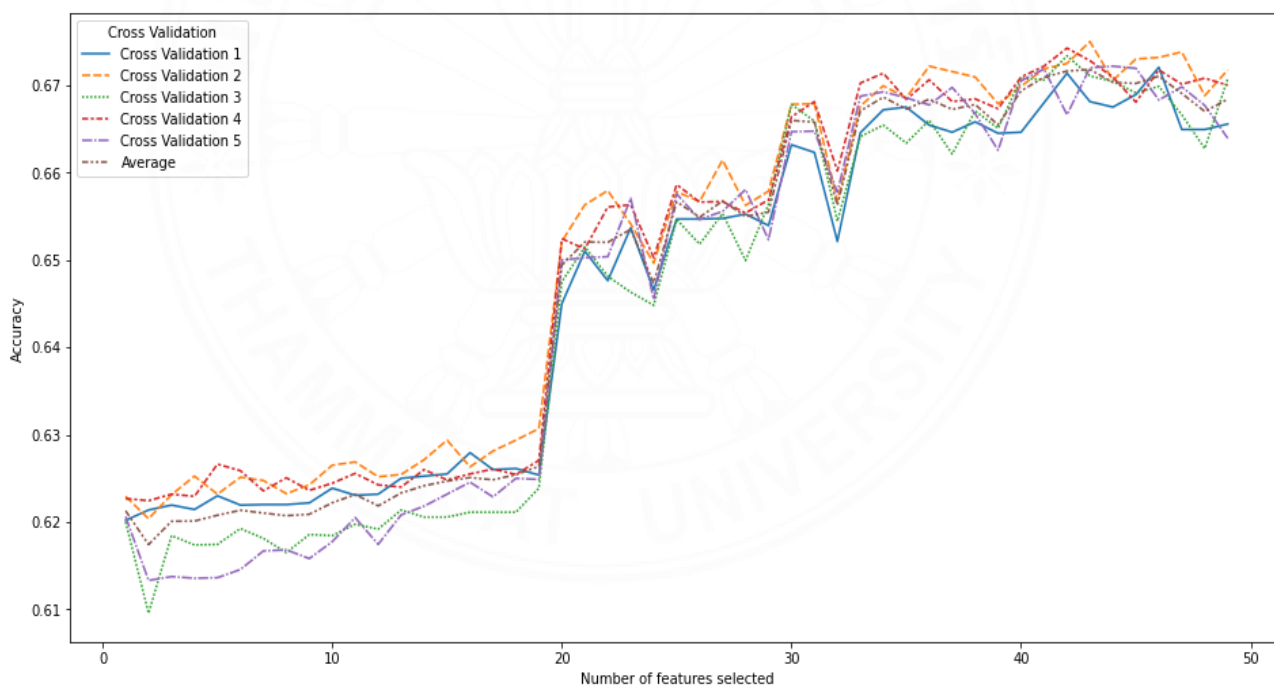
2. ค่า Variance Inflation Factor (VIF) เป็นขั้นตอนที่สองในการเลือกคุณลักษณะ โดยมีวัตถุประสงค์เพื่อกำหนดให้คุณลักษณะแต่ละตัวต้องเป็นอิสระต่อกัน ซึ่งหากคุณลักษณะต่างๆ ที่นำมาใช้สำหรับการหาความสัมพันธ์ระหว่างกันนั้นมีความสัมพันธ์กันในเกณฑ์ที่สูง หรือที่เรียกว่า การเกิด Multicollinearity จะส่งผลต่อการซ้ำซ้อนของข้อมูล เพื่อแก้ไขปัญหานี้เราจึงได้ใช้ค่าสถิติ Variance Inflation Factor ซึ่งหากค่า VIF มีค่ามากกว่า 5 แสดงว่า ระดับความสัมพันธ์ของคุณลักษณะที่ใช้มีมากเกินไป นั่นคือเกิดปัญหา Multicollinearity ผู้วิจัยจึงได้ทำการลบคุณลักษณะบางส่วนเพื่อแก้ปัญหา Multicollinearity

3. หลังจากการแก้ปัญหา Multicollinearity ผู้วิจัยได้ใช้วิธีการเลือกคุณลักษณะด้วยวิธีการ Recursive Feature Elimination with Cross-Validation (RFECV) (Lee, KKG et al., 2023) เพื่อกำจัดคุณลักษณะที่ไม่สำคัญ และมีความซ้ำซ้อน เพื่อให้คงเหลือเฉพาะคุณลักษณะจำนวนน้อยที่สุดที่ยังมีประสิทธิภาพในการทำนายที่สูง สำหรับการใส่ 5-Fold Cross-Validation ได้ทำการแบ่งชุดข้อมูลทดสอบเป็น 5 ส่วนเท่ากันๆ เพื่อป้องกันปัญหา Overfitting และเพื่อให้ได้มาซึ่งจำนวนของคุณลักษณะที่เหมาะสมที่สุด

ในงานวิจัยนี้ได้ใช้แบบจำลอง LightGBM ในการทำนายสถานะเงินกู้ เมื่อใช้วิธี RFECV ดังแสดงในรูปภาพที่ 3.2 ซึ่งจะเห็นได้ว่าแบบจำลองให้ความแม่นยำในการทำนายสูงที่สุดที่ 68% เมื่อเลือกคุณลักษณะที่ใช้มีจำนวน 43 คุณลักษณะ แต่อย่างไรก็ตาม เนื่องจากทรัพยากรที่ใช้ในการประมวลผลบน Google Colab Pro มีจำกัด ผู้วิจัยจึงเลือกคุณลักษณะทั้งหมด 41 คุณลักษณะ ซึ่งยังคงให้ผลการทำนายที่แม่นยำใกล้เคียงกับค่าสูงสุด ในส่วนของค่าความสำคัญของคุณลักษณะที่ใช้ในการทำนายความสามารถในการชำระหนี้ ผ่านแบบจำลอง LightGBM แสดงในภาพที่ 3.3 และในส่วนของคำอธิบายข้อมูลคุณลักษณะจำนวน 41 คุณลักษณะ แสดงในตารางที่ 3.2

ภาพที่ 3.2

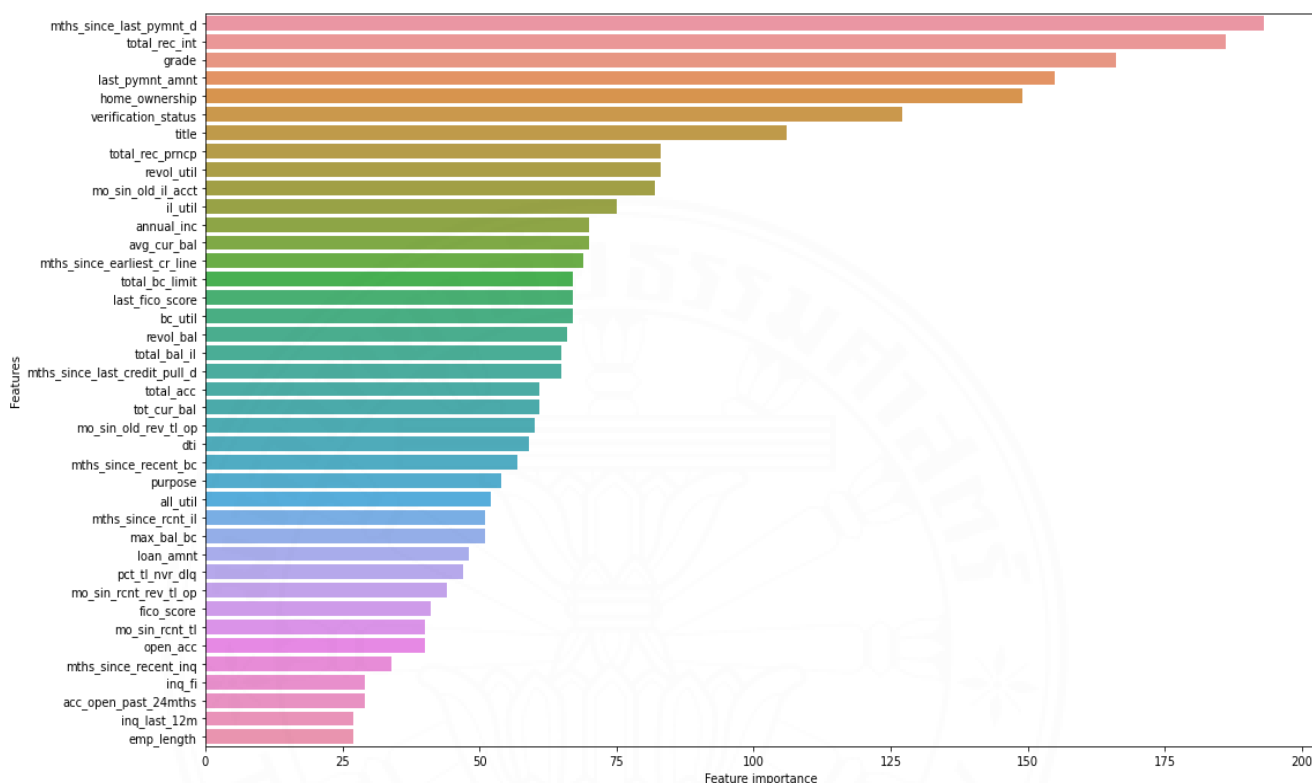
กราฟแสดงจำนวนคุณลักษณะต่อความแม่นยำในการทำนายสถานะเงินกู้ ด้วยแบบจำลอง LightGBM ด้วยวิธีการ Recursive Feature Elimination with 5-Cross-Validation



ภาพที่ 3.3

แผนภูมิแสดงค่าความสำคัญของคุณลักษณะที่ส่งผลต่อการทำนายสถานะเงินกู้ ผ่านแบบจำลอง

LightGBM



ตารางที่ 3.2

ตารางแสดงรายการคุณลักษณะข้อมูลที่ใช้ในงานวิจัย

ชื่อคุณลักษณะ	คำอธิบาย
mths_since_last_pymnt_d	จำนวนเดือนตั้งแต่รับการชำระเงินครั้งล่าสุด
total_rec_int	ดอกเบี้ยที่ได้รับจนถึงปัจจุบัน
grade	เกรดของสัญญากู้ยืมที่ได้รับการกำหนดโดย Lending Club
last_pymnt_amnt	จำนวนเงินทั้งหมดที่ได้รับจากการชำระเงินครั้งสุดท้าย
home_ownership	สถานะการเป็นเจ้าของบ้านซึ่งระบุโดยผู้กู้หรือได้รับจากรายงานเครดิต ค่าความหมายของเราคือ: เช่า, เป็นเจ้าของ, จดจำนอง, อื่น ๆ
verification_status	ผู้ขอสินเชื่อได้รับการตรวจสอบแหล่งที่มาของรายได้
title	ชื่อสัญญากู้ยืมที่ระบุโดยผู้กู้
total_rec_pncp	เงินต้นที่ได้รับจนถึงปัจจุบัน
revol_util	อัตราการใช้วงเงินเครดิตบัตรเครดิตหมุนเวียน
mo_sin_old_il_acct	จำนวนเดือนตั้งแต่บัญชีผ่อนชำระธนาคารเปิดใช้งานครั้งแรก

ตารางที่ 3.2

ตารางแสดงรายการคุณลักษณะข้อมูลที่ใช้ในงานวิจัย (ต่อ)

ชื่อคุณลักษณะ	คำอธิบาย
il_util	อัตราส่วนของยอดคงเหลือปัจจุบันต่อเครดิตสูงสุด/วงเงินเครดิตสำหรับบัญชีผ่อนชำระธนาคารทั้งหมด
annual_inc	รายได้ทั้งปีที่ผู้กู้ยื่นขอในขณะที่ลงทะเบียนกับเว็บไซต์ของ LendingClub
avg_cur_bal	ยอดคงเหลือเฉลี่ยของบัญชีทั้งหมด
mths_since_earliest_cr_line	จำนวนเดือนตั้งแต่เปิดบัญชีเครดิตครั้งแรกของผู้กู้
total_bc_limit	วงเงินเครดิตสูงสุดของบัตรเครดิตทั้งหมด
last_fico_score	คะแนน FICO ล่าสุดของผู้กู้เฉลี่ยระหว่างช่วงคะแนนต่ำสุดและสูงสุด
bc_util	อัตราส่วนของยอดคงเหลือปัจจุบันถึงวงเงินเครดิตสูงสุดสำหรับบัตรเครดิตทั้งหมด
revol_bal	อัตราส่วนของยอดคงเหลือปัจจุบันถึงวงเงินเครดิตสูงสุดสำหรับบัตรเครดิตทั้งหมด
total_bal_il	ยอดคงเหลือเฉลี่ยของบัญชีผ่อนชำระทั้งหมด
mths_since_last_credit_pull_d	จำนวนเดือนตั้งแต่ LC ดึงข้อมูลเครดิตครั้งล่าสุดสำหรับสินเชื่อ
total_acc	จำนวนเครดิตไลน์ทั้งหมดในไฟล์เครดิตของผู้กู้
tot_cur_bal	ยอดคงเหลือเทียบเท่าของบัญชีทั้งหมด
mo_sin_old_rev_tl_op	จำนวนเดือนตั้งแต่บัญชีหมุนเวียนเปิดใช้งานครั้งแรก
dti	อัตราส่วนที่คำนวณจากการชำระหนี้รายเดือนของผู้กู้ต่อหนี้สินรวมทั้งหมดที่ยกเว้นเงินกู้บ้านและเงินกู้จาก LC ที่ขอ
mths_since_recent_bc	จำนวนเดือนตั้งแต่บัตรเครดิตบัญชีล่าสุดถูกเปิดใช้งาน
purpose	หมวดหมู่ที่ผู้ขอสินเชื่อระบุสำหรับคำขอสินเชื่อ
all_util	ยอดคงเหลือต่อวงเงินเครดิตสูงสุดบนธุรกรรมทั้งหมด
mths_since_rcnt_il	จำนวนเดือนตั้งแต่บัญชีผ่อนชำระครั้งล่าสุดถูกเปิดใช้งาน
max_bal_bc	ยอดคงเหลือสูงสุดที่ต้องชำระบนบัญชีหมุนเวียนทั้งหมด
loan_amnt	จำนวนเงินที่ขอกู้ตามที่ระบุไว้โดยผู้กู้
pct_tl_nvr_dlq	เปอร์เซ็นต์ของธุรกรรมที่ไม่เคยผิดนัดชำระ
mo_sin_rcnt_rev_tl_op	จำนวนเดือนตั้งแต่บัญชีหมุนเวียนครั้งล่าสุดถูกเปิดใช้งาน
mo_sin_old_il_acct	จำนวนเดือนตั้งแต่บัญชีผ่อนชำระธนาคารเปิดใช้งานครั้งแรก
fico_score	ค่าเฉลี่ยระหว่างช่วงคะแนน FICO ต่ำสุดและสูงสุดของผู้กู้เงินที่ขอสินเชื่อ
mo_sin_rcnt_tl	จำนวนเดือนตั้งแต่บัญชีล่าสุดถูกเปิดใช้งาน
open_acc	จำนวนบัญชีเครดิตที่เปิดอยู่ในไฟล์เครดิตของผู้ขอสินเชื่อ
mths_since_recent_inq	จำนวนเดือนตั้งแต่การสอบถามล่าสุด
inq_fi	จำนวนการสอบถามการเงินส่วนบุคคล
acc_open_past_24mths	จำนวนการซื้อขายที่เปิดใช้งานในช่วง 24 เดือนที่ผ่านมา
inq_last_12m	จำนวนการสอบถามเครดิตในช่วง 12 เดือนที่ผ่านมา
emp_length	อายุงาน (0 หมายถึงอายุงานที่น้อยกว่า 1 ปี และ 10 หมายถึง อายุงาน 10 ปีขึ้นไป)
loan_status	สถานะปัจจุบันของสินเชื่อ

3.3 การแปลงข้อมูล (Data Transformation)

3.3.1. การแก้ปัญหาข้อมูลไม่สมดุล (Data Balancing)

การศึกษานี้มีวัตถุประสงค์เพื่อหาความสัมพันธ์ของผู้ขอสินเชื่อต่อแพลตฟอร์มการให้สินเชื่อออนไลน์ระหว่างบุคคล ซึ่งนำไปสู่การวิเคราะห์ความเสี่ยงเพื่อใช้ประเมินความน่าเชื่อถือของผู้ขอสินเชื่อ จึงมีการมุ่งเน้นไปยังคุณลักษณะสถานะเงินกู้ซึ่งมีความไม่สมดุลของข้อมูล ระหว่างกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid) จำนวน 538,917 คน และ กลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน (Charge Off) จำนวน 133,320 คน เพื่อจัดการความไม่สมดุลของข้อมูล ในงานวิจัยนี้จึงได้ใช้วิธีการสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) โดยการสุ่มเลือกข้อมูลกลุ่มน้อย ให้มีจำนวนเท่ากับข้อมูลกลุ่มมาก ในที่นี้คือกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน (Charge Off) ให้มีจำนวนเท่ากับกลุ่มที่มีจำนวนมาก คือ กลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid) ทำให้ข้อมูลมีความสมดุล ไม่มีความเอนเอียง และช่วยให้ใช้ในการวิเคราะห์ได้ดียิ่งขึ้น (Chen and Suzhou, 2021) ดังแสดงในตารางที่ 3.3

ตารางที่ 3.3

ตารางแสดงสัดส่วนข้อมูลสถานะเงินกู้ ก่อนและหลังการแปลงข้อมูลด้วยวิธีการ SMOTE

ก่อนการแก้ไขปัญหาข้อมูลไม่สมดุล			หลังการแก้ไขปัญหาข้อมูลไม่สมดุล		
สถานะสินเชื่อ	จำนวนข้อมูล	เปอร์เซ็นต์	สถานะสินเชื่อ	จำนวนข้อมูล	เปอร์เซ็นต์
Fully Paid	538,917	80.17%	Fully Paid	538917	50.00%
Charge Off	133,320	19.83%	Default	538917	50.00%
รวม	672,237	100.00%	รวม	1,077,834	100.00%

3.3.2 การแบ่งช่วงข้อมูล (Discretization)

ในการแบ่งช่วงข้อมูล ในงานวิจัยนี้ได้ใช้วิธีการแบ่งกลุ่มสองขั้นตอน (2-Step Clustering Method) ในโปรแกรม SPSS ซึ่งเป็นเทคนิคที่มีประสิทธิภาพในการจัดกลุ่มข้อมูล เมื่อเราไม่สามารถกำหนดจำนวนกลุ่ม หรือ จำนวนช่วงของข้อมูลได้ ซึ่งเทคนิคจะทำการจัดกลุ่มโดยอาศัยลักษณะของข้อมูลเป็นตัวกำหนด (S. C. Tan, 2018)

วิธีการแบ่งกลุ่มสองขั้นตอนประกอบด้วย ขั้นตอนแรกคือการวิเคราะห์จัดกลุ่มตามลำดับชั้น (hierarchical clustering analysis) เพื่อกำหนดจำนวนกลุ่มสูงสุดโดยใช้วัดระยะห่าง Log-likelihood และจะใช้เกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criterion : BIC) เพื่อประเมินผลลัพธ์การจัดกลุ่ม โดยเมื่อได้จำนวนกลุ่มสูงสุดแล้ว ขั้นตอนที่สองนี้จะใช้วิธีการแบ่งกลุ่มแบบเคมีน (K-means Clustering) โดยใช้จำนวนกลุ่มที่กำหนดไว้ ขั้นตอน k-means algorithm จะแบ่งข้อมูลเป็น k กลุ่มโดยทุกๆ กลุ่มจะถูกกำหนดให้เป็นกลุ่มที่มีค่าเฉลี่ยใกล้เคียงกัน ข้อมูลที่ได้หลังจากการทำการแบ่งช่วงข้อมูลด้วยวิธีการ 2-Step Clustering จะแสดงในตารางที่ 3.4

ตารางที่ 3.4

ตารางแสดงข้อมูลเชิงตัวเลขหลังจากทำการแบ่งช่วงข้อมูล

ชื่อคุณลักษณะ	ช่วงของข้อมูล	จำนวนกลุ่ม	ผลลัพธ์ที่ได้จากการแบ่งช่วงข้อมูล
mths_since_last_pymnt_d	[0, 56]	2	[0, 22.92] = mths_since_last_pymnt_d_low [22.93, 56] = mths_since_last_pymnt_d_high
total_rec_int	[0, 31714.37]	2	[0, 3122.73] = total_rec_int_low [3122.74, 31714.37] = total_rec_int_high
last_pymnt_amnt	[0, 42192.05]	2	[0, 7707.31] = last_pymnt_amnt_low [7707.32, 42192.05] = last_pymnt_amnt_high
total_rec_prncp	[0, 40000]	3	[0, 8112.99] = total_rec_prncp_low [8113, 21537.74] = total_rec_prncp_medium [21537.75, 40000] = total_rec_prncp_high
revol_util	[0, 165.8]	3	[0, 37.12] = revol_util_low [37.13, 62.38] = revol_util_medium [62.39, 165.8] = revol_util_high
mo_sin_old_il_acct	[0, 999]	3	[0, 103.57] = mo_sin_old_il_acct_low [103.58, 178.75] = mo_sin_old_il_acct_medium [178.76, 999] = mo_sin_old_il_acct_high
il_util	[0, 464]	4	[0, 49.036] = il_util_low [49.037, 73.177] = il_util_medium [73.178, 93.354] = il_util_high [93.355, 464] = il_util_very high

ตารางที่ 3.4

ตารางแสดงข้อมูลเชิงตัวเลขหลังจากทำการแบ่งช่วงข้อมูล (ต่อ)

ชื่อคุณลักษณะ	ช่วงของข้อมูล	จำนวน กลุ่ม	ผลลัพธ์ที่ได้จากการแบ่งช่วงข้อมูล
annual_inc	[20, 10999200]	4	[20, 83824] = annual_inc_low [83825, 160799] = annual_inc_medium [160800, 436761] = annual_inc_high [436762, 10999200] = annual_inc_very high
avg_cur_bal	[12, 425387]	2	[12, 18010] = avg_cur_bal_low [18011, 425387] = avg_cur_bal_high
mths_since_earliest_cr_line	[42, 1049]	3	[42, 211.15] = mths_since_earliest_cr_line_low [211.16, 326.814] = mths_since_earliest_cr_line_medium [326.815, 1049] = mths_since_earliest_cr_line_high
total_bc_limit	[100, 1105500]	2	[100, 34141] = total_bc_limit_low [34140, 1105500] = total_bc_limit_high
last_fico_score	[0, 639.18]	3	[0, 451.65] = last_fico_score_low [451.66, 639.18] = last_fico_score_medium [639.19, 847.5] = last_fico_score_high
bc_util	[0, 252.3]	3	[0, 40.15] = bc_util_low [40.16, 72.84] = bc_util_medium [72.85, 252.3] = bc_util_high
revol_bal	[0, 1696796]	2	[0, 39528.28] = revol_bal_low [39530, 1696796] = revol_bal_high
total_bal_il	[1, 1711009]	2	[1, 70414] = total_bal_il_low [70415, 1711009] = total_bal_il_high
mths_since_last_credit_pull_d	[0, 56]	4	[0, 7.932] = mths_since_last_credit_pull_d_low [7.933, 16.121] = mths_since_last_credit_pull_d_medium [16.122, 27.174] = mths_since_last_credit_pull_d_high [27.175, 56] = mths_since_last_credit_pull_d_very high
total_acc	[3, 176]	2	[3, 26.9] = total_acc_low [27, 176] = total_acc_high
tot_cur_bal	[52, 5445012]	4	[52, 94272] = tot_cur_bal_low [94273, 229132] = tot_cur_bal_medium [229131, 423093] = tot_cur_bal_high [423094, 5445012] = tot_cur_bal_very high

ตารางที่ 3.4

ตารางแสดงข้อมูลเชิงตัวเลขหลังจากทำการแบ่งช่วงข้อมูล (ต่อ)

ชื่อคุณลักษณะ	ช่วงของข้อมูล	จำนวนกลุ่ม	ผลลัพธ์ที่ได้จากการแบ่งช่วงข้อมูล
mo_sin_old_rev_tl_op	[1, 902]	3	[1, 131.86] = mo_sin_old_rev_tl_op_low [131.87, 230.70] = mo_sin_old_rev_tl_op_medium [230.71, 902] = mo_sin_old_rev_tl_op_high
dti	[0, 999]	5	[0, 14.54] = dti_very low [14.55, 21.60] = dti_low [21.61, 29.79] = dti_medium [29.80, 54.81] = dti_high [54.82, 999] = dti_very high
mths_since_recent_bc	[0, 569]	2	[0, 46.56] = mths_since_recent_bc_low [46.57, 569] = mths_since_recent_bc_high
all_util	[0, 204]	4	[0, 42.48] = all_util_low [42.49, 61.24] = all_util_medium [61.25, 77.37] = all_util_high [77.38, 204] = all_util_very high
mths_since_rcnt_il	[0, 426]	7	[0, 8.13] = mths_since_rcnt_il_extreamly low [8.14, 14.62] = mths_since_rcnt_il_very low [14.63, 22.28] = mths_since_rcnt_il_low [22.29, 32.88] = mths_since_rcnt_il_medium [32.89, 51.89] = mths_since_rcnt_il_high [51.90, 87.25] = mths_since_rcnt_il_very high [87.26, 426] = mths_since_rcnt_il_extreamly high
max_bal_bc	[0, 776843]	2	[0, 9555] = max_bal_bc_low [9556, 776843] = max_bal_bc_high
loan_amnt	[1000, 40000]	3	[1000, 12651] = loan_amnt_low [12652, 23352] = loan_amnt_medium [23353, 40000] = loan_amnt_high
pct_tl_nvr_dlq	[5, 100]	2	[5, 89.69] = pct_tl_nvr_dlq_low [89.70, 100] = pct_tl_nvr_dlq_high
mo_sin_rcnt_rev_tl_op	[0, 404]	2	[0, 24.39] = mo_sin_rcnt_rev_tl_op_low [24.40, 404] = mo_sin_rcnt_rev_tl_op_high
fico_score	[662, 847.5]	2	[662, 705.95] = fico_score_low [705.96, 847.5] = fico_score_high

ตารางที่ 3.4

ตารางแสดงข้อมูลเชิงตัวเลขหลังจากทำการแบ่งช่วงข้อมูล (ต่อ)

ชื่อคุณลักษณะ	ช่วงของข้อมูล	จำนวนกลุ่ม	ผลลัพธ์ที่ได้จากการแบ่งช่วงข้อมูล
mo_sin_rcnt_tl	[0, 27.60]	7	[0, 3.44] = mo_sin_rcnt_tl_extreamly low [3.44, 6.34] = mo_sin_rcnt_tl_very low [6.35, 9.36] = mo_sin_rcnt_tl_low [9.37, 12.89] = mo_sin_rcnt_tl_medium [12.90, 18.139] = mo_sin_rcnt_tl_high [18.14, 27.60] = mo_sin_rcnt_tl_very high [27.61, 147] = mo_sin_rcnt_tl_extreamly high
open_acc	[2, 88]	3	[2, 10.8] = open_acc_low [10.9, 18.91] = open_acc_medium [18.92, 88] = open_acc_high
mths_since_recent_inq	[0, 24]	2	[0, 7.69] = mths_since_recent_inq_low [7.70, 24] = mths_since_recent_inq_high
inq-fi	[0, 27.174]	3	[0, 7.93] = inq-fi_low [7.934, 16.12] = inq-fi_medium [16.13, 27.174] = inq-fi_high [27.175, 56] = inq-fi_very high
acc_open_past_24mths	[0, 61]	2	[0, 5.79] = acc_open_past_24mths_low [5.78, 61] = acc_open_past_24mths_high
inq_last_12m	[0, 67]	2	[0, 4.22] = inq_last_12m_low [4.23, 67] = inq_last_12m_high

3.4 การสร้างกฎความสัมพันธ์ (Association rules mining)

การสร้างกฎความสัมพันธ์ (Association Rules Mining - ARM) ด้วยอัลกอริทึมอปริโอรี (Apriori Algorithm) ถูกใช้ในงานวิจัยเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล โดยมีขั้นตอนหลัก 2 ขั้นตอน

ขั้นตอนที่ 1 จะทำการนับรายการความถี่ของรายการที่เกิดขึ้นบ่อย (Frequent Items Generation) โดยเซตของรายการต้องมีค่าความถี่มากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ (Minimum Support) ซึ่งในงานวิจัยได้กำหนดค่าสนับสนุนขั้นต่ำที่ 0.2

ขั้นตอนที่ 2 ค้นหากฎความสัมพันธ์ของกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย (Rule Generation) ที่มีค่าความเชื่อมั่นของกฎ (Confidence) ไม่น้อยกว่าค่าที่กำหนด ซึ่งในงานวิจัยนี้ได้ใช้

ค่าความเชื่อมั่นที่มากกว่า 50% เป็นเกณฑ์ในการเลือกกฎความสัมพันธ์ นอกจากนี้ยังมีการกำหนดค่าลิฟต์ (Lift) ซึ่งเป็นตัวบ่งชี้ว่าการเกิดรูปแบบว่าตัวแปรต้น และตัวแปรตามมีความสัมพันธ์กันมากเพียงใด โดยมีการกำหนดให้ค่าลิฟต์มีค่ามากกว่า 1

การกำหนดค่าในการสร้างกฎความสัมพันธ์นั้น ซึ่งประกอบไปด้วย ค่าสนับสนุน ≥ 0.2 ค่าความเชื่อมั่น $\geq 50\%$ และ ค่าลิฟต์ > 1 นั้น เป็นการคัดกรองกฎที่ไม่น่าสนใจบางส่วนออกไป ซึ่งกฎที่น่าสนใจที่ผ่านเกณฑ์ข้างต้นจะถูกนำไปประเมิน และวิเคราะห์ต่อเพื่อหาความสัมพันธ์ที่มีประโยชน์ต่อการปล่อยสินเชื่อผ่านแพลตฟอร์ม P2P Lending รวมไปถึงสามารถวิเคราะห์ความเสี่ยงจากการให้กู้ยืมเงินได้ในขั้นตอนต่อไป

3.5 การสร้างกฎสำหรับการวิเคราะห์ความเสี่ยง (Rule Construction For Risk Analysis)

จากกฎความสัมพันธ์ที่ได้จากการสร้างกฎความสัมพันธ์ในหัวข้อที่ 3.4 ทางผู้วิจัยได้ทำการเลือกผลลัพธ์จากเหตุการณ์ที่เกิดขึ้น (Consequence) ที่เกี่ยวข้องกับการวิเคราะห์ความเสี่ยงทางด้านเครดิต ซึ่งได้แก่ สถานะหนี้ (*loan_status*) และ คะแนน FICO (*fico_score*) โดยทำการเลือกกฎที่มีค่าความเชื่อมั่นที่สูงที่สุด 5 รายการ ในแต่ละระดับของค่าสนับสนุน เพื่อให้ได้กฎที่มีประสิทธิภาพต่อการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อผ่านแพลตฟอร์ม P2P Lending

3.6 วิเคราะห์กฎความสัมพันธ์ด้วยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ

จากกฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงที่ได้มาจากข้อที่ 3.5 นำมาให้ผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อทำการตีความและแปลผลเพื่อค้นหากฎความสัมพันธ์ที่สามารถนำไปใช้ประโยชน์ และวิเคราะห์ความเสี่ยงจากการให้สินเชื่อได้จริง

บทที่ 4

ผลการวิจัย

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเสนอกรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงของการให้กู้ยืมเงินผ่านแพลตฟอร์ม Peer-to-Peer Lending (P2P Lending) โดยใช้ชุดข้อมูล Lending Club ซึ่งเป็น แพลตฟอร์ม P2P Lending ที่ใหญ่ที่สุดในประเทศสหรัฐอเมริกา โดยใช้ข้อมูลในช่วงพ.ศ. 2550-2563 ซึ่งมีข้อมูลทั้งหมด 2,925,493 ข้อมูล และมีจำนวนคุณลักษณะทั้งหมด 141 คอลัมน์ ซึ่งข้อมูลบางส่วนอยู่ในรูปแบบที่ไม่มีประสิทธิภาพต่อการใช้หาความสัมพันธ์ เพื่อเป็นการเตรียมข้อมูลให้มีคุณภาพก่อนการใช้หาความสัมพันธ์จึงได้ทำการเตรียมข้อมูล, การทำวิศวกรรมคุณลักษณะ โดยใช้แบบจำลอง LightGBM ร่วมกับ Recursive Feature Elimination with 5-Cross-Validation เพื่อกำหนดจำนวนคุณลักษณะที่เหมาะสม และคัดเลือกคุณลักษณะต่างๆ จำนวน 41 คุณลักษณะที่มีผลต่อการวิเคราะห์ความเสี่ยงในการปล่อยสินเชื่อผ่านแพลตฟอร์ม P2P Lending ดังแสดงในภาพที่ 3.2 และ ภาพที่ 3.3 ในการแก้ปัญหาข้อมูลไม่สมดุลได้ใช้วิธีการสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) เพื่อเพิ่มข้อมูลกลุ่มสถานะหนี้ให้มีจำนวนที่เท่ากัน ดังตารางที่ 3.3 ซึ่งจะได้ข้อมูลรวมทั้งหมด 1,077,834 ข้อมูล เพื่อให้ข้อมูลอยู่ในรูปแบบที่สามารถสร้างหาความสัมพันธ์ด้วยอัลกอริทึมออร์โรว์ได้ ผู้วิจัยได้ทำการแบ่งช่วงข้อมูล ด้วยวิธีการแบ่งกลุ่มสองขั้นตอน (2-Step Clustering) ดังตารางที่ 3.4 ต่อมาเราได้ใช้อัลกอริทึมออร์โรว์ เพื่อสร้างหาความสัมพันธ์ด้วยค่าสนับสนุน ≥ 0.2 และ ค่าลิฟท์ >1 ซึ่งเราได้หาความสัมพันธ์รวมทั้งรวมทั้งหมด 9,158,342 หาความสัมพันธ์ และในขั้นตอนสุดท้ายเราได้ทำการค้นหาหาความสัมพันธ์ที่มีผลสถานะเงินกู้ และคะแนนเครดิต เพื่อให้ผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อสามารถใช้ในการวิเคราะห์ความเสี่ยงจากการให้กู้ยืมผ่านแพลตฟอร์ม P2P Lending ได้

4.1 หาความสัมพันธ์ที่ได้จากอัลกอริทึมออร์โรว์

จากตารางที่ 4.1 แสดงจำนวนหาความสัมพันธ์ในแต่ละช่วงค่าสนับสนุน และช่วงค่าความเชื่อมั่น ตารางนี้แสดงให้เห็นว่าเมื่อค่าสนับสนุนเพิ่มสูงขึ้น จำนวนหาความสัมพันธ์จะลดลง เช่นเดียวกันกับค่าความเชื่อมั่นหากค่าความเชื่อมั่นเพิ่มสูงขึ้น จำนวนหาความสัมพันธ์จะมีแนวโน้มที่ลดลงด้วยเช่นเดียวกัน แต่อย่างไรก็ตามจากตารางที่ 4.1 จะพบว่าหาความสัมพันธ์ส่วนใหญ่จะอยู่ในช่วงความเชื่อมั่น 0.6-0.7 ที่จำนวนหาความสัมพันธ์ 3,550,405 หา ซึ่งมากกว่าหาความสัมพันธ์ในช่วงความเชื่อมั่น 0.5-0.6 ที่มีจำนวน 2,450,485 หาความสัมพันธ์ เนื่องมาจาก การวิเคราะห์

ความสัมพันธ์ในข้อมูลนั้นมีความซับซ้อน และมีหลายปัจจัยที่มีผลต่อค่าความสัมพันธ์ เช่น จำนวนข้อมูล ความถี่ของรายการ และค่าความน่าจะเป็น เป็นต้น ในกรณีของงานวิจัยนี้ได้ทำการกำหนดค่าสนับสนุนขั้นต่ำ (Minimum Support threshold) ที่ 0.2 ซึ่งพบว่ามีจำนวนกฎความสัมพันธ์จำนวนมากถึง 9,158,342 กฎความสัมพันธ์ ซึ่งแสดงว่ามีรายการที่เกิดขึ้นบ่อยอยู่ในชุดข้อมูล และมีความสัมพันธ์กันอย่างมีนัยสำคัญ ซึ่งสามารถนำไปวิเคราะห์ได้ในขั้นตอนต่อไป

ตารางที่ 4.1

ตารางแสดงจำนวนกฎความสัมพันธ์ ในแต่ละช่วงของค่าสนับสนุน และค่าความเชื่อมั่น

ค่าสนับสนุน	จำนวนกฎความสัมพันธ์					
	ค่าความเชื่อมั่นเท่ากับ 1	ช่วงค่าความเชื่อมั่นระหว่าง 0.9 ถึงน้อยกว่า 1	ช่วงค่าความเชื่อมั่นระหว่าง 0.8 ถึงน้อยกว่า 0.9	ช่วงค่าความเชื่อมั่นระหว่าง 0.7 ถึงน้อยกว่า 0.8	ช่วงค่าความเชื่อมั่นระหว่าง 0.6 ถึงน้อยกว่า 0.7	ช่วงค่าความเชื่อมั่นระหว่าง 0.5 ถึงน้อยกว่า 0.6
0.2-0.3	570	539,161	919,967	1,435,224	3,404,638	2,307,622
0.3-0.4	7	46,781	76,372	108,231	131,286	131,466
0.4-0.5	0	5,696	8,949	11,549	12,814	10,893
≥ 0.5	0	1,146	1,716	2,083	1,667	504
จำนวนกฎความสัมพันธ์	577	592,784	1,007,004	1,557,087	3,550,405	2,450,485

4.2 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่ง (Strong Association)

ผู้วิจัยได้ทำการคัดเลือกกฎความสัมพันธ์ที่มีแข็งแกร่ง (Strong Association) โดยการคัดเลือกจากค่าสนับสนุน และค่าความเชื่อมั่นที่สูงที่สุด สำหรับแต่ละจำนวนไอเทมเซตด้านซ้ายของกฎความสัมพันธ์ (Left Hand Side: LHS) โดยเริ่มต้นจากจำนวน 2 ไอเทมเซตไปจนถึง 5 ไอเทมเซต ซึ่งจะแสดงในตารางที่ 4.2 ถึง ตารางที่ 4.5 เพื่อนำมาวิเคราะห์หากกฎความสัมพันธ์ที่มีประโยชน์ต่อแพลตฟอร์มการให้สินเชื่อออนไลน์ส่วนบุคคล โดยมีรายละเอียดดังนี้

4.2.1 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 2 ไอเทมเซต

จากการคัดเลือกกฎความสัมพันธ์โดยกำหนดจำนวนไอเทมเซตของกฎความสัมพันธ์ด้านซ้ายจำนวน 2 ไอเทมเซต และเลือกกฎความสัมพันธ์ด้วยค่าสนับสนุนที่มากกว่า 0.5 และค่าความเชื่อมั่นที่สูงที่สุด แสดงได้ในตารางที่ 4.2 เพื่อคัดเลือกกฎความสัมพันธ์ที่มีความแข็งแกร่ง

ตารางที่ 4.2

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 2 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'tot_cur_bal_low'}, {'total_bal_il_low'}	=> {'avg_cur_bal_low'}	0.50	0.99
2	{'inq_fi_low'}, {'total_bal_il_low'}	=> {'inq_last_12m_low'}	0.52	0.93
3	{'total_bc_limit_low'}, {'total_rec_prncp_low'}	=> {'last_pymnt_amnt_low'}	0.52	0.99
4	{'total_bc_limit_low'}, {'annual_inc_low'}	=> {'max_bal_bc_low'}	0.55	0.95
5	{'total_bc_limit_low'}, {'annual_inc_low'}	=> {'revol_bal_low'}	0.58	0.99
6	{'tot_cur_bal_low'}, {'avg_cur_bal_low'}	=> {'total_bal_il_low'}	0.50	0.96
7	{'max_bal_bc_low'}, {'fico_score_low'}	=> {'total_bc_limit_high'}	0.56	0.93

จากตารางที่ 4.2 อธิบายกฎความสัมพันธ์ได้ดังนี้

กฎข้อที่ 1 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มียอดคงค้างวงเงินสินเชื่อทั้งหมดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ แล้ว ผู้ขอสินเชื่อจะมีค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 99

กฎข้อที่ 2 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนการสอบถามการเงินส่วนบุคคลอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ แล้ว ผู้ขอสินเชื่อจะมีจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 52 และค่าความเชื่อมั่นร้อยละ 93

- กฎข้อที่ 3 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดที่ใช้งานอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ
แล้ว จำนวนเงินที่ผู้กู้ชำระครั้งล่าสุดให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 52 และค่าความเชื่อมั่นร้อยละ 99
- กฎข้อที่ 4 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีรายได้ประจำปีอยู่ในเกณฑ์ที่ต่ำ
แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending จะมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 55 และค่าความเชื่อมั่นร้อยละ 95
- กฎข้อที่ 5 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดที่อยู่ในเกณฑ์ที่ต่ำ และมีรายได้ประจำปีอยู่ในเกณฑ์ที่ต่ำ
แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending จะมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 58 และค่าความเชื่อมั่นร้อยละ 99
- กฎข้อที่ 6 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มียอดคงค้างวงเงินสินเชื่อทั้งหมดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ
แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending จะมียอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 96
- กฎข้อที่ 7 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และคะแนนเครดิต FICO อยู่ในเกณฑ์ที่ต่ำ
แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดที่ใช้งานอยู่ในเกณฑ์ที่สูง ด้วยค่าสนับสนุนที่ร้อยละ 56 และค่าความเชื่อมั่นร้อยละ 93

4.2.2 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 3 ไอเทมเซต

จากการคัดเลือกกฎความสัมพันธ์โดยกำหนดจำนวนไอเทมเซตของกฎความสัมพันธ์ด้านซ้ายจำนวน 3 ไอเทมเซต และเลือกกฎความสัมพันธ์ด้วยค่าสนับสนุนที่มากกว่า 0.5 และค่าความเชื่อมั่นที่สูงที่สุด แสดงได้ในตารางที่ 4.3 เพื่อคัดเลือกกฎความสัมพันธ์ที่มีความแข็งแกร่ง

ตารางที่ 4.3

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 3 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'mths_since_recent_bc_low'}, {'revol_bal_low'}, {'inq_fi_low'}	=> {'inq_last_12m_low'}	0.50	0.92
2	{'mths_since_recent_bc_low'}, {'revol_bal_low'}, {'total_rec_prncp_low'}	=> {'last_pymnt_amnt_low'}	0.53	0.99
3	{'mths_since_recent_bc_low'}, {'total_bc_limit_low'}, {'annual_inc_low'}	=> {'max_bal_bc_low'}	0.50	0.96
4	{'total_bc_limit_low'}, {'annual_inc_low'}, {'max_bal_bc_low'}	=> {'revol_bal_low'}	0.55	0.99
5	{'revol_bal_low'}, {'max_bal_bc_low'}, {'total_acc_low'}	=> {'total_bal_il_low'}	0.52	0.93
6	{'revol_bal_low'}, {'max_bal_bc_low'}, {'fico_score_low'}	=> {'total_bc_limit_high'}	0.55	0.93

จากตารางที่ 4.3 อธิบายกฎความสัมพันธ์ได้ดังนี้

- กฎข้อที่ 1 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนการสอบถามการเงินส่วนบุคคลอยู่ในเกณฑ์ที่ต่ำ แล้ว ผู้ขอสินเชื่อจะมีจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 92
- กฎข้อที่ 2 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ แล้ว จำนวนเงินที่ผู้กู้ชำระครั้งล่าสุดให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 53 และค่าความเชื่อมั่นร้อยละ 99

- กฎข้อที่ 3 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีรายได้ประจำปียอยู่ในเกณฑ์ที่ต่ำ
- แล้ว จะมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 96
- กฎข้อที่ 4 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีรายได้ประจำปียอยู่ในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ
- แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending จะมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 55 และค่าความเชื่อมั่นร้อยละ 95
- กฎข้อที่ 5 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และมีจำนวนบัญชีเครดิตทั้งหมดอยู่ในเกณฑ์ที่ต่ำ
- แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending จะมียอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 52 และค่าความเชื่อมั่นร้อยละ 93
- กฎข้อที่ 6 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และคะแนนเครดิต FICO อยู่ในเกณฑ์ที่ต่ำ
- แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดที่ใช้งานอยู่ในเกณฑ์ที่สูง ด้วยค่าสนับสนุนที่ร้อยละ 55 และค่าความเชื่อมั่นร้อยละ 93

4.2.3 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 4 ไอเทมเซต

จากการคัดเลือกกฎความสัมพันธ์โดยกำหนดจำนวนไอเทมเซตของกฎความสัมพันธ์ด้านซ้ายจำนวน 4 ไอเทมเซต และเลือกกฎความสัมพันธ์ด้วยค่าสนับสนุนที่มากกว่า 0.5 และค่าความเชื่อมั่นที่สูงที่สุด แสดงได้ในตารางที่ 4.4 เพื่อคัดเลือกกฎความสัมพันธ์ที่มีความแข็งแกร่ง

ตารางที่ 4.4

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 4 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'mths_since_recent_bc_low'}, {'revol_bal_low'}, {'total_rec_prncp_low'}, {'mo_sin_rcnt_rev_tl_op_low'}	=> {'last_pymnt_amnt_low'}	0.50	0.998
2	{'mths_since_recent_bc_low'}, {'avg_cur_bal_low'}, {'total_bc_limit_low'}, {'mo_sin_rcnt_rev_tl_op_low'}	=> {'max_bal_bc_low'}	0.51	0.961
3	{'avg_cur_bal_low'}, {'total_bc_limit_low'}, {'max_bal_bc_low'}, {'last_pymnt_amnt_low'}	=> {'revol_bal_low'}	0.51	0.994
4	{'mo_sin_rcnt_rev_tl_op_low'}, {'revol_bal_low'}, {'max_bal_bc_low'}, {'fico_score_low'}	=> {'total_bc_limit_high'}	0.50	0.936

จากตารางที่ 4.4 อธิบายกฎความสัมพันธ์ได้ดังนี้

กฎข้อที่ 1 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ

แล้ว ผู้จำนวนเงินที่ผู้กู้ชำระครั้งล่าสุดให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 99

กฎข้อที่ 2 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ

แล้ว จะมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 51 และค่าความเชื่อมั่นร้อยละ 96

กฎข้อที่ 3 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินที่ผู้กู้ชำระครั้งล่าสุดให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ
แล้ว จะมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 51 และค่าความเชื่อมั่นร้อยละ 99

กฎข้อที่ 4 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และคะแนนเครดิต FICO อยู่ในเกณฑ์ที่ต่ำ
แล้ว ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีวงเงินบัตรเครดิตทั้งหมดที่ใช้งานอยู่ในเกณฑ์ที่สูง ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 93

4.2.4 ผลการวิจัยกฎความสัมพันธ์ที่แข็งแกร่งที่มีจำนวน 5 ไอเทมเซต

จากการคัดเลือกกฎความสัมพันธ์โดยกำหนดจำนวนไอเทมเซตของกฎความสัมพันธ์ด้านซ้ายจำนวน 5 ไอเทมเซต และเลือกกฎความสัมพันธ์ด้วยค่าสนับสนุนที่มากกว่า 0.5 และค่าความเชื่อมั่นที่สูงที่สุด แสดงได้ในตารางที่ 4.5 เพื่อคัดเลือกกฎความสัมพันธ์ที่มีความแข็งแกร่ง

ตารางที่ 4.5

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 5 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'mths_since_recent_bc_low'}, {'revol_bal_low'}, {'total_bc_limit_low'}, {'mo_sin_rcnt_rev_tl_op_low'}, {'avg_cur_bal_low'}	=> {'max_bal_bc_low'}	0.50	0.962

ตารางที่ 4.5

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 5 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด (ต่อ)

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
2	{'mths_since_recent_bc_low'}, {'total_bc_limit_low'}, {'max_bal_bc_low'}, {'mo_sin_rcnt_rev_tl_op_low'}, {'avg_cur_bal_low'}	=> {'revol_bal_low'}	0.50	0.993

จากตารางที่ 4.5 อธิบายกฎความสัมพันธ์ได้ดังนี้

กฎข้อที่ 1 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ

แล้ว จะมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 96

กฎข้อที่ 2 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ

แล้ว จะมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 99

4.3 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ

ปัญหาการผิณฑชำระหนี้เป็นปัญหาสำคัญจากการให้สินเชื่อทุกรูปแบบ เพื่อให้กฎความสัมพันธ์สามารถนำไปวิเคราะห์พฤติกรรมการชำระหนี้ได้นั้น ในงานวิจัยฉบับนี้จึงได้ทำการ

คัดเลือกกฎความสัมพันธ์ที่ใช้ในการวิเคราะห์ความเสี่ยงจากการปล่อยสินเชื่อ ด้วยวิธีการกำหนดไอเทมด้านขวาของกฎความสัมพันธ์ (Right Hand Side: RHS) เป็น 2 ไอเทมดังนี้

ไอเทมที่ 1. สถานะปัจจุบันของสินเชื่อ (loan_status) เพื่อค้นหากฎความสัมพันธ์ที่มีผลต่อสถานะปัจจุบันของสินเชื่อ ซึ่งประกอบด้วยกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (Fully Paid) และกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน (Charge Off)

ไอเทมที่ 2. คะแนนเครดิต FICO ซึ่งเป็นตัวแปรที่สะท้อนถึงพฤติกรรมการชำระหนี้ที่ผ่านมาของผู้ขอสินเชื่อกับทุกสถาบันการเงิน ซึ่งประกอบด้วย 2 ไอเทมย่อย

1. ช่วงของค่าเฉลี่ยคะแนนเครดิต FICO สำหรับค่าต่ำสุดและค่าสูงสุดของผู้ขอสินเชื่อ เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club (fico_score)

2. ช่วงของค่าเฉลี่ยคะแนนเครดิต FICO สำหรับค่าต่ำสุดและค่าสูงสุดของผู้ขอสินเชื่อ ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club (last_fico_score)

ทั้งนี้กฎความสัมพันธ์ที่ได้จากการกำหนดไอเทมด้านขวาของกฎความสัมพันธ์ทั้ง 2 ไอเทมที่กล่าวข้างต้นนั้นมีกฎความสัมพันธ์จำนวนมาก จึงทำการคัดเลือกกฎความสัมพันธ์ที่มีความน่าสนใจหรือกฎความสัมพันธ์ที่มีแข็งแกร่ง (Strong Association) โดยการคัดเลือกจากค่าสนับสนุน และค่าความเชื่อมั่นที่สูงที่สุด สำหรับแต่ละจำนวนไอเทมเซตด้านซ้ายของกฎความสัมพันธ์ (Left Hand Side: LHS) โดยเริ่มต้นจากจำนวน 2 ไอเทมเซตไปจนถึง 5 ไอเทมเซต ซึ่งจะแสดงในตารางที่ 4.6 สำหรับกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ และตารางที่ 4.7 สำหรับกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นคะแนนเครดิต FICO

4.3.1 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ กรณีกำหนดไอเทมด้านขวาของกฎความสัมพันธ์เป็นสถานะปัจจุบันของสินเชื่อ

ตารางที่ 4.6

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'inq_fi_low'}, {'last_fico_score_high'}	=> {'loan_status_Fully Paid'}	0.42	0.941
2	{'total_rec_int_low'}, {'last_fico_score_high'}, {'inq_last_12m_low'}	=> {'loan_status_Fully Paid'}	0.31	0.948

ตารางที่ 4.6

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ (ต่อ)

ลำดับที่	กฎความสัมพันธ์		ค่า สนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{'inq_fi_low'}, {'last_fico_score_high'}	=> {'loan_status_Fully Paid'}	0.42	0.941
2	{'total_rec_int_low'}, {'last_fico_score_high'}, {'inq_last_12m_low'}	=> {'loan_status_Fully Paid'}	0.31	0.948
3	{'last_fico_score_high'}, {'revol_bal_low'}, {'total_rec_int_low'}, {'max_bal_bc_low'}	=> {'loan_status_Fully Paid'}	0.31	0.946
4	{'last_fico_score_medium'}, {'total_rec_prncp_low'}	=> {'loan_status_Charge off'}	0.36	0.940
5	{'mths_since_recent_bc_low'}, {'last_fico_score_medium'}, {'total_rec_prncp_low'}	=> {'loan_status_Charge off'}	0.33	0.942
6	{'mths_since_recent_bc_low'}, {'last_fico_score_medium'}, {'total_rec_prncp_low'}, {'last_pymnt_amnt_low'}	=> {'loan_status_Charge off'}	0.33	0.943
7	{'mths_since_recent_bc_low'}, {'total_rec_prncp_low'}, {'last_fico_score_medium'}, {'mo_sin_rcnt_rev_tl_op_low'}, {'last_pymnt_amnt_low'}	=> {'loan_status_Charge off'}	0.31	0.943
8	{'mths_since_recent_bc_low'}, {'revol_bal_low'}, {'max_bal_bc_low'}, {'last_fico_score_medium'}, {'mo_sin_rcnt_rev_tl_op_low'}, {'last_pymnt_amnt_low'}	=> {'loan_status_Charge off'}	0.31	0.912

จากตารางที่ 4.6 อธิบายกฎความสัมพันธ์ได้ดังนี้

- กฎข้อที่ 1 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนการสอบถามสถานะเครดิตส่วนบุคคลอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง
แล้ว สถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ด้วยค่าสนับสนุนที่ร้อยละ 42 และค่าความเชื่อมั่นร้อยละ 94
- กฎข้อที่ 2 : ถ้า จำนวนดอกเบี้ยที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ
แล้ว สถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 94
- กฎข้อที่ 3 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนดอกเบี้ยที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ
แล้ว สถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 94
- กฎข้อที่ 4 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ปานกลาง และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ
แล้ว สถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน ด้วยค่าสนับสนุนที่ร้อยละ 36 และค่าความเชื่อมั่นร้อยละ 94
- กฎข้อที่ 5 : ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่

ในเกณฑ์ปานกลาง และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ

แล้ว สถานะปัจจุบันของผู้ขอสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน ด้วยค่าสนับสนุนที่ร้อยละ 33 และค่าความเชื่อมั่นร้อยละ 94

กฎข้อที่ 6 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งสุดท้ายจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ปานกลาง และจำนวนจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินที่ผู้กู้ชำระครั้งสุดท้ายให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ

แล้ว สถานะปัจจุบันของผู้ขอสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน ด้วยค่าสนับสนุนที่ร้อยละ 33 และค่าความเชื่อมั่นร้อยละ 94

กฎข้อที่ 7 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งสุดท้ายจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ปานกลาง และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งสุดท้ายอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินที่ผู้กู้ชำระครั้งสุดท้ายให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ

แล้ว สถานะปัจจุบันของผู้ขอสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 94

กฎข้อที่ 8 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งสุดท้ายจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และมีค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ปานกลาง และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งสุดท้ายอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินที่ผู้กู้ชำระครั้งสุดท้ายให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำ

แล้ว สถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 91

4.3.2 กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อ กรณีกำหนดไอเทมด้านขวาของกฎความสัมพันธ์เป็นคะแนนเครดิต FICO

ตารางที่ 4.7

ตารางแสดงตัวอย่างของกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นคะแนนเครดิต FICO

ลำดับที่	กฎความสัมพันธ์		ค่าสนับสนุน	ค่าความเชื่อมั่น
	LHS	RHS		
1	{pct_tl_nvr_dlq_high'}, {'loan_status_Fully Paid'}	=> {'last_fico_score_high'}	0.36	0.904
2	{pct_tl_nvr_dlq_high'}, {'inq_last_12m_low'}, {'loan_status_Fully Paid'}	=> {'last_fico_score_high'}	0.31	0.909
3	{max_bal_bc_low'}, {'bc_util_high'}	=> {'fico_score_low'}	0.22	0.910
4	{mths_since_recent_bc_low'}, {'total_bc_limit_low'}, {'bc_util_high'}	=> {'fico_score_low'}	0.21	0.912
5	{mths_since_recent_bc_low'}, {'revol_bal_low'}, {'total_bc_limit_low'}, {'bc_util_high'}	=> {'fico_score_low'}	0.21	0.913

จากตารางที่ 4.7 อธิบายกฎความสัมพันธ์ได้ดังนี้

กฎข้อที่ 1 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีเปอร์เซ็นต์ของบัญชีเครดิตที่ไม่เคยผิดนัดชำระอยู่ในเกณฑ์ที่สูง และสถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข

แล้ว ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง ด้วยค่าสนับสนุนที่ร้อยละ 36 และค่าความเชื่อมั่นร้อยละ 90

กฎข้อที่ 2 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีเปอร์เซ็นต์ของบัญชีเครดิตที่ไม่เคยผิดนัดชำระอยู่ในเกณฑ์ที่สูง และจำนวนการสอบถามสถานะเครดิตในรอบ 12

เดือนอยู่ในเกณฑ์ที่ต่ำ และสถานะปัจจุบันของสินเชื่อกำลังอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข

แล้ว ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 90

กฎข้อที่ 3 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และสัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง

แล้ว ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 22 และค่าความเชื่อมั่นร้อยละ 91

กฎข้อที่ 4 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดที่ใช้ทำงานอยู่ในเกณฑ์ที่ต่ำ และสัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง

แล้ว ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 21 และค่าความเชื่อมั่นร้อยละ 91

กฎข้อที่ 5 : **ถ้า** ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดที่ใช้ทำงานอยู่ในเกณฑ์ที่ต่ำ และสัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง

แล้ว ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 21 และค่าความเชื่อมั่นร้อยละ 91

4.4 การวิเคราะห์กฎความสัมพันธ์ โดยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ

4.4.1 การวิเคราะห์กฎความสัมพันธ์ที่แข็งแกร่ง (Strong Association)

จากการค้นหากฎความสัมพันธ์ของข้อมูลการให้สินเชื่อผ่านแพลตฟอร์ม P2P Lending ของแพลตฟอร์ม Lending Club โดยใช้จำนวนไอเทมเซต ค่าสนับสนุน และค่าความเชื่อมั่นเป็นเกณฑ์นั้น พบว่ามีกฎความสัมพันธ์จำนวนมากที่มีความน่าสนใจ จากตารางที่ 4.3 แสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 3 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด สำหรับกฎข้อที่ 6 พบว่า ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ หรืออยู่ในช่วง $[0, 39,528.28]$ เหรียญดอลลาร์สหรัฐ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ หรืออยู่ในช่วง $[0, 9,555]$ เหรียญดอลลาร์สหรัฐ และคะแนนเครดิต FICO อยู่ในเกณฑ์ที่ต่ำ หรืออยู่ในช่วง $[662, 705.95]$ คะแนน แล้ว วงเงินบัตรเครดิตทั้งหมดที่ใช้งานอยู่ในเกณฑ์ที่สูง หรือ อยู่ในช่วง $[34,140, 1,105,500]$ เหรียญดอลลาร์สหรัฐ กฎนี้จะแสดงถึงความเสี่ยงที่สูงที่จะผิดนัดชำระหนี้ได้ ซึ่งแสดงถึงการก่อหนี้จำนวนมากจากวงเงินบัตรเครดิตทั้งหมดที่อยู่ในเกณฑ์ที่สูง รวมไปถึงคะแนนเครดิตที่อยู่ในเกณฑ์ที่ต่ำซึ่งบ่งบอกถึงการมีประวัติทางการเงินที่ไม่ดีนัก อย่างไรก็ตามสิ่งที่สำคัญต่อการพิจารณาควรมีปัจจัยอื่นร่วมด้วย เช่น รายได้ สถานการณ์จ้างงาน และอัตราส่วนหนี้สินต่อรายได้ เป็นต้น เพื่อทำการประเมินความน่าเชื่อถือของการพิจารณาการผิดนัดชำระหนี้อย่างครอบคลุม นอกจากนี้กฎข้อที่ 1 พบว่าผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนการสอบถามการเงินส่วนบุคคลอยู่ในเกณฑ์ที่ต่ำ แล้ว ผู้ขอสินเชื่อจะมีจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ ค่าความเชื่อมั่นร้อยละ 92 ซึ่งบ่งชี้ว่าเมื่อมีเหตุการณ์ก่อนหน้าแล้ว จะมีโอกาส 92% ที่ผลตามมาจะเกิดขึ้นด้วย กฎนี้น่าสนใจเพราะระบุความเชื่อมโยงที่เป็นไปได้ระหว่างพฤติกรรมสินเชื่อล่าสุดของผู้กู้และแนวโน้มที่จะสอบถามเกี่ยวกับสินเชื่อในรอบ 12 เดือนที่ผ่านมา

จากตารางที่ 4.4 แสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 4 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด ในกฎข้อที่ 3 พบว่าขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และจำนวนเงินที่ผู้กู้ชำระครั้งล่าสุดให้กับผู้ให้กู้จะอยู่ในเกณฑ์ที่ต่ำแล้ว จะมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 51 และค่าความเชื่อมั่นร้อยละ 99 ซึ่งค่าสนับสนุนที่ร้อยละ 51 หมายความว่าเหตุการณ์ก่อนหน้า (LHS) จะเกิดเหตุการณ์ที่ตามมา

(RHS) คิดเป็นสัดส่วน 51% ของธุรกรรมทั้งหมดในชุดข้อมูล ซึ่งเป็นอัตราการเกิดขึ้นที่ค่อนข้างสูง ในส่วนของค่าความเชื่อมั่นสูงที่ 99% แสดงว่าเมื่อมีเหตุการณ์ก่อนหน้า (LHS) มีความเป็นไปได้ 99 ที่ผลที่ตามมา (RHS) จะเกิดขึ้นด้วย กฎที่น่าสนใจเพราะเน้นความสัมพันธ์ที่เป็นไปได้ระหว่างการใช้เครดิตของผู้กู้กับยอดคงเหลือในบัญชีปัจจุบัน และพฤติกรรมการชำระหนี้ แสดงให้เห็นว่าผู้กู้ที่สามารถรักษายอดคงค้างให้ต่ำ และชำระหนี้ได้อย่างต่อเนื่องมีแนวโน้มที่จะมียอดคงเหลือหมุนเวียนที่ต่ำกว่า ซึ่งสามารถมองในเชิงบวกจากผู้ให้กู้ได้

จากตารางที่ 4.5 แสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 5 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด กฎข้อที่ 1 ถ้า ผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีจำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และมีวงเงินบัตรเครดิตทั้งหมดในเกณฑ์ที่ต่ำ และจำนวนเดือนนับตั้งแต่เปิดบัญชีหมุนเวียนครั้งล่าสุดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ แล้ว จะมีค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 96 กฎข้อนี้ที่น่าสนใจเพราะแสดงให้เห็นว่าพฤติกรรมการใช้เครดิตของผู้กู้ เป็นปัจจัยสำคัญที่แสดงถึงสภาพคล่องทางการเงิน จากอัตราการใช้สินเชื่อที่ต่ำ รวมไปถึงระยะเวลาตั้งแต่การเปิดบัญชีบัตรเครดิตครั้งล่าสุดที่อยู่ในเกณฑ์ที่ต่ำนั้น แสดงให้เห็นว่าการขอสินเชื่อเพิ่มมากขึ้น ไม่ได้แสดงพฤติกรรมการใช้สินเชื่อเพิ่มมากขึ้น เพราะยอดคงค้างสูงสุดบัญชีวงเงินหมุนเวียนในแต่ละบัญชียังคงอยู่ในเกณฑ์ที่ต่ำ ซึ่งสะท้อนไปยังสภาพคล่องของผู้ขอสินเชื่อ ซึ่งหากพิจารณาจากข้อมูลส่วนอื่นจะใช้ในการประเมินความเสี่ยงของผู้กู้ได้อย่างมีประสิทธิภาพ

นอกจากนี้กฎความสัมพันธ์บางกฎความสัมพันธ์นั้นเป็นข้อมูลเชิงตัวเลขโดยทั่วไป ยกตัวอย่างเช่น จากตารางที่ 4.2 แสดงตัวอย่างของกฎความสัมพันธ์ที่มีจำนวน 2 ไอเทมเซต และค่าสนับสนุน ≥ 0.5 และถูกเลือกโดยความเชื่อมั่นสูงสุด กฎข้อที่ 1 พบว่า ถ้า ยอดคงค้างวงเงินสินเชื่อทั้งหมดอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ แล้ว ยอดคงค้างวงเงินสินเชื่อหมุนเวียนจะอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 99 ซึ่งโดยปกติแล้วนั้นการที่ยอดคงค้างวงเงินสินเชื่อทั้งหมดอยู่ในเกณฑ์ที่ต่ำแล้วนั้น ย่อมส่งผลให้ยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำด้วยเช่นกัน เช่นเดียวกันกับกฎข้อที่ 6 การขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending ถ้า มียอดคงค้างวงเงินสินเชื่อทั้งหมดอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ แล้ว จะมียอดคงค้างวงเงินกู้ยืมอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 96 นี่เป็นการค้นพบที่สมเหตุสมผลโดยทั่วไป เนื่องจากวงเงินที่ต่ำสะท้อนข้อมูลเชิงตัวเลขของยอดวงเงิน การที่ค่าเฉลี่ยอยู่ในเกณฑ์ที่ต่ำแล้วนั้น ค่า

ของตัวเลขก็จะอยู่ในเกณฑ์ที่ต่ำเช่นเดียวกัน ซึ่งกฎความสัมพันธ์ในลักษณะนี้นั้นไม่ได้มีความน่าสนใจมากนักที่จะนำมาใช้ในการวิเคราะห์ข้อมูลการขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending

4.4.2 การวิเคราะห์กฎความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยง โดยผู้เชี่ยวชาญด้านการวิเคราะห์สินเชื่อ

จากกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อจำนวน 8 กฎความสัมพันธ์ข้างต้น ในตารางที่ 4.6 ประกอบด้วยกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไขจำนวน 3 กฎความสัมพันธ์ (กฎข้อที่ 1-3) และกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน จำนวน 5 กฎความสัมพันธ์ (กฎข้อที่ 4-8) ซึ่งมีกฎความสัมพันธ์ที่น่าสนใจสำหรับการวิเคราะห์ความเสี่ยงทางด้านเครดิตจำนวนมาก ซึ่งจะทำการแบ่งการวิเคราะห์เป็น 2 ส่วน คือ กฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไขจำนวน และ กฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน ดังนี้

4.4.2.1 การวิเคราะห์ความเสี่ยงจากกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข (กฎข้อที่ 1-3 จากตารางที่ 4.6)

กฎข้อที่ 1 ระบุว่าหากผู้ขอสินเชื่อมีจำนวนการสอบถามสถานะเครดิตส่วนบุคคลในรอบ 6 เดือนที่ผ่านมาอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อ อยู่ในเกณฑ์ที่สูง แล้วสถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ด้วยค่าสนับสนุนที่ร้อยละ 42 และค่าความเชื่อมั่นร้อยละ 94 สามารถวิเคราะห์ได้ว่าการที่จำนวนการตรวจสอบสถานะเครดิตอยู่ในเกณฑ์ที่ต่ำนั้น ผู้ขอสินเชื่อมีการติดต่อขอสินเชื่อกับธนาคาร หรือสถาบันการเงินอื่น ๆ ในจำนวนครั้งที่น้อย เพราะในแต่ละครั้งของการขอสินเชื่อใหม่จะมีการตรวจสอบสถานะเครดิต สะท้อนถึงพฤติกรรมของผู้ขอสินเชื่อที่ไม่ได้ขาดสภาพคล่องของเงินทุนจนนำไปสู่การขอสินเชื่อจำนวนมากกับสถาบันการเงินต่าง ๆ ในช่วงระยะเวลา 6 เดือนที่ผ่านมา ในส่วนต่อมาคะแนนเครดิต FICO นั้นถูกคำนวณโดยบริษัท FICO (Fair Isaac Corporation) ซึ่งเป็นผู้ให้บริการที่มีชื่อเสียงในการประเมินความเสี่ยงเครดิตของบุคคล โดยคำนวณมาจาก ประวัติการชำระเงินหนี้, ปริมาณหนี้ที่คงเหลือ, อายุของบัญชีเครดิต, ประวัติการขอสินเชื่อ หรือบัตรเครดิตใหม่ และปัจจัยอื่น ๆ การที่ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่ออยู่ในเกณฑ์ที่สูง สะท้อนถึงประวัติการผ่อนชำระหนี้ที่ผ่านมาของผู้ขอสินเชื่ออยู่ในเกณฑ์ที่ดี ด้วยตัวสองแปรที่กล่าวข้างต้นส่งผลให้ผู้ขอสินเชื่อกลุ่มนี้มีสถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ที่ความเชื่อมั่นสูงถึงร้อยละ 94

กฎข้อที่ 2 ระบุว่าถ้าดอกเบี้ยที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 94 กฎความสัมพันธ์ข้อที่ 2 มีความใกล้เคียงกับกฎความสัมพันธ์ข้อที่ 1 โดยมีตัวแปรที่เพิ่มขึ้นคือจำนวนดอกเบี้ยที่ทางผู้ให้สินเชื่อได้รับหากอยู่ในเกณฑ์ที่ต่ำนั้น อาจจะสะท้อนให้เห็นว่าผู้ขอสินเชื่ออาจจะมีสินเชื่อกับแพลตฟอร์ม Lending Club ได้ไม่นาน ทำให้จำนวนของดอกเบี้ยที่ชำระนั้นอยู่ในเกณฑ์ที่ต่ำ หรือมองได้อีกมุมคือเป็นกลุ่มลูกค้าที่ทางแพลตฟอร์ม Lending Club ประเมินความเสี่ยงจากการให้สินเชื่อว่าเป็นกลุ่มลูกค้าที่มีความเสี่ยงต่ำ ทำให้การกำหนดดอกเบี้ยของลูกค้ารายนี้อยู่ในเกณฑ์ที่ไม่สูง ส่งผลให้บริษัทเก็บอัตราดอกเบี้ยได้น้อยจากลูกค้ารายนี้ ซึ่งด้วยตัวแปรจำนวนดอกเบี้ยที่แพลตฟอร์ม Lending Club ได้รับนั้นอยู่ในเกณฑ์ที่ต่ำ ร่วมกับ ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ ส่งผลให้ผู้ขอสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ดี หรือ ลูกหนี้ที่สามารถชำระหนี้ได้ตามเงื่อนไข ที่ความเชื่อมั่นที่สูงถึงร้อยละ 94

กฎข้อที่ 3 ระบุว่า ถ้าผู้ขอสินเชื่อผ่านแพลตฟอร์ม P2P Lending มีค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และมียอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ และจำนวนดอกเบี้ยที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และค่าสูงสุดของยอดคงค้างบัญชีวงเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ แล้วสถานะปัจจุบันของสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ด้วยค่าสนับสนุนที่ร้อยละ 31 และค่าความเชื่อมั่นร้อยละ 94 ความสัมพันธ์ข้อที่ 3 มีความใกล้เคียงกับกฎความสัมพันธ์ข้อที่ 2 โดยมีตัวแปรที่เพิ่มขึ้นคือตัวแปรยอดคงค้างวงเงินสินเชื่อหมุนเวียน การที่ยอดคงค้างวงเงินสินเชื่ออยู่ในเกณฑ์ที่ต่ำนั้น อาจจะสะท้อนถึงการที่ผู้ขอสินเชื่อมีจำนวนวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่ต่ำ ทำให้ไม่สามารถวงเงินอยู่ในจำนวนมากได้ หรือในทางตรงกันข้ามสามารถมองได้ว่าผู้ขอสินเชื่อรายนี้มีพฤติกรรมการใช้วงเงินสินเชื่อหมุนเวียนที่ต่ำ แสดงถึงผู้ขอสินเชื่อมีสภาพคล่องหรือเงินทุนหมุนเวียนที่เพียงพออยู่แล้ว ส่งผลให้ไม่จำเป็นต้องใช้วงเงินหมุนเวียนมากนัก ซึ่งด้วยตัวแปรยอดใช้วงเงินหมุนเวียนที่อยู่ในเกณฑ์ที่ต่ำ ร่วมกับ ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง และจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนอยู่ในเกณฑ์ที่ต่ำ และจำนวนดอกเบี้ยที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำส่งผลให้ผู้ขอสินเชื่อจะอยู่ในกลุ่มลูกหนี้ที่ดี หรือ ลูกหนี้ที่สามารถชำระหนี้ได้ตามเงื่อนไขที่ความเชื่อมั่นที่สูงถึงร้อยละ 94

จากกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข จำนวน 3 กฎความสัมพันธ์ข้างต้น แสดงให้เห็นว่าจำนวนการสอบถามสถานะเครดิต ซึ่งแสดงการพฤติกรรมการขอสินเชื่อของผู้ขอสินเชื่อ และคะแนนเครดิต FICO ที่อยู่ในเกณฑ์ที่สูงนั้นจะปรากฏอยู่ในทุกกฎความสัมพันธ์ แสดงให้เห็นว่าทั้ง 2 คุณลักษณะ เป็นคุณลักษณะที่สำคัญต่อสถานะการชำระหนี้ของผู้ขอสินเชื่อที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข ที่ความเชื่อมั่นที่สูงถึงร้อยละ 94

4.4.2.2 การวิเคราะห์ความเสี่ยงจากกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 (กฎข้อที่ 4-8 จากตารางที่ 4.6)

กฎความสัมพันธ์ข้อที่ 4-8 ในตารางที่ 4.6 จะเห็นว่า คุณลักษณะที่เป็นตัวกำหนดกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขนั้น จะประกอบด้วย 2 คุณลักษณะสำคัญ ได้แก่ คุณลักษณะค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club ที่อยู่ในเกณฑ์ปานกลาง และคุณลักษณะจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันที่อยู่ในเกณฑ์ที่ต่ำ ซึ่งจะเห็นได้ว่าคุณลักษณะ 2 คุณลักษณะข้างต้นปรากฏในทุกกฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน เมื่อวิเคราะห์คุณลักษณะค่าเฉลี่ยช่วงคะแนนเครดิต FICO ครั้งล่าสุดของผู้ขอสินเชื่อ นั้น จะเห็นได้ว่าเมื่อกำหนดไอเทมด้านขวาของกฎความสัมพันธ์เป็นกลุ่มของลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไข ค่าเฉลี่ยช่วงคะแนนเครดิต FICO ครั้งล่าสุดของผู้ขอสินเชื่อ นั้น จะอยู่ในเกณฑ์ปานกลาง ซึ่งแตกต่างจากกรณีที่กำหนดไอเทมด้านขวาเป็นกลุ่มของลูกหนี้ที่สามารถชำระหนี้ได้ตามเงื่อนไข ซึ่งค่าเฉลี่ยช่วงคะแนนเครดิต FICO ครั้งล่าสุดอยู่ในเกณฑ์ที่สูง ซึ่งถือได้ว่าตัวแปรค่าเฉลี่ยช่วงคะแนนเครดิต FICO เป็นตัวแปรสำคัญในการระบุความสามารถในการชำระหนี้ของผู้ขอสินเชื่อได้ ในส่วนของคุณลักษณะจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันที่อยู่ในเกณฑ์ที่ต่ำนั้น สามารถอธิบายได้ว่าผู้ขอสินเชื่ออาจจะมิสนใจกับแพลตฟอร์ม Lending Club ได้ไม่นาน ทำให้จำนวนของเงินต้นที่ชำระนั้นอยู่ในเกณฑ์ที่ต่ำ หรืออาจจะเป็นผลมาจากผู้ขอสินเชื่ออยู่ในกลุ่มผู้ขอสินเชื่อที่มีความเสี่ยงต่อการผิดนัดชำระหนี้ที่สูงทำให้การกำหนดจำนวนวงเงินที่อนุมัติให้กับผู้ขอสินเชื่อ นั้นอยู่ในเกณฑ์ที่ต่ำ และเป็นผลสืบเนื่องทำให้ผู้ขอสินเชื่อผิดนัดชำระหนี้ได้ในท้ายที่สุด

นอกจากนี้ กฎความสัมพันธ์ข้อที่ 4-8 ยังมีคุณลักษณะอื่นๆ ที่ส่งผลต่อสถานะสินเชื่อ นอกจาก 2 คุณลักษณะที่กล่าวในข้างต้น เมื่อกำหนดไอเทมด้านขวาเป็นสถานะสินเชื่อที่ไม่สามารถชำระหนี้ได้ตามเงื่อนไขได้ จำนวน 4 คุณลักษณะ ได้แก่ จำนวนเดือนนับตั้งแต่เปิดบัญชีบัตรเครดิตครั้งล่าสุดจนถึงปัจจุบันอยู่ในเกณฑ์ที่ต่ำ และยอดคงค้างวงเงินสินเชื่อหมุนเวียนอยู่ในเกณฑ์ที่

ต่ำ และค่าสูงสุดของยอดคงค้างบัญชีเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ และค่าสูงสุดของยอดคงค้างบัญชีเงินหมุนเวียนในแต่ละบัญชีอยู่ในเกณฑ์ที่ต่ำ

จากกฎความสัมพันธ์ที่ไอเทมด้านขวาเป็นลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน จำนวน 5 กฎความสัมพันธ์ข้างต้น แสดงให้เห็นว่าค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club ที่อยู่ในเกณฑ์ปานกลาง และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันที่อยู่ในเกณฑ์ที่ต่ำนั้นจะปรากฏอยู่ในทุกกฎความสัมพันธ์ แสดงให้เห็นว่าทั้ง 2 คุณลักษณะ เป็นคุณลักษณะที่สำคัญต่อสถานะการชำระหนี้ของผู้ขอสินเชื่อที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลานานกว่า 120 วัน ที่ความเชื่อมั่นที่สูงถึงร้อยละ 94

4.4.2.3 การวิเคราะห์คุณลักษณะที่มีผลต่อคะแนนเครดิต FICO เมื่อกำหนดไอเทมด้านขวาเป็นคะแนนเครดิต FICO

ในชุดข้อมูลของ Lending Club พบว่าคุณลักษณะคะแนนเครดิต FICO ประกอบด้วย 2 คุณลักษณะ ได้แก่ คะแนนเครดิต FICO ซึ่งแสดงคะแนนเครดิต FICO ในวันแรกที่ผู้ขอสินเชื่อได้ดำเนินการขอสินเชื่อกับแพลตฟอร์ม Lending Club (*fico_score*) และ คะแนนเครดิต FICO ในครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club (*last_fico_score*) โดยคะแนนเครดิต FICO เป็นปัจจัยสำคัญที่ส่งผลต่อสถานะการขอสินเชื่อ ในงานวิจัยฉบับนี้จึงได้ทำการค้นหากฎความสัมพันธ์ที่มีผลต่อคุณลักษณะคะแนนเครดิต FICO ดังต่อไปนี้

กฎข้อที่ 1-2 ในตารางที่ 4.7 แสดงให้เห็นว่าคุณลักษณะที่เป็นตัวกำหนดให้ค่าเฉลี่ยคะแนนเครดิต FICO ในครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง (*last_fico_score*) นั้นจะประกอบด้วย 2 คุณลักษณะสำคัญที่ปรากฏทั้งในกฎข้อที่ 1 และ 2 ได้แก่ สถานะปัจจุบันของสินเชื่อที่ผู้ขอสินเชื่ออยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข และเปอร์เซ็นต์ของบัญชีเครดิตที่ไม่เคยผิดนัดชำระอยู่ในเกณฑ์ที่สูง จะเห็นได้ว่าทั้ง 2 คุณลักษณะแสดงถึงกลุ่มผู้ขอสินเชื่อที่มีพฤติกรรมการชำระหนี้ที่ผ่านมาอยู่ในเกณฑ์ที่ดี นอกจากนี้คุณลักษณะจำนวนการสอบถามสถานะเครดิตในรอบ 12 เดือนที่อยู่ในเกณฑ์ที่ต่ำนั้น ซึ่งแสดงถึงการที่ผู้ขอสินเชื่อไม่ได้มีความต้องการที่จะก่อหนี้จำนวนมาก ร่วมกับ 2 คุณลักษณะข้างต้น ดังแสดงในกฎที่ 2 ตารางที่ 4.7 จะส่งผลให้คะแนนเครดิต FICO อยู่ในเกณฑ์ที่สูงที่ความเชื่อมั่นสูงถึงร้อยละ 90

กฎข้อที่ 3-5 ในตารางที่ 4.7 แสดงให้เห็นว่าคุณลักษณะที่เป็นตัวกำหนดให้ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่ต่ำ (*fico_score*) มีเพียง 1 คุณลักษณะที่ปรากฏในทุกกฎความสัมพันธ์ ได้แก่ สัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง ซึ่งสะท้อนถึง

พฤติกรรมผู้ขอสินเชื่อที่ขาดสภาพคล่องทางการเงิน ส่งผลทำให้ต้องนำเงินจากบัญชีหมุนเวียน หรือ วงเงินเครดิตมาใช้ในจำนวนมาก และท้ายที่สุดอาจส่งผลให้ไม่สามารถชำระหนี้ได้ตามเงื่อนไข จึงเป็น คุณลักษณะที่สำคัญที่ส่งผลให้คะแนนเครดิต FICO อยู่ในเกณฑ์ที่ต่ำ

จากผลการวิเคราะห์ข้างต้นจะเห็นได้ว่าคุณลักษณะต่างๆที่ปรากฏในกฎความสัมพันธ์เมื่อมีการกำหนดไอเทมด้านขวาเป็นค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ในครั้งแรกเมื่อมีการขอสินเชื่อกับแพลตฟอร์ม Lending Club (*fico_score*) พบว่าคุณลักษณะที่เป็นไอเทม เช็ดด้านซ้ายนั้นค่อนข้างมีความแตกต่าง และไม่ซ้ำกันเหมือนกับกรณีของการกำหนดไอเทมด้านขวา เป็นสถานะสินเชื่อ หรือ ค่าเฉลี่ยคะแนนเครดิต FICO ในครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่าน แพลตฟอร์ม Lending Club (*last_fico_score*) สาเหตุน่าจะมาจากการที่คะแนนเครดิต FICO เมื่อมี การขอสินเชื่อครั้งแรกกับแพลตฟอร์ม Lending Club นั้น อาจเกิดขึ้นมาเป็นระยะเวลาานาน ทำให้ คุณลักษณะนี้ไม่ได้สะท้อนพฤติกรรมหรือลักษณะของผู้ขอสินเชื่อในปัจจุบัน แต่อย่างไรก็ตามจะเห็น ได้ว่าสัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง ยังคงเป็น ตัวกำหนดคะแนนเครดิต FICO ที่ดี

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยฉบับนี้มีนำเสนอเพื่อกรอบพัฒนาวิธีหาความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงของการให้กู้ยืมเงินผ่านแพลตฟอร์ม Peer-to-Peer Lending (P2P Lending) โดยใช้ชุดข้อมูล Lending Club ซึ่งเป็น แพลตฟอร์ม P2P Lending ที่ใหญ่ที่สุดในประเทศสหรัฐอเมริกา โดยใช้ข้อมูลในช่วงพ.ศ. 2550-2563 ซึ่งมีข้อมูลทั้งหมด 2,925,493 ข้อมูล และมีจำนวนคุณลักษณะทั้งหมด 141 คอลัมน์ และอาศัยเทคนิคการเตรียมข้อมูลต่างๆ เพื่อให้ข้อมูลมีคุณภาพก่อนกระบวนการหาความสัมพันธ์ จึงได้เสนอการทำวิศวกรรมคุณลักษณะ โดยใช้แบบจำลอง LightGBM ร่วมกับ Recursive Feature Elimination with 5-Cross-Validation เพื่อกำหนดจำนวนคุณลักษณะที่เหมาะสม และคัดเลือกคุณลักษณะต่างๆ จำนวน 41 คุณลักษณะที่มีผลต่อการวิเคราะห์ความเสี่ยงในการปล่อยสินเชื่อผ่านแพลตฟอร์ม P2P Lending ในส่วนของการแก้ปัญหาข้อมูลไม่สมดุลได้ใช้วิธีการสังเคราะห์ข้อมูลเพิ่ม (Synthetic Minority Oversampling Technique: SMOTE) เพื่อเพิ่มข้อมูลกลุ่มสถานะหนี้ให้มีจำนวนที่เท่ากัน ซึ่งจะได้ข้อมูลรวมทั้งหมด 1,077,834 ข้อมูล และเพื่อให้ข้อมูลอยู่ในรูปแบบที่สามารถนำไปสร้างกฎความสัมพันธ์ด้วยอัลกอริทึมอปริโริได้ จึงเสนอวิธีการแบ่งช่วงข้อมูล ด้วยวิธีการแบ่งกลุ่มสองขั้นตอน (2-Step Clustering) ในขั้นตอนต่อมาการสร้างกฎความสัมพันธ์ได้เสนอใช้อัลกอริทึมอปริโริ และเพื่อคัดกรองกฎความสัมพันธ์เบื้องต้น จึงทำการกำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 2 และค่าลิฟท์มากกว่า 1 พบว่าค้นพบกฎความสัมพันธ์รวมทั้งหมด 9,158,342 กฎความสัมพันธ์

ในส่วนของ การค้นหากฎความสัมพันธ์ที่มีความน่าสนใจหรือกฎความสัมพันธ์ที่มีแข็งแกร่ง (Strong Association) ได้ทำการคัดเลือกจากค่าสนับสนุน และค่าความเชื่อมั่นที่สูงที่สุด สำหรับแต่ละจำนวนไอเทมเซตด้านซ้ายของกฎความสัมพันธ์ (Left Hand Side: LHS) โดยเริ่มต้นจากจำนวน 2 ไอเทมเซตไปจนถึงจำนวนไอเทมเซตสูงสุด ซึ่งกฎความสัมพันธ์ที่ค้นพบหลายกฎความสัมพันธ์เป็นกฎที่มีความน่าสนใจ แสดงความสัมพันธ์ที่สะท้อนให้เห็นถึงพฤติกรรมการใช้เครดิตของผู้ขอสินเชื่อ พฤติกรรมการชำระหนี้ ซึ่งสามารถนำไปวิเคราะห์ลักษณะของผู้ขอสินเชื่อได้ แต่ทั้งนี้ก็มีกฎความสัมพันธ์บางส่วนที่แสดงกฎความสัมพันธ์ของข้อมูลเชิงตัวเลขโดยทั่วไป ซึ่งไม่มีความน่าสนใจในการวิเคราะห์ข้อมูลมากนัก

สำหรับกฎความสัมพันธ์ที่ใช้ในการวิเคราะห์ความเสี่ยง เพื่อค้นหาพฤติกรรมของผู้ขอสินเชื่อที่จะนำไปสู่ปัญหาการผิดนัดชำระหนี้ได้นั้น ในงานวิจัยฉบับนี้ได้กำหนดเงื่อนไขรูปแบบไอเทม

ด้านขวาของกฎความสัมพันธ์ให้เป็นสถานะปัจจุบันของสินเชื่อ และคะแนนเครดิต FICO ซึ่งเกี่ยวข้องกับการชำระหนี้ที่ผ่านมาของผู้ขอสินเชื่อ และให้ผู้เชี่ยวชาญในการวิเคราะห์สินเชื่อทำการวิเคราะห์ซึ่งสามารถสรุปผลลัพธ์ที่ได้จากกฎความสัมพันธ์ได้ดังนี้

1. กฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข พบว่า 2 คุณลักษณะที่สำคัญ ได้แก่ จำนวนการสอบถามสถานะเครดิตในเกณฑ์ที่ต่ำ และคะแนนเครดิต FICO ในเกณฑ์ที่สูง ซึ่งเป็นคุณลักษณะที่ส่งผลให้ผู้ขอสินเชื่อมีแนวโน้มที่จะชำระหนี้ได้ตามเงื่อนไข เนื่องจากคุณลักษณะ 2 คุณลักษณะนี้ อยู่ในทุกกฎความสัมพันธ์เมื่อกำหนดไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ชำระหนี้ได้ตามเงื่อนไข โดยมีค่าสนับสนุนสูงสุดที่ร้อยละ 42 และค่าความเชื่อมั่นสูงถึงร้อยละ 94

2. กฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นสถานะปัจจุบันของสินเชื่อ ในกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไขเป็นระยะเวลาเกินกว่า 120 วัน (Charge Off) พบว่า 2 คุณลักษณะที่สำคัญ ได้แก่ ค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO ครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club ที่อยู่ในเกณฑ์ปานกลาง และจำนวนเงินต้นที่ Lending Club ได้รับตั้งแต่ผู้ขอสินเชื่อเริ่มชำระหนี้จนถึงปัจจุบันที่อยู่ในเกณฑ์ที่ต่ำ เนื่องจากคุณลักษณะ 2 คุณลักษณะนี้ อยู่ในทุกกฎความสัมพันธ์เมื่อกำหนดไอเทมด้านขวาเป็นกลุ่มลูกหนี้ที่ไม่สามารถชำระหนี้คืนได้ตามเงื่อนไข โดยมีค่าสนับสนุนสูงสุดที่ร้อยละ 36 และค่าความเชื่อมั่นสูงถึงร้อยละ 94

3. กฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นค่าเฉลี่ยคะแนนเครดิต FICO ในครั้งล่าสุดที่ผู้ขอสินเชื่อขอสินเชื่อผ่านแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่สูง พบว่า 2 คุณลักษณะที่สำคัญ ได้แก่ สถานะปัจจุบันของสินเชื่อที่ผู้ขอสินเชื่ออยู่ในกลุ่มลูกหนี้ที่สามารถชำระหนี้เต็มจำนวนตามเงื่อนไข และเปอร์เซ็นต์ของบัญชีเครดิตที่ไม่เคยผิดนัดชำระอยู่ในเกณฑ์ที่สูง เนื่องจากคุณลักษณะ 2 คุณลักษณะนี้อยู่ในทุกกฎความสัมพันธ์ โดยมีค่าสนับสนุนสูงสุดที่ร้อยละ 36 และค่าความเชื่อมั่นสูงถึงร้อยละ 90

4. กฎความสัมพันธ์ที่กำหนดไอเทมด้านขวาเป็นค่าเฉลี่ยระหว่างช่วงคะแนนเครดิต FICO เมื่อมีการขอสินเชื่อในครั้งแรกกับแพลตฟอร์ม Lending Club อยู่ในเกณฑ์ที่ต่ำ) พบว่า 1 คุณลักษณะที่สำคัญ ได้แก่ สัดส่วนของยอดค้างปัจจุบันต่อวงเงินของบัตรเครดิตทั้งหมดอยู่ในเกณฑ์ที่สูง เนื่องจากคุณลักษณะนี้อยู่ในทุกกฎความสัมพันธ์ โดยมีค่าสนับสนุนสูงสุดที่ร้อยละ 22 และค่าความเชื่อมั่นสูงถึงร้อยละ 91

ผลลัพธ์ที่ได้จากกฎความสัมพันธ์สำหรับวิเคราะห์ความเสี่ยงด้านเครดิต แสดงให้เห็นว่ากรอบพัฒนาวิธีหาความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงของการให้กู้ยืมเงินผ่านแพลตฟอร์ม P2P Lending ที่นำเสนอขึ้นมีประสิทธิภาพในการระบุรูปแบบพฤติกรรมที่สามารถเชื่อมโยงไปสู่การประเมินความเสี่ยงด้านเครดิตทางการเงินได้ ด้วยการค้นพบรูปแบบและความสัมพันธ์ระหว่าง

พฤติกรรมการชำระหนี้ รวมไปถึงลักษณะเฉพาะของผู้กู้ ซึ่งจะเป็นประโยชน์สำหรับผู้ให้กู้ในการกำหนดนโยบายต่างๆ เพื่อให้ประกอบการตัดสินใจให้กู้ยืมเงินได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

5.2 ข้อเสนอแนะ

งานวิจัยฉบับนี้ได้ใช้ข้อมูลจากแพลตฟอร์มเลนดิงคลับซึ่งเป็นแพลตฟอร์มจากประเทศสหรัฐอเมริกา ในการนำเสนอกรอบพัฒนาวิธีหาความสัมพันธ์สำหรับการวิเคราะห์ความเสี่ยงในการให้สินเชื่อออนไลน์ระหว่างบุคคล (P2P Lending) เนื่องจากในประเทศไทยยังมีข้อมูลสินเชื่อออนไลน์ระหว่างบุคคลไม่มากนัก โดยผลลัพธ์ที่ได้จากการวิเคราะห์ความเสี่ยงจากการให้สินเชื่อได้ที่ได้จากงานวิจัยฉบับนี้นั้น อาจจะไม่สะท้อนพฤติกรรมการชำระหนี้ของผู้ขอสินเชื่อในประเทศอื่นๆที่ไม่ใช่ประเทศของแหล่งที่มาของข้อมูล เนื่องจากพฤติกรรมการขอสินเชื่อ และการชำระหนี้ของบุคคลในแต่ละพื้นที่จะมีความแตกต่างกัน เพราะฉะนั้นเพื่อเป็นการประยุกต์ใช้กรอบพัฒนานี้ให้เกิดประสิทธิภาพมากที่สุดควรนำผลลัพธ์ที่ได้ประยุกต์ใช้ในพื้นที่เดียวกันกับแหล่งข้อมูล

รายการอ้างอิง

- จิตติมา สอาดเอี่ยม. (2563). ปัจจัยที่ส่งผลต่อความตั้งใจใช้บริการแพลตฟอร์มกู้ยืมเงินระหว่างบุคคลกับบุคคลในมุมมองของผู้กู้. ในการค้นคว้าอิสระ (วิทยาศาสตรมหาบัณฑิต). คณะพาณิชยศาสตร์และการบัญชี, มหาวิทยาลัยธรรมศาสตร์.
- พรชนก เทพขาม (2019). เศรษฐกิจแบ่งปัน: บทเรียนการกู้ยืมผ่านช่องทางอิเล็กทรอนิกส์ (Peer-to-peer Lending). ในรายงานธนาคารแห่งประเทศไทย สายนโยบายการเงิน ปี 2019
- พัชรนิกันต์ พงษ์ธนู และวรารัตน์ สงฆ์แป้น (2561). การเลือกคุณลักษณะเพื่อเพิ่มประสิทธิภาพของการจำแนกประเภท: การวิจัยเชิงสำรวจ. คณะวิทยาศาสตร์, สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์. มหาวิทยาลัยขอนแก่น.
- วชิญวิสิฐ เกษรสิทธิ์, และ ดร.วชิต หล่อจิระขุนทด. (2561). การแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลสำหรับการจำแนกผู้ป่วยโรคเบาหวาน. KRU RESEARCH JOURNAL (GRADUATE STUDY), Vol. 18 NO. 3: July-September 2018. วารสารวิจัย มหาวิทยาลัย ขอนแก่น (ฉบับบัณฑิตศึกษา).
- ภัทรารุณี แสงศิริ, ชนกธมน สุขศรี, และ พรนภา เจมส์. (2557). การค้นพบกฎความสัมพันธ์จากระบบสนับสนุนการแก้ปัญหาการใช้งาน. ในรายงานสืบเนื่องจากการประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และ เทคโนโลยีสารสนเทศ ครั้งที่ 10. (หน้า 631-637). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- บุษราภรณ์ มัทธนชัย, ครรชิต มัลย์วงศ์, เสมอแซ สมหอม, และ ญัฐยา ตันตรานนท์ (2558). ความสัมพันธ์ของรายวิชาที่มีผลต่อการพัฒนาศักยภาพนักศึกษาโดยใช้อัลกอริทึมฟรอนทอน. ในการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏกำแพงเพชร ครั้งที่ 3 (ฉบับที่ 1)
- ธนาคารแห่งประเทศไทย (2563). รายงานผลสำรวจการเข้าถึงบริการด้านการเงินภาคครัวเรือนปี 2563. (หน้า 11-13)
- ธรรณัฐ ชูแสง, สายัญญ์ เทพแดง, และ สุวิจักขณ์ โศชนะกุล. (2559). การหาความสัมพันธ์ของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย ระหว่างดัชนีตลาดหลักทรัพย์ในต่างประเทศ ราคาทองคำโลก และราคาน้ำมันดิบโลก ด้วยเทคนิคเหมืองข้อมูล. ในรายงานสืบเนื่องจากการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏภูเก็ต ครั้งที่ 6: 2559 45 ปี มหาวิทยาลัยราชภัฏภูเก็ต สร้างสรรค์การศึกษา พัฒนางานวิจัยภูเก็ต: มหาวิทยาลัยราชภัฏภูเก็ต.

- สุรวัชโยธิน, อัศวิน, & ไพรีเกรง, วรภัทร. (2564). การสร้างตัวแบบการทำนายในการเลือกศึกษาต่อในระดับอุดมศึกษา โดยใช้เทคนิคแบบบูรณาการในการแก้ปัญหาการจำแนกข้อมูลไม่สมดุลของกลุ่มผู้เรียน. *Journal of Information Science and Technology*, 11(1), 65-79. ISSN: 2651-1053 (Online).
- พรชนก เทพขาม (2019). เศรษฐกิจแบ่งปัน: บทเรียนการกู้ยืมผ่านช่องทางอิเล็กทรอนิกส์ (Peer-to-peer Lending). ในรายงานธนาคารแห่งประเทศไทย สายนโยบายการเงิน ปี 2019.
- วชิณวิสิฐ เกษรสิทธิ์, และ ดร.วชิต หล่อจ๊ะระชุมท์กล. (2561). การแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลสำหรับการจำแนกผู้ป่วยโรคเบาหวาน. *KKU RESEARCH JOURNAL (GRADUATE STUDY)*, Vol. 18 NO. 3: July-September 2018. วารสารวิจัย มหาวิทยาลัย ขอนแก่น (ฉบับบัณฑิตศึกษา).
- Acumen Research and Consulting. (2023). P2P Lending Market Size - Global Industry, Share, Analysis, Trends and Forecast 2022 – 2030. Retrieved from <https://www.acumen-researchandconsulting.com/table-of-content/p2p-lending-market>
- Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y. and Haase, A. (2019). “Recursive feature elimination in random forest classification supports nanomaterial grouping.” paper presented at NanolImpact 2019, 15, 100179. ISSN 2452-0748. <https://doi.org/10.1016/j.impa ct.2019.100179>
- Bank of Thailand. (2022). รายชื่อผู้ให้บริการระบบ Peer-to-Peer Lending Platform ที่ทดสอบภายใต้ Regulatory Sandbox. Retrieved from <https://www.bot.or.th/Thai/Payment%20Systems/FinTech/Pages/P2PLendingSandbox.aspx>
- Bao, F., Mao, L., Zhu, Y., Xiao, C., and Xu, C. (2022). “An Improved Evaluation Methodology for Mining Association Rules.” paper presented at the Axioms 2022, 11, 17. <https://doi.org/10.3390/axioms11- 010017>
- Chen, T. and Suzhou, P. R. (2021). “Credit Default Risk Prediction of Lenders with Resampling Methods.” paper presented at the International Conference on Machine Learning, Big Data and Business Intelligence. <https://doi.org/10.1109/MLBDBI54094.2021.00032>
- Cheng, Y.-C., Chang, H.-T., & Lin, C.-Y. (2021). “Predicting Credit Risk in Peer-to-Peer Lending: A Machine Learning Approach with Few Features.” paper presented at

- 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 2021, pp. 295-300, <https://doi.org/10.1109/TAAI54685.2021.00064>
- Chulawate, N. and Kiattisin, S. (2023). "Success Factors Influencing Peer-to-Peer Lending to Support Financial Innovation." paper presented at Sustainability 2023, 15, 4028. <https://doi.org/10.3390/su15054028>
- Desai, M. D. B. and Kaiwade, A. (2018). "Application of Apriori Algorithm for Analyzing Customer Behavior to Improve Deposits in Banks." paper presented at the International Conference on Advances in Computer Technology and Management. (ICACTM) In Association with Novateur Publications IJPET-ISSN No: 2454-7875. ISBN No. 978-81-921768-9-5.
- Hegazy, Mohamed. (2016). "Enhanced Fraud Miner: Credit Card Fraud Detection using Clustering Data Mining Techniques." paper presented at Journal of Computer Science and Technology. 40. 72.
- Havrylchyk, Olena., Mariotto, Carlotta., Rahim, Talal-Ur-. and Verdier, Marianne. (2016) "What drives the expansion of the peer-to-peer lending?." paper presented at SSRN. <https://ssrn.com/abstract=2841316> or <http://dx.doi.org/10.2139/ssrn.2841316>
- Hsueh, S.-C. and Kuo, C.-H. (2017). "Effective Matching for P2P Lending by Mining Strong Association Rules." paper presented at the 3rd International Conference on Industrial and Business Engineering, August 2017, Pages 30–33. <https://doi.org/10.1145/3133811.3133823>
- Jesus Silva, N. Varela, L. A. Borrero López, and R. H. Rojas Millán. (2019). "Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm." paper presented at Procedia Computer Science, 151, 1207–1212.
- K. Chengeta and E. R. Mabika. (2021). "Peer To Peer Social Lending Default Prediction with Convolutional Neural Networks," paper presented at 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data

- Communication Systems (icABCD), Durban, South Africa, 2021, pp. 1-10, doi: 10.1109/icABCD51485.2021.9519309.
- Ko, P.-C., Lin, P.-C., Do, H.-T. and Huang, Y.-F. (2022). "P2P Lending Default Prediction Based on AI and Statistical Models." paper presented at Entropy 2022, 24(6), 801. <https://doi.org/10.3390/e24060801>.
- Lee, K.K.G., Kasim, H., Zhou, W.J., Sirigina, R.P. and Hung, G.G.T. (2023). "Feature redundancy assessment framework for subject matter experts." paper presented at Engineering Applications of Artificial Intelligence, 117(A), 105456. <https://doi.org/10.1016/j.engappai.2022.105456>.
- Misra, P. and Yadav, A. (2020). "Improving the Classification Accuracy using Recursive Feature Elimination with Cross-Validation." paper presented at International Journal on Emerging Technologies, 11(3), 659-665.
- Mohamad, Nadirah., Ahmad, N., Jawawi, Dayang. and Mohd Hashim, Siti. (2020). "Feature Engineering for Predicting MOOC Performance." paper presented at IOP Conference Series: Materials Science and Engineering. 884. 012070. 10.1088/1757-899X/884/1/012070.
- P. Mukherjee and Y. Badr. (2022). "Detection of Defaulters in P2P Lending Platforms using Unsupervised Learning." paper presented at 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), Barcelona, Spain, 2022, pp. 1-5, doi: 10.1109/COINS54846.2022.9854964.
- P. Pramokchon. and P. Punpiti. (2017). "Filter-Based Feature Selection for Data Classification in IoT." paper presented at International Journal of Far Eastern University 2017, vol. 11, no. 3, pp. 98-113.
- S. Datta and K. Mali. (2021). "Significant Association Rule Mining with High Associability." paper presented at 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1159-1164, doi: 10.1109/ICICCS51141.2021.9432237.
- S. Kareem., R. Binti Ahmad. and A. B. Sarlan. (2017). "Framework for the identification of fraudulent health insurance claims using association rule mining." paper presented at 2017 IEEE Conference on Big Data and Analytics (ICBDA), Kuching, Malaysia, 2017, pp. 99-104, doi: 10.1109/ICBDAA.2017.8284114.

- Suchada, C. and Devasia, J. V. (2022). "Peer-to-Peer Lending: Risk Prediction Using Machine Learning on An Imbalanced Dataset." paper presented at the Third International Conference on Intelligent Computing, Instrumentation and Control Technologies. doi: 10.1109/ICICT54557.2022.991 7708
- S. C. Tan. (2018). "Improving Association Rule Mining Using Clustering-based Discretization of Numerical Data." paper presented at 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC), Mon Tresor, Mauritius, 2018, pp. 1-5, doi: 10.1109/ICONIC.2018.8601291.
- Tanantong, T. and Ramjan, S. (2021). "An Association Rule Mining Approach to Discover Demand and Supply Patterns Based on Thai Social Media Data." paper presented at International Journal of Knowledge and Systems Science (IJKSS), 12(2), 1-16.
- X. Mei and Y. Jiang. (2016). "Association rule-based feature selection for credit risk assessment." paper presented at 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS), Chongqing, China, 2016, pp. 301-305, doi: 10.1109/ICOACS.2016.7563102.
- Y. -C. Cheng., H. -T. Chang., C. -Y. Lin. and H. -Y. Chang. (2021). "Predicting Credit Risk in Peer-to-Peer Lending: A Machine Learning Approach with Few Features." paper presented at 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 2021, pp. 295-300, doi: 10.1109/TAAI54685.2021.00064.
- Yan, K. and Zhang, D. (2015). "Feature selection and analysis on correlated gas sensor data with recursive feature elimination." paper presented at Sensors and Actuators B: Chemical, 212, 353-363. ISSN 0925-4005. <https://doi.org/10.1016/j.snb.2015.02.025>.
- Yash. (2020). "Lending Club 2007-2020Q1." Retrieved from Kaggle, <https://www.kaggle.com/ethon0426/lending-club-20072020q1>
- Zhang, Y. (2021). "Application of Data Mining Technology in Financial Risk Management." paper presented at the IEEE Conference on Telecommunications, Optics and Computer Science. doi:10.1109/tocs53301 .2021.9689039

- Zheng, A. and Casari, A. (2018). "Feature Engineering for Machine Learning: Principles and techniques for Data Scientists." O'Reilly Media, Inc., Sebastopol.
- Zhang, W., Seeja, K. R. and Zareapoor, M. (2014). "FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining." paper presented at The Scientific World Journal, 2014, 252797. ISSN 2356-6140. <https://doi.org/10.1155/2014/252797>.
- Z. Wang, Z., Yu, Z., Liu, Y., Liu and R. Han. (2019). "Online P2P Lending User Profile Model Based on Multidimensional Data Analysis." paper presented at The 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2019, pp. 56-59, doi: 10.1109/ICAIBD.2019.8836983.



ประวัติผู้เขียน

ชื่อ	ภาคิน เลิศวิพุธ
วุฒิการศึกษา	ปีการศึกษา 2559: ปริญญาตรี สาขาเศรษฐศาสตร์ เศรษฐศาสตร์บัณฑิต มหาวิทยาลัยเกษตรศาสตร์
ตำแหน่ง	Credit Risk Model Validation บริษัท เงินดีดล้อ จำกัด (มหาชน)

ผลงานทางวิชาการ

Loetwiphut. P., & Tanantong, T. (2023). “Association Rule Mining Framework for Financial Credit Risk Analysis based on Peer-to-Peer Lending Data” paper presented at The 4th Asia Joint Conference on Computing (AJCC)