

การศึกษาและพัฒนาต้นแบบ (Prototype) ของระบบการทำนายสถานะการผัดนัดชำระหนี้ของผู้กู้ใน
P2P Lending ด้วยวิธีการทำเหมืองข้อมูล

สุพิตา วิพัฒน์กิจไพศาล

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
บัณฑิตวิทยาลัย สาขาวิชาเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีไทย-ญี่ปุ่น
ปีการศึกษา 2562

THE STUDY AND PROTOTYPE DEVELOPMENT OF PEER-TO-PEER LENDING PLATFORM
USING DATA MINING TECHNIQUE TO PREDICT LOAN STATUS
FOR EACH LOAN RISK CLASS

Supita Wipatkitpaisal



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Information Technology
Graduate School
Thai-Nichi Institute of Technology
Academic Year 2019

หัวข้อวิทยานิพนธ์

การศึกษาและพัฒนาต้นแบบ (Prototype) ของระบบ
การทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P
Lending ด้วยวิธีการทำเหมืองข้อมูล

โดย

สุพิตา วิพัฒน์กิจไพศาล

สาขาวิชา

เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. พิชิต สุขเจริญพงษ์)

วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. วรภัทร ไพรีเกรง)

.....กรรมการ

(ดร. ณัฐกิตติ์ จิตรเอื้อตระกูล)

.....กรรมการ

(ดร. ชฎาพร เกตุมณี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม)

สุพิทา วิพัฒน์กิจไพศาล: การศึกษาและพัฒนาต้นแบบ (Prototype) ของระบบการทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ด้วยวิธีการทำเหมืองข้อมูล. อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เฉิดโฉม, 91 หน้า.

งานวิจัยนี้เป็นการศึกษาและพัฒนาต้นแบบ (Prototype) ของระบบการทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ด้วยวิธีการทำเหมืองข้อมูล ในแต่ละกลุ่มความเสี่ยง คือ กลุ่มความเสี่ยงต่ำและสูง โดยมีวัตถุประสงค์ดังนี้ 1) เพื่อศึกษาตัวแปรที่ใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ 2) เพื่อสร้างแบบจำลองพยากรณ์แนวโน้มการผิดนัดชำระหนี้ จากเทคนิค classification ต่างๆ ได้แก่ Logistic Regression, Decision Tree (ID3), Decision Tree (C4.5) และ Naïve Bayes รวมถึงคัดเลือกตัวแปร ได้แก่ Pearson's Correlation, Recursive Feature Elimination (RFE) และ Factor Analysis 3) เพื่อพัฒนาต้นแบบ (Prototype) ของระบบการทำนายสถานะการผิดนัดชำระหนี้ ผลการวิจัยพบว่าปัจจัยที่ส่งผล ได้แก่ 1) ด้านผู้กู้และสินเชื่อ 2) ด้านเศรษฐกิจมหภาค โดยแบบจำลองที่มีประสิทธิภาพสำหรับผู้กู้ทั้ง 2 กลุ่มคัดเลือกตัวแปรด้วย Factor Analysis และใช้ synthetic minority oversampling technique (SMOTE) ผ่านเทคนิค Decision Tree (C4.5) ซึ่งมีค่า Accuracy Score และ F-1 Score ที่ 0.8843 และ 0.9383 ตามลำดับ สำหรับกลุ่มความเสี่ยงต่ำ และค่า Accuracy Score และ F-1 Score อยู่ที่ 0.7778 และ 0.8621 ตามลำดับ สำหรับกลุ่มความเสี่ยงสูง อย่างไรก็ตาม แบบจำลอง 2 กลุ่มมีตัวแปรที่ถูกคัดเลือกต่างกัน อีกทั้งยังได้นำข้อมูลอดีตจากผู้ประกอบการไทยมาทดสอบ พบว่าค่า Accuracy Score อยู่ในเกณฑ์ยอมรับได้ที่ 0.8541 และ 0.6956 ตามลำดับ แสดงว่าแบบจำลองสามารถใช้เป็นแนวทางให้ผู้ประกอบการไทยได้

บัณฑิตวิทยา

สาขาวิชา เทคโนโลยีสารสนเทศ

ปีการศึกษา 2562

ลายมือชื่อนักศึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา.....

SUPITA WIPATKITPAISAL: THE STUDY AND PROTOTYPE DEVELOPMENT OF PEER-TO-PEER LENDING PLATFORM USING DATA MINING TECHNIQUE TO PREDICT LOAN STATUS FOR EACH LOAN RISK CLASS. ADVISOR : ASST. PROF. DR. PRAJAK CHERTCHOM, 91 PP.

This research is the study and prototype development of Peer-to-Peer (P2P) Lending platform using data mining technique to predict loan status for each loan risk class which are low and high loan risk classes. The main objective is to 1) study the key attributes for borrowers' loan status prediction 2) create data mining model for borrowers' loan status prediction in P2P Lending platform for each borrowers' loan risk classes through classification techniques including Logistic Regression, Decision Tree (ID3), Decision Tree (C4.5) and Naïve Bayes, as well as different feature selection technique comparison including Pearson's Correlation, Recursive Feature Elimination (RFE) and Factor Analysis 3) develop prototype of loan status prediction system. From this study, it is found that the main factors for loan status prediction in P2P Lending are: 1) borrower and loan factors 2) macroeconomic factors. The study has found that the effective model for both loan risk classes are predictive model selected features through Factor Analysis, through Synthetic Minority Oversampling Technique (SMOTE) and Decision Tree (C4.5) technique. The Accuracy Score and F-1 Score of predictive model for low risk borrowers are 0.8843 and 0.9383 respectively, while for high risk borrowers are 0.7778 and 0.8621 respectively. However, both models have different selected features. After test with historical data of one Thai P2P Lending company, the Accuracy Scores are still acceptable for both models at 0.8541 and 0.6956 respectively. This is indicated that the predictive model can be guideline for Thai company.

Graduate School

Field of Study Information Technology

Academic Year 2019

Student's Signature.....

Advisor's Signature.....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์นี้จะสำเร็จลุล่วงไปได้ด้วยดีนั้น ผู้วิจัยขอกราบขอบพระคุณในความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร. ประจักษ์ เติมโณม ซึ่งได้สละเวลาในการให้คำแนะนำ คำปรึกษา การสนับสนุน รวมถึงตรวจสอบข้อบกพร่องทุกขั้นตอนและกำลังใจเต็มเปี่ยมตลอดระยะเวลาในการทำวิทยานิพนธ์ฉบับนี้

ทั้งนี้ ผู้วิจัยขอขอบพระคุณบัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่นที่ให้ทุนสนับสนุน ผู้วิจัยเป็นทุนการศึกษาเต็มจำนวนในระดับปริญญาโท สาขาวิชาเทคโนโลยีสารสนเทศ รวมถึงให้การสนับสนุนในการทำวิทยานิพนธ์เล่มนี้

นอกจากนี้ ผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบทุกท่าน ได้แก่ ผู้ช่วยศาสตราจารย์ ดร. วรภัทร ไพรีเกรง ดร. ณัฐกิตติ จิตรเอื้อตระกูล และ ดร. ชญาพร เกตุมณี ที่ได้กรุณาให้ข้อเสนอแนะที่เป็นประโยชน์และช่วยตรวจสอบข้อบกพร่องของวิทยานิพนธ์ฉบับนี้ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จและมีความสมบูรณ์

ผู้วิจัยหวังเป็นอย่างยิ่งว่า วิทยานิพนธ์ฉบับนี้จะเกิดประโยชน์ต่อผู้ที่มีความสนใจ สามารถนำไปต่อยอด องค์ความรู้เพื่อประยุกต์ใช้งานการสร้างแบบจำลองเหมืองข้อมูลเพื่อการพยากรณ์ แนวโน้มการผิมนัดชำระหนี้ของผู้กู้ ในแพลตฟอร์ม Peer-to-Peer Lending เพื่อให้เกิดคุณค่าและสามารถใช้งานได้จริง

สุพิตา วิพัฒน์กิจไพศาล

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	4
1.3 สมมติฐานและเงื่อนไขของการวิจัย.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.5 ขอบเขตของการวิจัย.....	5
1.6 นิยามศัพท์เฉพาะ.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
2.1 เอกสารและงานวิจัยที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending.....	7
2.2 เอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยที่ส่งผลต่อแนวโน้มการ ผิดนัดชำระหนี้ของผู้กู้.....	8
2.3 เอกสารและงานวิจัยที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูล.....	11
2.4 เอกสารและงานวิจัยที่เกี่ยวข้องกับความสอดคล้องกันระหว่างข้อมูล ที่ใช้กับข้อมูลที่เกิดขึ้นจริงในประเทศไทย.....	20
2.5 เอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้าง predictive model เฉพาะเจาะจงสำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้.....	22

สารบัญ (ต่อ)

บทที่		หน้า
4	4.3 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีต จากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบสำหรับกลุ่มผู้ที่มีความเสี่ยง ต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B).....	50
	4.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง จากวิธีคัดเลือกตัวแปรและ เทคนิคเหมืองข้อมูลแบบต่างๆ สำหรับกลุ่มผู้ที่มีความเสี่ยงสูง (กลุ่มเกรด สินเชื่อ C-G).....	51
	4.5 การประเมินค่าความสำคัญของตัวแปรที่ใช้ ในแบบจำลองที่มีประสิทธิภาพ มากที่สุดในการทำงานสถานะการผิวนัดชำระหนี้ สำหรับกลุ่มผู้ ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G).....	52
	4.6 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีต จากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ สำหรับกลุ่มผู้ ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G).....	53
	4.7 การออกแบบ User Experience (UX) และ User Interface (UI).....	53
5	บทสรุป ข้อเสนอแนะ บทวิจารณ์.....	79
	5.1 สรุปและอภิปรายผลการวิจัย.....	79
	5.2 ปัญหาและอุปสรรคในการทำงานวิจัย.....	81
	5.3 ข้อเสนอแนะงานวิจัย.....	81
	บรรณานุกรม.....	83
	ภาคผนวก.....	88
	ภาคผนวก ก. การตีพิมพ์ผลงานวิจัย	
	ประวัติย่อผู้วิจัย.....	91

สารบัญตาราง

ตาราง	หน้า
2.1 การเปรียบเทียบการควบคุมดูแลแพลตฟอร์ม P2P Lending ในประเทศ สหรัฐอเมริกาและไทย.....	20
2.2 การเปรียบเทียบองค์ประกอบด้านการใช้งาน สิ่งที่น่าสนใจและช่องว่าง ที่สามารถพัฒนาเพิ่มเติมจากการศึกษาแพลตฟอร์ม P2P Lending ที่มี ชื่อเสียงระดับโลก.....	25
2.3 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านข้อมูลประวัติทางการเงิน.....	27
2.4 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านข้อมูลส่วนตัว.....	29
2.5 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านเศรษฐกิจเชิงมหภาพ.....	30
2.6 สรุปเอกสารที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูลและการเตรียมข้อมูล.....	32
3.1 สรุปตัวแปรต้นที่เกี่ยวข้องที่ใช้ในงานวิจัย.....	36
3.2 สรุปตัวแปรตามที่เกี่ยวข้องที่ใช้ในงานวิจัย.....	40
3.3 คำถามที่พบบ่อยจากแพลตฟอร์ม P2P Lending.....	45
4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้กู้ความเสี่ยงต่ำ ถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) ด้วยเกณฑ์ Accuracy Score.....	48
4.2 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้กู้ความเสี่ยงต่ำ ถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) ด้วยเกณฑ์ F-1 Score.....	48
4.3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) ด้วยเกณฑ์ Accuracy Score.....	51
4.4 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) ด้วยเกณฑ์ F-1 Score.....	51
4.5 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Register.....	61
4.6 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที่ 1.....	67
4.7 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที่ 2.....	70

สารบัญรูป

รูป	หน้า
1.1 จำนวนผู้ประกอบการ P2P Lending ที่เกิดปัญหาสภาพคล่องหรือปิดกิจการ (ราย).....	2
2.1 เทคนิคการสังเคราะห์ข้อมูลใหม่.....	12
2.2 เทคนิคการกำจัดข้อมูลผิดปกติ.....	13
2.3 เทคนิคถดถอยโลจิสติกส์ทวี.....	17
2.4 เทคนิคต้นไม้ตัดสินใจ.....	18
2.5 แผนภาพแสดงสัดส่วนของสินเชื่อในแพลตฟอร์ม P2P Lending ในแต่ละประเทศ.....	21
2.6 แผนภาพแสดงสัดส่วนของสินเชื่อในประเทศไทย.....	22
3.1 กรอบแนวคิดวิจัยสำหรับแบบจำลองพยากรณ์แนวโน้มการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending.....	35
3.2 แผนผังของการสร้างแบบจำลองพยากรณ์จากตัวแปรที่ใช้ในงานวิจัย.....	42
4.1 กราฟแสดงค่าความสำคัญของตัวแปรที่คัดเลือกสำหรับแบบจำลอง.....	50
4.2 กราฟแสดงค่าความสำคัญของตัวแปรที่คัดเลือกสำหรับแบบจำลอง.....	52
4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (a).....	54
4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (b).....	55
4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (c).....	56
4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (d).....	57
4.4 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Home.....	58
4.5 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Login.....	59
4.6 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Register.....	60

สารบัญรูป (ต่อ)

รูป	หน้า
4.7 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Forgot Password.....	64
4.8 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า User Profile.....	65
4.9 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้าที่ 1.....	66
4.10 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้าที่ 2.....	69
4.11 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้าที่ 3.....	72
4.12 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Check Loan Status.....	73
4.13 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า About P2P Lending.....	75
4.14 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า FAQ.....	76
4.15 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Contact Us.....	77
4.16 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัด ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Logout.....	78

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การกู้เงินแบบบุคคลต่อบุคคล (Peer-to-Peer Lending หรือ P2P Lending) ได้ถูกนิยามว่าเป็นการปล่อยกู้ระหว่างผู้ให้กู้หรือนักลงทุน กับผู้กู้ผ่านแพลตฟอร์มทางออนไลน์ โดยไม่มีสถาบันทางการเงินเป็นตัวกลาง ซึ่งจะช่วยตอบสนองความต้องการกู้ยืมของผู้รายย่อย รวมถึงธุรกิจขนาดกลางและขนาดเล็กที่มักได้รับผลกระทบจากความเข้มงวดในการปล่อยกู้ที่มากขึ้นของสถาบันทางการเงิน [1]

นอกจากนี้ P2P Lending เป็นการเพิ่มช่องทางการปล่อยกู้ผ่านทางออนไลน์ ซึ่งช่วยลดค่าใช้จ่ายด้านธุรกรรมสำหรับผู้กู้ และเพิ่มผลตอบแทนการลงทุนสำหรับนักลงทุน เนื่องจากเป็นช่องทางที่จะให้อัตราดอกเบี้ยตอบแทนที่สูงกว่าการลงทุนช่องทางอื่น เช่น เงินฝากประจำ หุ้นกู้ หรือผลิตภัณฑ์ทางการเงินอื่นๆ ของธนาคาร จึงกล่าวได้ว่า P2P Lending เป็นการปล่อยกู้ที่ให้ผลประโยชน์ทั้งกับผู้กู้และนักลงทุน อีกทั้งยังสร้างโอกาสในการเข้าถึงแหล่งเงินกู้ที่ง่ายขึ้นสำหรับธุรกิจขนาดกลางและขนาดเล็ก [2] โดยนักลงทุนสามารถกระจายการลงทุนได้ในผู้กู้หลายราย ขณะที่ผู้กู้ก็สามารถได้รับแหล่งเงินทุนจากนักลงทุนหลายรายอีกด้วย [3]

ทั้งนี้ P2P Lending ประกอบด้วยผู้ที่เกี่ยวข้องดังต่อไปนี้

ผู้กู้ คือ ผู้ที่ทำการขออนุมัติเงินกู้ โดยระบุจำนวนและระยะเวลาที่ต้องการกู้ รวมถึงให้รายละเอียดเกี่ยวกับประวัติทางการเงิน และข้อมูลส่วนตัวแก่ผู้ประกอบการ P2P Lending

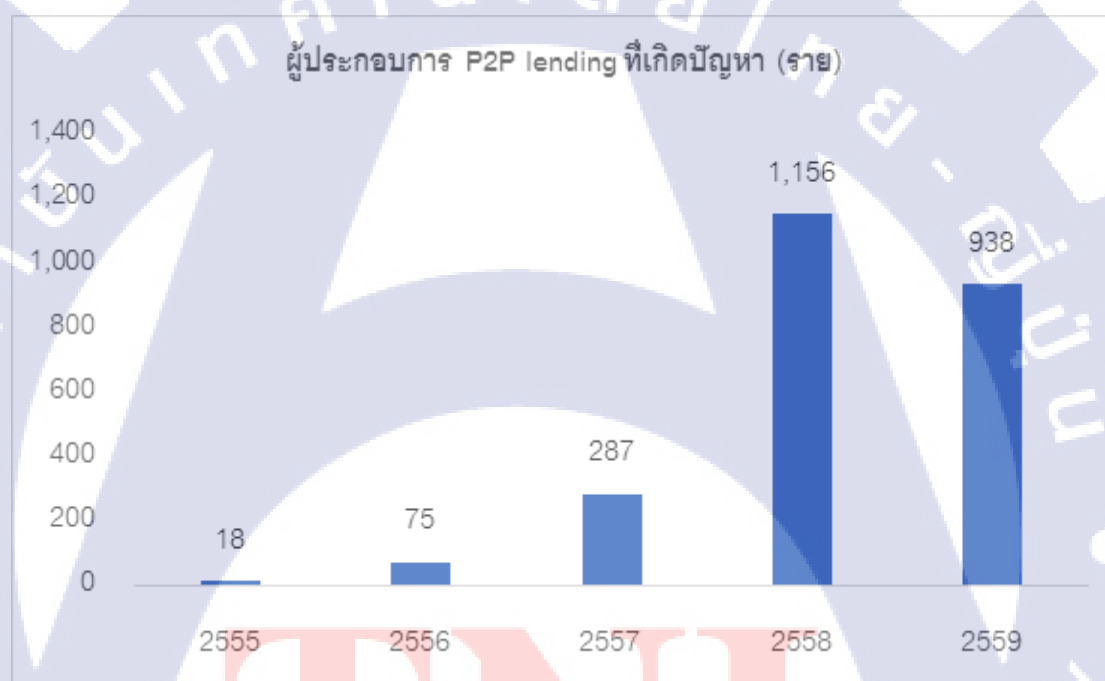
ผู้ประกอบการ P2P Lending คือ ตัวกลางที่เชื่อมระหว่างผู้กู้และนักลงทุน ซึ่งจะทำหน้าที่ประเมินระดับความเสี่ยงและอัตราดอกเบี้ยของผู้กู้แต่ละราย ซึ่งผู้ประกอบการบางรายใช้ระดับคะแนนความเสี่ยง และอัตราขาดทุนที่คาดหวัง เพื่อใช้แบ่งประเภทของคำขออนุมัติเงินกู้ เพื่ออำนวยความสะดวกแก่นักลงทุน

นักลงทุน คือ ผู้ปล่อยกู้ ซึ่งจะได้รับข้อมูลจากผู้ประกอบการ P2P Lending เพื่อใช้ในการตัดสินใจว่าจะลงทุนในผู้กู้รายใดและจำนวนเท่าใด [4]

งานวิจัยของ Transparency Market Research แสดงถึงสถิติของมูลค่าตลาดทั่วโลกของ P2P Lending ที่ระดับ 2.62 หมื่นล้านดอลลาร์สหรัฐในปี 2558 และคาดการณ์ว่ามูลค่าตลาดจะเพิ่มขึ้นอยู่ที่ระดับ 8.98 แสนล้านดอลลาร์สหรัฐภายในปี 2567 ด้วยอัตราเติบโตเฉลี่ยต่อปีที่ 48.2% นอกจากนี้รายงานที่จัดทำโดยมหาวิทยาลัยเคมบริดจ์และ KPMG ยังระบุว่าในปี 2558 P2P Lending ผู้รายย่อยมีมูลค่าตลาดอยู่ที่ 367 ล้านดอลลาร์สหรัฐ ซึ่งมีขนาดที่ใหญ่ที่สุดในบรรดาการกู้

ทางเลือกช่องทางอื่นๆ ขณะที่สำหรับผู้ประกอบการขนาดกลางและขนาดเล็กมีมูลค่าตลาดอยู่ที่ 213 ล้านดอลลาร์สหรัฐ ซึ่งมีขนาดใหญ่เป็นอันดับสองในบรรดาการกู้ยืมทางเลือกช่องทางอื่นๆ โดยประเทศที่มีมูลค่าตลาดของ P2P Lending ใหญ่ที่สุดคือประเทศจีน ซึ่งมีจำนวนผู้ประกอบการกว่า 4,000 ราย เมื่อเทียบกับ 50 รายในช่วงปลายปี 2554 จึงจัดได้ว่า P2P Lending เป็นช่องทางการลงทุนที่เติบโตเร็วที่สุดในบรรดาการลงทุนในผลิตภัณฑ์ทางการเงินต่างๆ [5]

อย่างไรก็ตามสถิติจาก Online Lending House ที่ระบุถึงจำนวนที่เพิ่มขึ้นอย่างมีนัยสำคัญตั้งแต่ปี 2556 ของผู้ประกอบการ P2P Lending ที่เกิดปัญหาต่างๆ เช่น ปัญหาการฉ้อโกง ขำระหนี้ของผู้กู้ ปัญหาสภาพคล่องของเงินทุน หรือแม้กระทั่งปัญหาการปิดกิจการของผู้ประกอบการ เป็นต้น ดังกราฟที่แสดงต่อไปนี้ [6]



รูปที่ 1.1 จำนวนผู้ประกอบการ P2P Lending ที่เกิดปัญหาสภาพคล่องหรือปิดกิจการ (ราย)

จากสถิติดังกล่าว สะท้อนให้เห็นถึงปัญหาที่จะนำไปสู่การขาดทุนของนักลงทุน ซึ่งส่งผลกระทบต่อพัฒนาและการเติบโตของอุตสาหกรรม P2P Lending ซึ่งผู้ประกอบการที่เกิดปัญหาส่วนใหญ่มีสาเหตุมาจากปัจจัยหลักดังต่อไปนี้

1.1.1 ความไม่เท่าเทียมของข้อมูลระหว่างผู้กู้กับนักลงทุน (Information Asymmetry)

ความกังวลส่วนใหญ่ของผู้ลงทุนใน P2P Lending เกิดจากความไม่เท่าเทียมของข้อมูลระหว่างผู้กู้กับนักลงทุน ซึ่งมาจากข้อเท็จจริงที่ว่าผู้กู้มีข้อมูลมากกว่านักลงทุนอย่างน้อย 2 เงื่อนไขที่

จะส่งผลถึงการตัดสินใจปล่อยกู้ของนักลงทุน และความเสี่ยงในการกลับมาลงทุนอีกครั้ง เงื่อนไขแรก คือ ความสามารถและความต้องการที่แท้จริงในการชำระหนี้คืนของผู้กู้ และอีกเงื่อนไข คือ แนวโน้มของการชำระหนี้ล่วงหน้า โดยผู้ประกอบการ P2P Lending จำเป็นต้องมีการรวบรวมข้อมูลที่มีผลต่อการเข้าใจพฤติกรรมการณ์ผิดนัดชำระหนี้ของผู้กู้ และนำข้อมูลดังกล่าวมาแสดงให้เห็นนักลงทุนรับทราบ เพื่อลดความเสี่ยงต่อการเกิดปัญหาดังกล่าว [7]

1.1.2 การขาดผู้เชี่ยวชาญทางการเงิน

จากเดิมที่สถาบันทางการเงินเป็นตัวกลางในการปล่อยเงินกู้ และมีบทบาทสำคัญในการลดความไม่เท่าเทียมกันของข้อมูล เนื่องจากสถาบันทางการเงินดังกล่าวมีผู้เชี่ยวชาญทางการเงิน ที่มีประสบการณ์ในการประเมินความน่าเชื่อถือของผู้กู้แต่ละราย ขณะที่ P2P Lending เป็นการปล่อยกู้ผ่านแพลตฟอร์มออนไลน์ที่จะช่วยลดค่าใช้จ่ายทางธุรกรรม โดยการกำจัดตัวกลางที่เป็นสถาบันทางการเงิน ซึ่งจะทำให้ขาดผู้เชี่ยวชาญทางการเงินในขณะเดียวกัน และอาจนำไปสู่ความไม่เท่าเทียมของข้อมูล ดังนั้นผู้ประกอบการ P2P Lending จำเป็นต้องมีการเปิดเผยและวิเคราะห์ข้อมูลของผู้กู้ด้วยความละเอียดรอบคอบ

1.1.3 การขาดระบบการวัดระดับเครดิตทางการเงินของผู้กู้ที่น่าเชื่อถือ

ผู้ประกอบการบางรายในประเทศจีนได้ใช้มาตรฐานการวัดระดับเครดิตทางการเงินของตนเองในการประเมินความน่าเชื่อถือของผู้กู้ ผ่านข้อมูลที่ผู้กู้ส่งให้ด้วยตนเอง ซึ่งทำให้ขาดความน่าเชื่อถือในการบันทึกระดับเครดิตของผู้กู้ และง่ายต่อการผิดพลาดของข้อมูลที่ได้รับ ผ่านการอนุมัติเงินกู้ทางออนไลน์ ส่งผลให้นักลงทุนอาจไม่เชื่อถือการประเมินระดับเครดิตด้วยวิธีดังกล่าว ดังนั้นผู้ประกอบการ P2P Lending จำเป็นต้องมีกลไกที่ช่วยในการยืนยันความถูกต้องของข้อมูลที่ได้รับ และการประเมินระดับเครดิตของผู้กู้ที่น่าเชื่อถือมากขึ้น ตัวอย่างเช่น ผู้ประกอบการในจีนมีการใช้ช่องทางออฟไลน์มาช่วยใน P2P Lending ผ่านการเข้าพบผู้กู้โดยตรง หรือผ่านการอ้างอิงจากบุคคลที่สาม เป็นต้น [1]

1.1.4 การขาดการควบคุมตามกฎหมายที่มีประสิทธิภาพ

ตัวอย่างของผู้ประกอบการ P2P Lending ในประเทศจีนที่ขาดหน่วยงานที่ควบคุมดูแล P2P Lending โดยตรง ทำให้กฎหมายไม่มีความรัดกุมอย่างเพียงพอ ด้วยเทคนิคสำหรับการจัดการความเสี่ยงที่ล่าสมัย ส่งผลให้เกิดความเสี่ยงเชิงระบบกับผู้ประกอบการที่มีความเกี่ยวเนื่องกันเป็นวงกว้าง รวมไปถึงการลุกลามของความเสี่ยงด้านสินเชื่อ ด้านปฏิบัติการ ด้านสภาพคล่อง และด้านจริยธรรม

ดังนั้นผู้ประกอบการ P2P Lending จำเป็นต้องมีมาตรการในการป้องกันและควบคุมความเสี่ยงเชิงระบบที่อาจเกิดขึ้น [8]

ดังที่กล่าวมาข้างต้น เพื่อจัดการและควบคุมปัญหาที่อาจเกิดขึ้น สิ่งสำคัญสำหรับผู้ประกอบการ P2P Lending คือ การมีเครื่องมือที่สามารถใช้วัดแนวโน้มการลงทุนและคัดกรองผู้กู้ อย่างมีประสิทธิภาพ ตามระดับความน่าเชื่อถือของผู้กู้ที่แท้จริง โดยการใช้แบบจำลองการพยากรณ์ (Predictive Model) เพื่อช่วยทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ สำหรับแพลตฟอร์ม P2P Lending

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาตัวแปรที่ใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ สำหรับแพลตฟอร์ม P2P Lending

1.2.2 เพื่อสร้างแบบจำลองพยากรณ์สำหรับทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ด้วยวิธีการทำเหมืองข้อมูล

1.2.3 เพื่อพัฒนาต้นแบบ (Prototype) ของระบบการทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ด้วยวิธีการทำเหมืองข้อมูล

1.3 สมมติฐานและเงื่อนไขของการวิจัย

1.3.1 ระดับความเสี่ยงของผู้กู้ มีผลต่อการตัดสินใจในการลงทุน P2P Lending

1.3.2 ข้อมูลที่นำมาใช้ในการทำแบบทำนายเป็นข้อมูลจริง

1.3.3 แบบจำลองทำนายพยากรณ์ที่สร้างจากข้อมูลจริงจากผู้ประกอบการ P2P Lending ในสหรัฐ สามารถนำมาใช้เป็นแนวทางสำหรับผู้ประกอบการไทยได้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ทำให้ทราบถึงปัจจัยที่ใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ สำหรับแพลตฟอร์ม P2P Lending

1.4.2 ทำให้ทราบถึงแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ หลังจากการลงทุนใน P2P Lending

1.4.3 ผู้ประกอบการสามารถนำไปใช้เป็นแนวทางในการรวบรวมข้อมูลเกี่ยวกับผู้กู้และเงินกู้ที่เพียงพอต่อการตัดสินใจของนักลงทุน รวมถึงสามารถใช้เป็นแนวทางในคัดกรองผู้กู้ เพื่อลดปัญหาการผิดนัดชำระหนี้ และนำไปใช้ในการพัฒนาระบบแพลตฟอร์มให้มีประสิทธิภาพต่อไป

1.4.4 นักลงทุนสามารถนำไปใช้เป็นแนวทางประกอบการตัดสินใจในการลงทุน P2P Lending

1.4.5 ผู้กู้สามารถนำไปใช้เป็นแนวทางในการเปิดเผยข้อมูลที่เพียงพอต่อการตัดสินใจของนักลงทุน เพื่อเป็นประโยชน์ต่อการขออนุมัติเงินกู้ใน P2P Lending

1.5 ขอบเขตของการวิจัย

1.5.1 ขอบเขตด้านข้อมูล

งานวิจัยนี้ใช้กลุ่มตัวอย่างจากข้อมูลสินเชื่อจริงจาก Lending Club ซึ่งเป็นผู้ประกอบการ P2P Lending รายใหญ่ในสหรัฐอเมริกา โดยใช้ข้อมูลจากผู้กู้จำนวน 503,421 ราย ในปี 2009-2015

ข้อมูลด้านเศรษฐกิจมหภาค ตามช่วงเวลาที่คุณ์ออกคำขออนุมัติเงินกู้ สำหรับผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) อัตราการว่างงาน อัตราเงินเฟ้อ ดัชนีราคาอสังหาริมทรัพย์ของสหรัฐ จาก Organization for Economic Co-operation and Development (OECD Database) และผลตอบแทนจากดัชนีหลักทรัพย์ S&P 500 จาก Morning Star

1.5.2 ขอบเขตจากความไม่เท่าเทียมกันของข้อมูลที่ได้รับ

เนื่องจากนักลงทุนใน P2P Lending สามารถเข้าถึงข้อมูลของผู้กู้ที่แท้จริงได้อย่างจำกัด จากปัญหาความไม่เท่าเทียมของข้อมูลระหว่างผู้กู้กับนักลงทุน (Information Asymmetry) ดังนั้น การทำวิจัยนี้ จึงสร้างแบบจำลองพยากรณ์จากข้อมูลของผู้กู้และสินเชื่อ รวมถึงข้อมูลด้านนักลงทุนที่สามารถเข้าถึงได้เท่านั้น

1.5.3 ขอบเขตด้านตัวแปรที่จะทำการศึกษา

(1) ตัวแปรด้านผู้กู้และสินเชื่อ ได้แก่ ตัวแปรต้นด้านข้อมูลประวัติทางการเงิน เช่น จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ ระยะเวลาของการชำระสินเชื่อ อัตราดอกเบี้ยของสินเชื่อ งวดชำระต่อเดือน สถานะการเป็นเจ้าของบ้าน รายได้ต่อปี วัตถุประสงค์ที่ขอสินเชื่อ อัตราหนี้สินต่อรายได้ เป็นต้น รวมถึงตัวแปรต้นด้านข้อมูลประวัติส่วนตัว เช่น อาชีพ ระยะเวลาทำงาน ที่อยู่อาศัย เป็นต้น

(2) ตัวแปรทางเศรษฐศาสตร์มหภาค เช่น ผลิตภัณฑ์มวลรวมภายในประเทศ อัตราการว่างงาน อัตราเงินเฟ้อ ผลตอบแทนของดัชนีหลักทรัพย์ ดัชนีราคาอสังหาริมทรัพย์ เป็นต้น

(3) ตัวแปรตาม คือ สถานะที่ผู้กู้จะผิดนัดชำระหนี้หรือไม่

1.6 นิยามศัพท์เฉพาะ

1.6.1 **Peer-to-Peer Lending** หมายถึง ระบบหรือเครือข่ายอิเล็กทรอนิกส์สำหรับธุรกรรมสินเชื่อระหว่างบุคคลกับบุคคล (Peer-to-Peer Lending Platform) ซึ่งจะทำหน้าที่เป็นตัวกลางในการทำสัญญากู้ยืมเงินระหว่างผู้ให้กู้และผู้ขอกู้ที่เป็นบุคคลธรรมดา

1.6.2 **Data Mining** หมายถึง การค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ โดยใช้วิธีการวิเคราะห์ทางธุรกิจที่นอกเหนือจากการนับจำนวน การอธิบายเชิงลักษณะ การรายงาน หรือวิธีการต่างๆที่ยึดตามกฎเกณฑ์เดิมของธุรกิจ แต่รวมถึงการวิเคราะห์เชิงสถิติและ Machine Learning เพื่อให้ได้องค์ความรู้ใหม่ (Knowledge Discovery) ที่อาจจะอยู่ในรูปแบบ กฎเกณฑ์ (Rule)

1.6.3 **Classification** หมายถึง เทคนิค Data Mining สำหรับการจำแนกข้อมูลผ่านการเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่ทราบค่าประเภทของตัวอย่างในชุดข้อมูลสอนล่วงหน้า เป็นการสร้างโมเดลผ่านตัวจำแนกข้อมูล (Classifier) เพื่อทำนายหมวดหมู่ของข้อมูล (Class) โดยมีตัวอย่างในชุดข้อมูลสอน (Training Set) ที่ใช้ จะมีคุณลักษณะหนึ่ง ซึ่งบอกค่าประเภทของตัวอย่างนั้น (Class Label) ซึ่งเป็นค่าข้อมูลแบบ Categorical ทั้งนี้ การประยุกต์ใช้งานแบบ Classification มักพบใน การขออนุมัติคำขอมิบัตรเครดิต (Credit Approval)

1.6.4 **Information Asymmetry** หมายถึง ความไม่เท่าเทียมของข้อมูลระหว่างผู้กู้กับนักลงทุน ซึ่งมาจากข้อเท็จจริงที่ว่าผู้กู้มีข้อมูลมากกว่านักลงทุนอย่างน้อย 2 เงื่อนไขที่จะส่งผลถึงการตัดสินใจปล่อยกู้ของนักลงทุน และความเสี่ยงในการกลับมาลงทุนอีกครั้ง เงื่อนไขแรก คือ ความสามารถและความต้องการที่แท้จริงในการชำระหนี้สินของผู้กู้ และอีกเงื่อนไข คือ แนวโน้มของการชำระหนี้ล่วงหน้า

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีการศึกษาแนวคิด ทฤษฎี จากบทความวิชาการ หนังสือ รวมถึง ข้อมูลทางสารสนเทศ ต่างๆ ผ่านสื่ออินเทอร์เน็ต โดยผู้วิจัยได้ทำการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้อง ตามหัวข้อดังต่อไปนี้

- 2.1 เอกสารและงานวิจัยที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending
- 2.2 เอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยที่ส่งผลต่อแนวโน้มการผิณฑ์ชำระหนี้ของผู้กู้
- 2.3 เอกสารและงานวิจัยที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูล
- 2.4 เอกสารและงานวิจัยที่เกี่ยวข้องกับความสอดคล้องกันระหว่างข้อมูลที่ใช้กับข้อมูลที่เกิดขึ้นจริงในประเทศไทย
- 2.5 เอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้าง Predictive Model เฉพาะเจาะจงสำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้
- 2.6 เอกสารและงานวิจัยที่เกี่ยวข้องกับการออกแบบ User Experience (UX) และ User Interface (UI) เพื่อใช้เป็นแนวทางในการพัฒนาด้านแบบของระบบแพลตฟอร์ม P2P Lending
- 2.7 สรุปเอกสารและงานวิจัยที่ใช้

2.1 เอกสารและงานวิจัยที่เกี่ยวข้องกับแพลตฟอร์ม P2P Lending

Zopa เป็นแพลตฟอร์มแรกของ P2P Lending ที่ก่อตั้งขึ้นตั้งแต่ปี 2005 หลังจากนั้นตลาด P2P Lending ได้มีการพัฒนารูปแบบของแพลตฟอร์มจนกระทั่งมีหลากหลายรูปแบบในเวลาต่อมา โดยแต่ละแพลตฟอร์มจะมีกลไกการทำงานที่แตกต่างกัน และสามารถแบ่งประเภทของแพลตฟอร์มได้ตามแง่มุมต่างๆ ดังนี้

2.1.1 อนุกรมวิธาน (Taxonomy)

(1) แพลตฟอร์มที่มีวัตถุประสงค์ทั่วไป (General Platform) เพื่อทำให้ผู้กู้รายย่อย รวมถึงบริษัทขนาดกลางและเล็กเข้าถึงแหล่งเงินทุนได้มากขึ้น ตัวอย่างเช่น Prosper และ Lending Club เป็นต้น

(2) แพลตฟอร์มที่มีวัตถุประสงค์เฉพาะด้าน (Professional Platform) ซึ่งเกิดจากการพัฒนารูปแบบในภายหลัง ตัวอย่างเช่น AgFunder ที่เป็นแพลตฟอร์มออนไลน์ที่ทำให้นักลงทุนสามารถลงทุนในธุรกิจการเกษตร รวมถึงเทคโนโลยีทางการเกษตร หรือ Kiva ที่มีวัตถุประสงค์เพื่อให้ผู้มีรายได้น้อยในประเทศกำลังพัฒนา สามารถเข้าถึงแหล่งเงินทุนจากประชาชนทั่วไปได้

2.1.2 รูปแบบการซื้อขาย (Trading Rule)

(1) แพลตฟอร์มที่ใช้การประมูล เช่น Prosper เป็นต้น

(2) แพลตฟอร์มที่ไม่ใช่การประมูล เช่น Lending Club และ Kiva เป็นต้น

(3) ผลตอบแทน (Reward) ซึ่งสามารถแบ่งแพลตฟอร์มได้ 4 ประเภทหลัก ได้แก่ แพลตฟอร์มที่เป็นรูปแบบการบริจาค ไม่มีผลตอบแทน (Donation-Based) โดยผู้บริจาคจะให้เงินทุนแก่กิจการร่วมค้า (Joint Venture) เพื่อสนับสนุนในการออกผลิตภัณฑ์หรือบริการ หรือเพื่อสนับสนุนในวัตถุประสงค์ต่างๆ โดยที่ไม่รับผลตอบแทน นอกจากสิทธิประโยชน์ภาษี เช่น Sellaband และ Kickstarter เป็นต้น นอกจากนี้ยังมีแพลตฟอร์มที่ให้ผลตอบแทนที่เป็นของรางวัล หรือสิทธิจากสินค้าและบริการ (Reward-based) โดยจะไม่มีผลตอบแทนที่เป็นรูปแบบเงินแก่นักลงทุน รวมถึงแพลตฟอร์มที่ให้ผลตอบแทนที่เป็นหุ้นของบริษัท (Equity-based) โดยผลตอบแทนที่ได้ จะขึ้นกับผลประกอบการของบริษัทที่ลงทุน และแพลตฟอร์มที่ให้ผลตอบแทนที่เป็นดอกเบี้ยจากการให้กู้ยืม (Lending-based) ซึ่งจะเป็นลักษณะของแพลตฟอร์ม P2P Lending โดยทั่วไป ตัวอย่างเช่น Prosper Lending Club และ Kiva เป็นต้น [9]

2.2 เอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยที่ส่งผลต่อแนวโน้มการผิณฑชำระหนี้ของผู้กู้

2.2.1 ปัจจัยด้านผู้กู้และสินเชื่อ

ปัจจัยด้านผู้กู้และสินเชื่อที่มีผลต่อแนวโน้มการผิณฑชำระหนี้ของผู้กู้ใน P2P Lending ซึ่งมีปัจจัยหลัก 3 ด้าน ได้แก่

- (1) ปัจจัยด้านสินเชื่อ เช่น เงินกู้ที่ต้องชำระทั้งหมด จำนวนเงินลงทุน วันที่ออกเงินกู้ จำนวนเงินกู้ ความถี่ของเงินกู้ที่ต้องชำระแต่ละงวด วัตถุประสงค์ของการขอเงินกู้ เป็นต้น
- (2) ปัจจัยด้านผู้กู้ เช่น รายได้ต่อปีของผู้กู้ ระยะเวลาในการทำงาน ที่อยู่อาศัยของผู้กู้ ระดับความเสี่ยงของผู้กู้ เป็นต้น
- (3) ปัจจัยด้านประวัติการกู้ยืม เช่น วงเงินกู้ปัจจุบันของทุกบัญชีเงินกู้ จำนวนวงเงินสูงสุดที่ผู้กู้เคยได้รับการอนุมัติ จำนวนผิณฑชำระหนี้ที่มีประวัติผิณฑชำระหนี้ เป็นต้น [10]

นอกจากนี้ ปัจจัยด้านผู้กู้และสินเชื่อที่จะส่งผลต่อแนวโน้มที่ผู้กู้แต่ละรายจะมีการผิณฑชำระหนี้หรือไม่ สามารถแบ่งได้เป็น 2 ประเภทหลักดังนี้

- (1) ข้อมูลด้านฐานะทางการเงิน (Hard Information) ซึ่งสามารถวัดได้จากระดับคะแนนของสินเชื่อ (Credit Grade) อัตราส่วนหนี้สินต่อรายได้ จำนวนบัตรเครดิตของธนาคาร จำนวนสินเชื่อที่ขออนุมัติ ประวัติการขอสินเชื่อ ความเป็นเจ้าของบ้าน โดย Credit Grade และอัตราส่วนหนี้สินต่อรายได้เป็นปัจจัยที่สำคัญที่สุดที่นักลงทุนใช้ในการตัดสินใจลงทุน ผู้กู้ที่มี Credit Grade สูง ขณะที่หนี้สินต่อรายได้ต่ำ จะมีโอกาสในการได้รับอนุมัติเงินกู้ได้ง่ายกว่า เนื่องจากมี

แนวโน้มที่จะผิดนัดชำระหนี้มีน้อยกว่า ส่วนความเป็นเจ้าของบ้านมีส่วนในการตัดสินใจลดลง นับตั้งแต่วิกฤติสินเชื่อซับไพรม์ในปี 2550 [11]

ปัจจัยในการตัดสินใจลงทุน P2P Lending ทางด้านผู้กู้ว่า Hard Information มีความสำคัญอย่างมีนัยสำคัญต่อโอกาสที่ผู้กู้จะได้รับการอนุมัติสินเชื่อ โดยระดับคะแนนสินเชื่อ (Credit Rating) มีความสัมพันธ์ในทิศทางเดียวกับโอกาสที่จะได้รับเงินทุน กล่าวคือ ผู้กู้ที่มี Credit Rating สูง จะมีความเสี่ยงทางสินเชื่อต่ำ และมีโอกาสได้รับเงินทุนมากกว่า อีกทั้งนักลงทุนมักจะพิจารณาว่าผู้กู้ที่มี Credit Rating สูง จะมีโอกาสในการผิดนัดชำระหนี้มีน้อยลงอีกด้วย อีกตัวแปรที่มีความสำคัญเช่นกัน คือ อัตรารายได้ต่อรายได้ (Payment-to-Income Ratio) และอัตราหนี้สินต่อรายได้ (Debt-to-Income Ratio) โดยนักลงทุนถือว่า ผู้ที่มีอัตราส่วนทางการเงินดังกล่าวสูงมักจะมีโอกาสในการผิดนัดชำระหนี้มากกว่า ทำให้ผู้กู้กลุ่มนี้มีโอกาสในการได้รับการอนุมัติเงินกู้มีน้อยกว่า ดังนั้นในงานวิจัยนี้จึงมีการนำ Credit Rating และอัตราหนี้สินต่อรายได้ มาเป็นหนึ่งในปัจจัยที่ใช้ในการสร้าง Predictive Model [7]

นอกจากนี้ รายได้ต่อเดือน (Annual Income) ของผู้กู้ ยังมีส่วนในการตัดสินใจของนักลงทุน โดยผู้ที่มีรายได้ต่อเดือนสูง มีแนวโน้มที่จะได้รับเงินทุนจากนักลงทุนมากกว่า รวมถึงยังมีโอกาสในการผิดนัดชำระหนี้มีน้อยกว่า อีกทั้งผู้ที่มีบ้านเป็นของตนเอง (Home Ownership) จะมีแนวโน้มที่ผู้กู้นั้นจะได้รับการอนุมัติเงินกู้มากกว่า เพราะนักลงทุนพิจารณาว่าจะมีโอกาสในการผิดนัดชำระหนี้มีน้อยกว่า ดังนั้นในงานวิจัยนี้จึงมีการนำรายได้ต่อเดือนและการมีบ้านเป็นของตนเอง มาเป็นหนึ่งในปัจจัยที่ใช้ในการสร้าง Predictive Model อีกทั้งยังกล่าวอีกปัจจัยที่สำคัญ คือ วัตถุประสงค์การกู้ (Loan Purpose) โดยผู้ที่มีวัตถุประสงค์ในการกู้ยืมเพื่อทำบัตรเครดิตและเพื่อใช้ในการทำธุรกิจขนาดเล็ก จะมีแนวโน้มที่จะเกิดการผิดนัดชำระหนี้มากกว่า ทำให้มีโอกาสในการได้รับเงินทุนจากนักลงทุนน้อยกว่า ดังนั้นในงานวิจัยนี้จึงมีการนำวัตถุประสงค์ในการกู้ มาเป็นหนึ่งในปัจจัยที่ใช้ในการสร้าง Predictive Model

(2) ข้อมูลส่วนตัว (Soft Information) ซึ่งสามารถวัดได้จากข้อมูลด้านประชากร ข้อมูลด้านเครือข่ายทางสังคม รูปภาพ และคำที่ใช้ในแบบคำขออนุมัติเงินกู้ โดยผู้กู้ที่มีเครือข่ายทางสังคมจำนวนมาก จะมีแนวโน้มที่ได้รับการอนุมัติเงินกู้ได้ง่ายกว่า ขณะที่แบบคำขอที่มีข้อความที่สะกดผิด จำนวนความยาวในการเขียนคำขอ รวมถึงถ้อยคำที่สื่อถึงอารมณ์ในการเขียนคำขอ จะส่งผลต่อความยากง่ายในการขออนุมัติ นอกจากนี้ผู้กู้ที่มีการเปิดเผยข้อมูลมากกว่า มีโอกาสที่จะได้รับการอนุมัติได้ง่ายกว่า ด้วยอัตราดอกเบี้ยที่ต่ำกว่า เนื่องจากมีแนวโน้มที่จะผิดนัดชำระหนี้มีน้อยกว่า

นักลงทุนยังใช้ Soft Information ของผู้กู้ในการประกอบการตัดสินใจลงทุน P2P Lending เนื่องจากข้อมูลด้าน hard Information ของผู้กู้เพียงอย่างเดียว อาจไม่เพียงพอต่อการตัดสินใจ ดังนั้นข้อมูลดังกล่าวจึงถือว่าเป็นบทบาทต่อการตัดสินใจอย่างมาก โดยเฉพาะในกลุ่มผู้กู้ที่มี

Credit Rating ต่ำ เนื่องจากมีความเสี่ยงที่สูง และต้องใช้ข้อมูลที่หลากหลายในการพิจารณาสำหรับการลงทุน ทั้งนี้ข้อมูลด้านอายุ การศึกษา สถานภาพ และอาชีพ ส่งผลสำคัญต่อโอกาสในการได้รับการอนุมัติเงินกู้ โดยผู้กู้ที่มีอายุมากกว่า มีการศึกษาที่สูงกว่า สถานภาพแต่งงานแล้ว มีอาชีพ และมีการทำงานที่ยาวนานกว่า จะมีแนวโน้มที่จะได้รับเงินทุนจากนักลงทุนมากกว่า เนื่องจากนักลงทุนพิจารณาว่าเป็นกลุ่มผู้กู้ที่มีความเสี่ยงต่อการผิดนัดชำระหนี้ต่ำกว่า อย่างไรก็ตาม จากข้อจำกัดด้านข้อมูล งานวิจัยนี้จึงใช้เพียงอาชีพ และระยะเวลาในการทำงาน มาเป็นหนึ่งในปัจจัยที่ใช้ในการสร้าง Predictive Model [1]

อย่างไรก็ตาม พบว่าข้อมูลด้านเพศของผู้กู้ จะไม่ส่งผลต่อโอกาสในการได้รับอนุมัติเงินกู้ และแนวโน้มการผิดนัดชำระหนี้ในกรณีที่ผู้กู้มี Credit Rating ที่เท่ากัน ผู้กู้เพศหญิงจะมีโอกาสในการผิดนัดชำระหนี้ได้เท่ากับผู้กู้เพศชาย ดังนั้นในงานวิจัยนี้จึงไม่นำปัจจัยด้านเพศในการสร้าง Predictive Model [7]

2.2.2 ปัจจัยด้านเศรษฐกิจมหภาค

การผิดนัดชำระหนี้สามารถเกิดจากปัจจัยทางด้านสิ่งแวดล้อมในการกู้ยืม รวมถึงปัจจัยทางด้านมหภาค เช่น ผลิตภัณฑ์มวลรวมภายในประเทศ อัตราการว่างงาน การเปลี่ยนแปลงของตลาดหุ้นหรือตลาดอสังหาริมทรัพย์ เป็นต้น เมื่อเศรษฐกิจเชิงมหภาคชะลอตัว จะทำให้ผู้ว่างงานมากขึ้น ทำให้มีผู้กู้ที่ต้องการความช่วยเหลือด้านการเงินเพิ่มมากขึ้น และมีผู้กู้จำนวนเพิ่มขึ้นที่ไม่ได้รับการอนุมัติให้กู้ผ่านธนาคาร เนื่องจากความรัดกุมในการปล่อยสินเชื่อ ส่งผลให้ผู้กู้กลุ่มดังกล่าวเข้ามาขอกู้ยืมผ่านแพลตฟอร์ม P2P Lending เพิ่มขึ้น และมีแนวโน้มของการผิดนัดชำระหนี้เพิ่มขึ้นในที่สุด

หากตลาดหุ้นมีผลตอบแทนที่ลดลง จะทำให้รายได้ของประชาชนลดลง ส่งผลให้มีความต้องการในการลงทุนในกลุ่มระดับความเสี่ยงสูง ที่ให้ผลตอบแทนดอกเบี้ยสูงเพิ่มขึ้น ส่งผลให้แนวโน้มการผิดนัดชำระหนี้ของผู้กู้เพิ่มสูงขึ้น อย่างไรก็ตาม เนื่องจากแพลตฟอร์ม P2P Lending ถือว่าเป็นช่องทางเลือกสำหรับการลงทุน นอกเหนือจากตลาดหลักทรัพย์ ดังนั้นการที่ตลาดหุ้นให้ผลตอบแทนที่ลดลง จะทำให้นักลงทุนหันมาลงทุนในแพลตฟอร์ม P2P Lending เพิ่มขึ้นได้เช่นเดียวกัน

สำหรับการราคาของที่อยู่อาศัยลดลง จะทำให้หลักประกันของผู้กู้ในธนาคารมีมูลค่าลดลง ซึ่งเรียกว่าปรากฏการณ์ Collateral Effect ส่งผลให้ผู้กู้ได้รับการอนุมัติเงินกู้จากธนาคารลดลง การกู้ยืมผ่านแพลตฟอร์ม P2P Lending เพิ่มขึ้น และทำให้แนวโน้มของการผิดนัดชำระหนี้เพิ่มสูงขึ้น

นอกจากนี้ หากรัฐบาลออกมาตรการในการควบคุมความเสี่ยงในการกู้ยืมผ่านแพลตฟอร์ม P2P Lending ตัวอย่างเช่น หลายสถาบันการเงินในประเทศจีนได้ร่วมกันออกมาตรการในการส่งเสริมการพัฒนา แพลตฟอร์มอย่างมีประสิทธิภาพ จะเป็นการลดความเสี่ยงของการผิดนัดชำระหนี้ในการกู้ยืมผ่าน P2P Lending โดยการกำหนดให้แพลตฟอร์มจะต้องยื่นงบการเงิน เพื่อให้สอดคล้องกับ

เกณฑ์ของกระทรวงอุตสาหกรรมและเทคโนโลยีสารสนเทศ จากมาตรการดังกล่าวส่งผลให้การเปิดเผยข้อมูลโปร่งใสเพิ่มขึ้นและลดการเผยแพร่ข้อมูลปลอมบนเว็บไซต์ของ P2P Lending [12]

2.3 เอกสารและงานวิจัยที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูล

2.3.1 การทำเหมืองข้อมูล

เหมืองข้อมูล (Data Mining) หมายถึง การค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ [13] โดยการทำเหมืองข้อมูลเป็นวิธีการวิเคราะห์ทางธุรกิจที่นอกเหนือจาก การนับจำนวน การอธิบายเชิงลักษณะ การรายงาน หรือวิธีการต่างๆที่ยึดตามกฎเกณฑ์เดิมๆของธุรกิจ แต่รวมถึงการวิเคราะห์เชิงสถิติและ Machine Learning เพื่อให้ได้องค์ความรู้ใหม่ (Knowledge Discovery) ที่อาจจะอยู่ในรูปแบบ กฎเกณฑ์ (Rule) ทั้งนี้ Big Data เป็นตัวกระตุ้นให้เกิดการใช้ Data Mining เพิ่มขึ้น เพราะสามารถที่จะจัดการกับข้อมูลที่มีจำนวนและมูลค่ามหาศาลได้ [14] ทั้งนี้ขั้นตอนการทำ Data Mining ที่นิยมใช้ คือ Cross-Industry Standard Process for Data Mining (CRISP-DM) ซึ่งใช้ในแปลงปัญหาทางธุรกิจให้เป็นขั้นตอนต่างๆของ Data Mining โดยขั้นตอนของ CRISP-DM มีดังนี้

(1) ขั้นตอนการทำความเข้าใจธุรกิจ (Business Understanding) คือ การจัดหาจุดประสงค์ทางธุรกิจที่เฉพาะเจาะจง ที่ระบุโดยผู้ที่มีสิทธิ์หลักในการตัดสินใจ และรวบรวมข้อมูลพื้นฐานของสภาพธุรกิจในปัจจุบัน

(2) ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding) คือ การจัดเก็บข้อมูลเบื้องต้นทั้งที่มาจากข้อมูลเดิม ซ้ำข้อมูล สุ่ม หรือเริ่มติดตามข้อมูลเพิ่มเติม รวมถึงอธิบายข้อมูลทั้งเชิงปริมาณและคุณภาพ ว่าข้อมูลมีมากพอแค่ไหน และมีเงื่อนไขอะไรบ้าง

(3) ขั้นตอนการจัดเตรียมข้อมูล (Data preparation) เช่น การจัดเรียงข้อมูลสำหรับการทำโมเดล แล้วตัดหรือแทนค่าว่าง หรือค่าที่มีการผิดพลาด

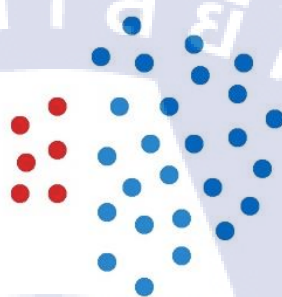
(4) ขั้นตอนการทำโมเดล (Modelling) ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised Learning) ตัวอย่างเช่น Classification ซึ่งเป็นการจัดกลุ่มข้อมูลด้วยตัวแปรที่กำหนดไว้แล้ว เพื่อสร้างแบบจำลองมาทำนายข้อมูลในอนาคต เช่น Decision Tree Naïve Bayes K-Nearest Neighbours Neural Network เป็นต้น และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ตัวอย่างเช่น Clustering ซึ่งเป็นการจัดข้อมูลที่มีลักษณะคล้ายกันในกลุ่มเดียวกัน โดยไม่จัดกลุ่มตัวอย่างไว้ล่วงหน้า

(5) ขั้นตอนการประเมินผล (Evaluation) ซึ่งเป็นการประเมินผลของโครงการกับเงื่อนไขที่ตั้งไว้ตั้งแต่ขั้นตอนการทำความเข้าใจธุรกิจ

(6) ขั้นตอนการประยุกต์ใช้ (Deployment) ซึ่งเป็นการนำองค์ความรู้ใหม่ที่ได้มาพัฒนา หรือแสดงถึงการเปลี่ยนแปลงที่เกิดขึ้นได้ [15]

2.3.2 เทคนิคการจัดเตรียมข้อมูล (Data Preparation)

(1) เทคนิคการสังเคราะห์ข้อมูลใหม่ (Synthetic Minority Oversampling Technique: SMOTE) โดย SMOTE คือ เทคนิคที่ใช้ในการจัดกลุ่มข้อมูลที่มีความไม่สมดุล (Imbalanced data) โดยจะทำการสุ่มข้อมูลในกลุ่มรองเพื่อสร้างข้อมูลใหม่ของกลุ่มรองให้มีจำนวนเพิ่มมากขึ้น จนใกล้เคียงหรือเท่ากับจำนวนกลุ่มหลักที่เดิมมีมากกว่า ทำให้เทคนิคนี้สามารถแก้ปัญหาที่เกิดขึ้น และเพิ่มความแม่นยำให้กับเทคนิค classification แบบต่างๆได้ รวมถึงเทคนิค Random Forest [16]



รูปที่ 2.1 เทคนิคการสังเคราะห์ข้อมูลใหม่

ทั้งนี้ เทคนิค SMOTE จะทำการสุ่มเพิ่ม โดยสังเคราะห์ข้อมูลขึ้นมาใหม่จากข้อมูลเดิม โดยมีขั้นตอนคือ การระบุเพื่อนบ้านที่ใกล้ที่สุด k ค่าของข้อมูลเดิมหรือจุด M จากนั้นสุ่มเลือกจุดใหม่ระหว่าง 2 จุด คือ $M(a_1, a_2, \dots, a_n)$ และ $M(b_1, b_2, \dots, b_n)$ เพื่อนำมาสังเคราะห์เป็นข้อมูลใหม่หรือจุด $m(c_1, c_2, \dots, c_n)$ [17]

$$C_1 = a_1 + (b_1 - a_1) \times \text{rand} \text{ ('ตัวแปรข้อมูล')}$$

$$C_2 = a_2 + (b_2 - a_2) \times \text{rand} \text{ ('ตัวแปรข้อมูล')}$$

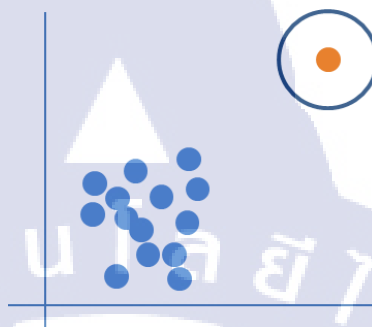
....

....

$$C_n = a_n + (b_n - a_n) \times \text{rand} \text{ ('ตัวแปรข้อมูล')}$$

เมื่อ	M	คือ ข้อมูลเดิม
	M	คือ ข้อมูลใหม่
	K	คือ จำนวนเพื่อนบ้านที่ใกล้เคียงที่สุดของข้อมูลเดิม (จุด M)

(2) เทคนิคการกำจัดข้อมูลผิดปกติ (Outlier) โดยเทคนิคการจัดกลุ่มข้อมูลแบบ DBSCAN (Density-Based Spatial Clustering of Applications with Noise: DBSCAN Clustering)



รูปที่ 2.2 เทคนิคการกำจัดข้อมูลผิดปกติ

ทั้งนี้ เทคนิค DBSCAN เป็นเทคนิคแบบ Clustering ที่ใช้ในการแบ่งกลุ่มและแยก Outlier ที่มีในข้อมูลโดยใช้ตัวแปรหลัก 2 ตัวแปร คือ เอปซิลอน (Epsilon: Eps) ซึ่งเป็นขนาดรัศมีของข้อมูลเพื่อนบ้านของจุดหลัก x (Neighbourhood of X) และจำนวนจุดขั้นต่ำ (Minimum Points: MinPts) ซึ่งเป็นจำนวนเพื่อนบ้านขั้นต่ำภายในรัศมี eps นอกจากนี้เมื่อใช้เทคนิค DBSCAN Clustering ควบคู่กับเทคนิค SMOTE จะช่วยเพิ่มความแม่นยำให้กับ Predictive Model หลังจาก Outlier ถูกกำจัดออกไป การจัดเตรียมข้อมูลเพื่อเพิ่มความแม่นยำของแบบจำลองพยากรณ์ ด้วยวิธีการทำเหมืองข้อมูลควบคู่กับเทคนิค SMOTE เพื่อใช้ในการแก้ปัญหา imbalanced data นอกจากนี้ยังใช้เทคนิค DBSCAN Clustering ในการกำจัด outlier ในชุดข้อมูล [18]

2.3.3 เทคนิคการลดตัวแปร

การลดตัวแปรด้วยการคัดเลือกตัวแปร (Feature Selection) สำหรับแบบจำลอง Classification ด้วยวิธีต่างๆ ได้แก่

(1) Filter Model ซึ่งประกอบด้วย 2 ขั้นตอนหลัก คือ การเรียงลำดับความสำคัญของตัวแปรจากเงื่อนไขที่ระบุ จากนั้นตัวแปรที่มีความสำคัญที่สุดจะถูกคัดเลือกนำมาใช้ในแบบจำลองพยากรณ์ โดยเงื่อนไขที่นิยมนำมาใช้ ได้แก่ การจัดลำดับให้ตัวแปรที่อยู่กลุ่มเดียวกันมีค่าเท่ากัน ตัวแปรต่างกลุ่มมีค่าต่างกัน [19] แล้วคัดตัวแปรที่มีค่า Pearson's Correlation สูง ซึ่งเป็นวิธีที่ใช้เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปร ผ่านหลักการว่า หากตัวแปรมีความเกี่ยวข้องสูงกับประเภทของ

ผลลัพธ์ข้อมูล มีแนวโน้มที่จะเป็นตัวแปรที่ดีในการใช้จัดประเภทข้อมูล อย่างไรก็ตามการคัดเลือกตัวแปรด้วยวิธี Filter Model มีข้อเสีย เนื่องจากวิธีนี้จะไม่สนใจผลกระทบของตัวแปรที่เลือก ต่อประสิทธิภาพของแอลกอริทึม (Algorithm) ที่ใช้ ทำให้ตัวแปรที่คัดเลือก จะขึ้นกับเงื่อนไขที่มีอคติ (Bias) และ Algorithm ที่เลือกใช้ [20]

(2) Wrapper Model ซึ่งถูกพัฒนามาเพื่อแก้ปัญหาที่เกิดจากการคัดเลือกตัวแปรแบบ Filter Method โดยใช้ตัวจัดคลาส (Classifier) เพื่อประเมินคุณภาพของการคัดเลือกตัวแปร รวมถึงการใช้วิธีการต่างๆ เพื่อแก้ปัญหาในการคัดเลือกตัวแปร โดยไม่มีผลกระทบจากการเลือกใช้ Algorithm ด้วยวิธีที่แตกต่างกัน ทำให้แบบจำลองมีประสิทธิภาพมากกว่าการคัดเลือกตัวแปรผ่านวิธี Filter Model แต่ยังคงมีปัญหาด้านการประมวลผลที่มีต้นทุนสูงกว่า Filter Model ทั้งนี้ Wrapper Model มีวิธีที่หลากหลาย ได้แก่ Recursive Feature Elimination (RFE) [21] ซึ่งเป็นวิธีที่ใช้การนำตัวแปรออกจากโมเดลเรื่อยๆ และสร้างโมเดลจากตัวแปรที่เหลืออยู่ โดยใช้ค่าความถูกต้อง (Accuracy Matric) ในการจัดลำดับค่าความสำคัญของตัวแปร เรียงจากตัวแปรที่สำคัญที่สุดจะมีค่า 1 หรือใช้ค่าที่บอกถึงตัวแปรที่เกี่ยวข้อง (True) และตัวแปรที่ไม่เกี่ยวข้อง (False) [22]

(3) Embedded Model ซึ่งเป็นการคัดเลือกตัวแปรที่ไม่ได้ใช้การนำเข้า หรือนำออกจากโมเดล แต่เป็นการคัดเลือกภายในตัวแปร ทั้งนี้การคัดเลือกตัวแปรด้วยวิธีนี้ ทำให้มีข้อได้เปรียบทั้งจากวิธี Wrapper Model ที่สามารถเพิ่มประสิทธิภาพของโมเดล ไม่ว่าจะใช้โมเดล Classification แบบใด ขณะเดียวกันก็ยังคงมีการประมวลผลได้ด้วยต้นทุนที่ถูกเช่นเดียวกับ Filter Model แต่มีความช้าในการประมวลผล [21]

นอกจากนี้ยังมีการใช้วิธีสำหรับการลดตัวแปรแบบ Factor Analysis ซึ่งเป็นการวิเคราะห์ตัวแปรที่ใช้ในการทำนาย โดยการค้นหาตัวแปรที่มีความสำคัญ โดยการดึงค่าความแปรปรวนที่ตัวแปรทุกตัวร่วมกัน (Common Variance) สูงสุดจากทุกตัวแปร เพื่อใส่เข้าไปในค่าคะแนนที่ใช้ร่วมกัน (Common Score) ซึ่งวิธีนี้นิยมนำมาใช้ในการวิจัยทางการตลาด ข้อมูลด้านการโฆษณา ข้อมูลด้านจิตวิทยา รวมถึงข้อมูลทางการเงิน เช่นเดียวกับงานวิจัยนี้ โดยวิธีดังกล่าวจะใช้กลไกในการค้นหาว่าตัวแปรที่สังเกต (Observable Factors) ใดมีความเกี่ยวข้องเชิงเส้นกับตัวแปรที่ไม่ได้สังเกต (Unobservable Factors) ที่มีจำนวนน้อยกว่า ทำให้วิธีนี้มีข้อดีในการใช้งานคือ ใช้ได้ดีกับข้อมูลขนาดใหญ่ เพื่อหาความเกี่ยวเนื่องกันของตัวแปร สำหรับการลดตัวแปรที่สังเกตให้เป็นตัวแปรที่ไม่ได้สังเกตที่มีจำนวนน้อยลง เพื่อเป็นการเพิ่มประสิทธิภาพของแบบจำลอง อย่างไรก็ตามวิธีดังกล่าวยังมีข้อจำกัดที่ว่า ไม่ควรมีข้อมูลผิดปกติ (Outlier) ขนาดของข้อมูลควรมากกว่าจำนวนตัวแปร ตัวแปรไม่ควรมีความเกี่ยวเนื่องกันอย่างสมบูรณ์ (Perfect Multicollinearity) และไม่ควรมีความแปรปรวนที่เท่ากัน (Homoscedasticity) ระหว่างแต่ละตัวแปร [23]

ดังนั้นในงานวิจัยนี้จะทำการทดลองเปรียบเทียบประสิทธิภาพของแบบจำลองที่ผ่านการลดตัวแปร จาก 3 วิธี คือ Pearson's Correlation Recursive Feature Elimination (RFE) และ Factor Analysis

2.3.4 การแก้ปัญหาข้อมูลที่ขาดหาย (Missing Value)

การแก้ปัญหาข้อมูลที่ขาดหายหรือ Missing Value ที่เกิดขึ้นในข้อมูลที่น่ามาวิเคราะห์ สำหรับการสร้างแบบจำลองในการทำนาย ด้วยการนำแถวของข้อมูลที่มีข้อมูลในแต่ละตัวแปรน้อยกว่าเกณฑ์ที่กำหนดออกไป เช่น แถวที่มีข้อมูลส่วนใหญ่ 70-75% เป็น Missing Value เป็นต้น ซึ่งวิธีดังกล่าวเป็นการทำให้แบบจำลองที่ใช้ทำนายมีความถูกต้องแม่นยำเพิ่มมากขึ้น เนื่องจากได้นำแถวที่ไม่มีข้อมูลที่เป็นนัยสำคัญ และมีน้ำหนักของข้อมูลน้อยออกไปจากแบบจำลอง อย่างไรก็ตามวิธีดังกล่าว อาจทำให้เกิดการสูญเสียข้อมูลได้ จึงเหมาะสำหรับการใช้งานกับข้อมูลที่มีจำนวนแถวของข้อมูลที่มี Missing Value น้อยเมื่อเทียบกับข้อมูลทั้งหมด หรือมีข้อมูลจำนวนที่เพียงพอ ดังเช่นข้อมูลที่ใช้ในการทดลองนี้

นอกจากนี้ยังได้แนะนำการใช้ค่าเฉลี่ยของข้อมูลกลุ่มที่มีค่าใกล้เคียงกับกลุ่มของข้อมูล เพื่อเติมข้อมูล Missing Value ซึ่งจะใช้กับตัวแปรที่เป็นตัวเลข (Numeric Feature) ซึ่งวิธีดังกล่าวสามารถป้องกันการสูญเสียข้อมูลที่เกิดขึ้นจากการนำแถวที่มีข้อมูล Missing Value จำนวนมากออกไปได้ อย่างไรก็ตามวิธีดังกล่าวอาจเป็นการเพิ่มความแปรปรวน (Variance) และอคติ (Bias) กับข้อมูลที่น่ามาใช้ได้ [24]

2.3.5 การเตรียมข้อมูลเชิงปริมาณ (Scale Variable)

การเตรียมข้อมูลเชิงปริมาณด้วยวิธี Min-Max Scaling หรือที่เรียก Normalization ผ่านการทำให้ข้อมูลเรียงจากค่า 0 ไป 1 ซึ่งวิธีดังกล่าวจะทำให้ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ลดลง และทนต่อผลกระทบจากข้อมูลที่ผิดปกติ (Outlier) อีกทั้งยังเป็นการลดขนาดของข้อมูลผ่านการแปลงค่า [25] ด้วยสมการด้านล่าง ดังต่อไปนี้

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

เมื่อ X_{sc} คือ ค่าที่ผ่านการเตรียมข้อมูลด้วยวิธี Min-Max Scaling เรียบร้อยแล้ว
 X คือ ค่าที่ก่อนการเตรียมข้อมูลด้วยวิธี Min-Max Scaling
 X_{min} คือ ค่าที่น้อยที่สุดในข้อมูลของตัวแปร Scale Variable แต่ละตัวแปร

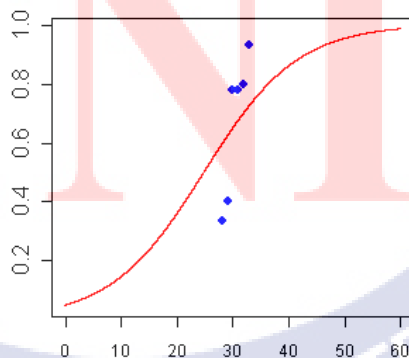
$X_{\max}$ คือ ค่าที่มากที่สุดที่สุดในข้อมูลของตัวแปร Scale Variable แต่ละตัวแปร

2.3.6 การเตรียมข้อมูลเชิงคุณภาพ (Nominal และ Ordinal Variable)

ด้วยการสร้างตัวแปรจำลอง (Dummy หรือ Indicator Variable) เพื่อใช้แทนค่าของตัวแปรแบบจัดกลุ่ม (Nominal Variable) หรือตัวแปรที่สามารถไล่ระดับได้ (Ordinal Variable) ซึ่งวิธีดังกล่าวเป็นการทำให้ตัวแปรเชิงคุณภาพ ถูกแปลงเป็นตัวเลข และสามารถนำมาใช้วิเคราะห์ในสมการถดถอย (Regression) ได้ ผ่านการแทนค่าของตัวแปรแต่ละค่าเป็นค่าตัวเลข โดยจำนวนของ Dummy Variable จะเท่ากับจำนวนกลุ่มหรือระดับของตัวแปรนั้นๆ หักออกด้วยหนึ่งเสมอ ซึ่ง Dummy Variable หนึ่งตัวแปรจะไม่สามารถเป็นค่าคงที่ของความสัมพันธ์เชิงเส้นอันอื่นได้ และหากต้องการแทนค่าเพื่อแสดงความสัมพันธ์ระหว่าง 2 ตัวแปร เช่น เพศและสถานภาพ จะสามารถแทนค่าด้วย Dummy Variable ค่าที่สาม ซึ่งเท่ากับผลคูณของค่า Dummy Variable ของสองตัวแปรคู่กัน

2.3.7 เทคนิคการทำเหมืองข้อมูลแบบต่างๆ

(1) เทคนิคถดถอยโลจิสติกส์ทวิ (Binary Logistic Regression) ซึ่งเป็นการทำเหมืองข้อมูลที่มีผู้สอน (supervised Learning) ใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวแปร และตัวแปรตามแบบจัดกลุ่ม (Categorical Dependent Variable) รวมถึงการพยากรณ์ความน่าจะเป็นที่จะเกิดขึ้นของเหตุการณ์ (Odd) โดยการบรรจุข้อมูลลงในกราฟโลจิสติกส์ (Logistic Curve) โดยจะมีตัวแปรตาม 2 ค่า ตามจำนวนเหตุการณ์ที่เป็นไปได้ ในการวิเคราะห์แบบจำลองประเภทถดถอยโลจิสติกส์ ค่าสัมประสิทธิ์ที่คำนวณได้ [26] จะถูกนำมาเขียนเป็นสมการถดถอยโลจิสติกส์ เพื่อที่จะทำนายความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ ดังนี้



รูปที่ 2.3 เทคนิคถดถอยโลจิสติกส์ทวิ

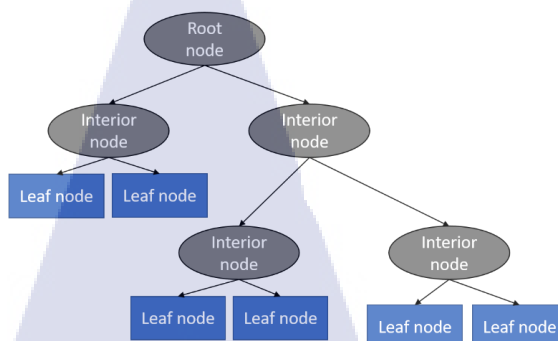
$$P_{(y=1)} = \frac{1}{1+e^{-z}}$$

- เมื่อ
- p (y) คือ ความน่าจะเป็นของการเกิดเหตุการณ์ y
 - e คือ exponential function (e = 2.71828)
 - z คือ linear combination ของตัวแปร; $z = b_0 + b_1x_1 + \dots + b_px_p$

โดยเทคนิคดังกล่าวสามารถใช้ได้ดีกับข้อมูลที่มีขนาดใหญ่ และมีตัวแปรต้นจำนวนมาก เช่นเดียวกับข้อมูลที่ใช้ในงานวิจัยนี้ อีกทั้งยังมีประสิทธิภาพในการจัดการข้อมูลที่ไม่ได้มีการแจกแจงแบบปกติ (Non-Normal Distribution) ซึ่งทำให้วิธีสมการถดถอย (Regression) แบบทั่วไปใช้ไม่ได้ผล

เหตุผลที่นำวิธีดังกล่าวมาใช้ คือ ใช้ได้ดีกับข้อมูลที่มีขนาดใหญ่ มีตัวแปรต้นจำนวนมาก สามารถใช้ได้กับการแจกแจงที่ไม่ปกติ (Non-Normal Distribution) ได้ อีกทั้งยังใช้ได้กับตัวแปรตามแบบกลุ่ม (Categorical Variable) และมีแนวโน้มที่จะเกิด Over-Fitting น้อยหากมีข้อมูลจำนวนที่มากเพียงพอ อย่างไรก็ตามวิธีดังกล่าวใช้เวลาในการประมวลผลช้า และไม่สามารถจัดการกับข้อมูลที่มีข้อมูลขาดหาย (Missing Value) และข้อมูลผิดปกติ (Outlier) ได้ รวมถึงหากข้อมูลมีจำนวนน้อยกว่าจำนวนตัวแปร อาจเกิดปัญหา Over-fitting ได้ [26]

(2) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยเป็นเทคนิคที่ใช้การแบ่งขั้นตอนการตัดสินใจจนกว่าจะได้ชุดข้อมูลที่สอดคล้องกับผลลัพธ์ที่ต้องการจัดประเภทมากที่สุด ทั้งนี้ ส่วนประกอบของต้นไม้ตัดสินใจ ได้แก่ ปมราก (Root Node) ปมภายใน (Interior Node) และปมใบไม้ (Leaf Node) ซึ่งแต่ละปมของต้นไม้ตัดสินใจ จะเชื่อมต่อกันด้วยกิ่งที่ใช้กำหนดทิศทาง (Directional Branch) สำหรับ Root Node จะเป็นปมแรกที่จะเป็นจุดเริ่มต้น ซึ่งจะไม่มีกิ่งใดชี้เข้าหา ส่วน Interior Node จะเป็นปมที่มีกิ่งชี้เข้าหาและออกจากจุดนี้ และ Leaf Node คือปมที่มีแต่กิ่งที่ชี้เข้าหา ไม่มีกิ่งที่ออกมาจากปมนี้ ทั้งนี้กฎของการจัดประเภทของต้นไม้ตัดสินใจ จะถูกกำหนดไว้ที่แต่ละกิ่ง หรือเส้นทางของต้นไม้ตัดสินใจ สำหรับปมที่ไม่มี Leaf Node จะเรียกว่า ตัวแปร (Attributes) โดยเทคนิคต้นไม้ตัดสินใจนี้จะใช้แนวคิดของ ถ้าหาก-ดังนั้น (IF-THEN) ซึ่งเส้นทางที่เป็น IF จะประกอบด้วยตัวแปรต้นที่อยู่ตามเส้นทางของต้นไม้ตัดสินใจ ซึ่งเป็นส่วนที่ทำให้เกิดเหตุการณ์ต่างๆตามมา ขณะที่เส้นทางที่เป็น THEN จะประกอบด้วยตัวแปรตามที่เป็นผลลัพธ์ที่เกิดจากตัวแปรต้นต่างๆที่อยู่ในเส้นทางที่เป็น IF [27]



รูปที่ 2.4 เทคนิคต้นไม้ตัดสินใจ

สำหรับเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคที่เข้าใจและใช่ง่าย ทำให้เกิดความรวดเร็วในการคำนวณ อีกทั้งยังเหมาะที่จะนำมาใช้กับข้อมูลที่มีขนาดใหญ่และมี Outlier ดังเช่นข้อมูลนี้ โดย Decision Tree ที่นิยมใช้ได้แก่

ID3 ซึ่งเป็นวิธี Decision Tree พื้นฐานที่ใช้ในการจัดประเภทผลลัพธ์ โดยการทดสอบคุณสมบัติ ผ่านการสร้างต้นไม้ตัดสินใจจากระดับกว้างไปที่ระดับเฉพาะเจาะจง (Top-Down) โดยแต่ละหน่วยของต้นไม้ตัดสินใจ จะมีคุณสมบัติที่ใช้คัดกรองให้ได้ค่าเกนความรู้ (Information Gain) ที่มากที่สุด และความไม่เป็นระเบียบของระบบ (Entropy) ที่น้อยที่สุด ซึ่งผลที่ได้จะถูกนำมาแบ่งแยกประเภทของข้อมูลซ้ำๆ จนกระทั่งหน่วยย่อยของต้นไม้ตัดสินใจ (Sub-Tree) เป็นหนึ่งเดียวกัน เหตุผลที่นำวิธีดังกล่าวมาใช้ คือ ผลลัพธ์ของการทำนายเข้าใจง่าย สามารถประมวลผลได้รวดเร็ว ด้วยการสร้างต้นไม้ตัดสินใจแบบสั้นๆ ผ่านการใช้ตัวแปรที่เพียงพอต่อการจัดประเภท นอกจากนี้ยังมีการค้นหาข้อมูลทั้งหมดสำหรับการสร้างต้นไม้ตัดสินใจอีกด้วย อย่างไรก็ตามวิธีดังกล่าวมีข้อเสียคือ อาจทำให้เกิดปัญหา Overfitting หากข้อมูลทดสอบมีจำนวนน้อยเกินไป สามารถทดสอบได้เพียง 1 ตัวแปรต่อครั้งที่ทำการทดสอบเพื่อการตัดสินใจ และไม่สามารถจัดการกับตัวแปรเชิงปริมาณ ข้อมูลผิดปกติ (Outlier) และข้อมูลที่ขาดหาย (Missing Value) หากมีปริมาณมากได้ อีกทั้งยังไม่สามารถลดทอนผลของต้นไม้ตัดสินใจได้ ถึงแม้เป็นการเพิ่มความผิดพลาดสำหรับข้อมูลสอน แต่จะเป็นการลดความผิดพลาดในข้อมูลทดสอบที่มองไม่เห็นได้ อีกทั้งยังสามารถแก้ปัญหาเรื่องตัวแปรเชิงปริมาณ ข้อมูลที่ขาดหาย (Missing Value) รวมถึงข้อมูลผิดปกติ (Outlier) ต่างๆ

C4.5 เป็นวิธีต้นไม้ตัดสินใจที่ถูกพัฒนาขึ้นมา โดยการวิเคราะห์ข้อมูลเชิงลึกก่อน (Depth-First Strategy) ที่สามารถคัดเลือกข้อมูลที่มีอัตราส่วนของเกนความรู้ (Gain Ratio) มากที่สุด รวมถึงยังกำจัดอคติที่เกิดขึ้นจากวิธี ID3 เนื่องจากสามารถลดทอนผลของต้นไม้ตัดสินใจได้ เหตุผลที่นำวิธีดังกล่าวมาใช้ คือ สามารถจัดการได้ทั้งตัวแปรแบบ Continuous และ Discrete สามารถแก้ปัญหาในตัวแปร Continuous ได้โดยการสร้างค่าเกณฑ์ (Threshold) เพื่อแบ่งแยกประเภทได้ รวมถึง

จัดการกับข้อมูลที่หายไป (Missing Value) ได้ดี อีกทั้งยังสามารถนำกึ่งต้นไม้ตัดสินใจที่ไม่จำเป็นออกได้ อย่างไรก็ตามข้อเสียของวิธีดังกล่าวคือ อาจมีการสร้างกิ่งว่างเปล่า จึงจำเป็นที่จะต้องให้ความสำคัญในการสร้างเกณฑ์ นอกจากนี้อาจเกิดปัญหา Over-Fitting หากใช้กับข้อมูลที่มีลักษณะเฉพาะหรือข้อมูลที่มีข้อมูลผิดปกติ (Outlier) จำนวนมาก [28]

(3) เทคนิคการเรียนรู้แบบง่าย (Naïve Bayes) โดยเป็นเทคนิคสาย Bayesian Network ที่ง่ายและนิยมใช้ ภายใต้สมมติฐานที่ว่าทุกตัวแปรสามารถนำไปสู่การจัดประเภทข้อมูลได้ และมีความเกี่ยวเนื่องร่วมกัน ซึ่งสามารถใช้ทำนายความเป็นไปได้ ตามประเภทของข้อมูลที่ระบุได้ อ้างอิงจากสมการด้านล่าง ดังนี้

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

โดย	$P(C X)$	คือ ความเป็นไปได้หลังเกิดเหตุการณ์ของผลลัพธ์ที่ใช้ทำนาย
	$P(C)$	คือ ความเป็นไปได้ก่อนเกิดเหตุการณ์ของผลลัพธ์ที่ใช้ทำนายของประเภท
	$P(X C)$	คือ ความเป็นไปได้หลังเกิดเหตุการณ์ของตัวแปรที่ใช้ทำนาย
	$P(X)$	คือ ความเป็นไปได้ก่อนเกิดเหตุการณ์ของตัวแปรที่ใช้ทำนาย

เทคนิคดังกล่าวมีข้อดีคือ สามารถประมวลผลได้ในเวลาอันสั้น รวมถึงเพิ่มประสิทธิภาพของการจัดประเภทผ่านการนำตัวแปรที่ไม่เกี่ยวข้องออกไป อีกทั้งยังมีประสิทธิภาพที่ดีในการทำนาย อย่างไรก็ตามวิธีดังกล่าวจำเป็นที่จะต้องใช้ข้อมูลจำนวนมากอย่างเพียงพอ และอาจมีค่าความถูกต้อง (Accuracy) ที่น้อยลง เมื่อเทียบกับวิธีการจัดประเภทอื่นๆ ในบางข้อมูล [27]

2.4 เอกสารและงานวิจัยที่เกี่ยวข้องกับความสอดคล้องกันระหว่างข้อมูลที่ใช้กับข้อมูลที่เกิดขึ้นจริงในประเทศไทย

งานวิจัยนี้ใช้ข้อมูลจาก Lending Club ซึ่งเป็นผู้ประกอบการ P2P Lending ในสหรัฐอเมริกา โดยธุรกิจ P2P Lending ในสหรัฐอเมริกา จะอยู่ในการควบคุมดูแลของสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ ซึ่งจะมีความเข้มงวดในการบังคับให้มีการเปิดเผยข้อมูลอย่างเพียงพอ ทำให้อุตสาหกรรมการปล่อยกู้ผ่าน P2P Lending ในสหรัฐอเมริกามีความเป็นระเบียบ โดยมีการจำกัดวงเงินลงทุนไม่เกิน 2,500 ดอลลาร์สหรัฐต่อครั้ง หากนักลงทุนมีรายได้ไม่เกิน 70,000 ดอลลาร์สหรัฐต่อปี ซึ่งสอดคล้องกับการกำกับดูแลผู้ประกอบการ P2P Lending ในประเทศไทย ภายใต้กฎระเบียบที่ออกโดยธนาคารแห่งประเทศไทย ที่มีการจำกัดวงเงินของนักลงทุนรายย่อยให้ลงทุนไม่เกิน 500,000 บาทต่อปีเช่นกัน ขณะที่กลุ่มนักลงทุนรายใหญ่จะไม่

ถูกจำกัดวงเงิน อีกทั้งยังกำหนดให้นักลงทุนผู้ให้กู้ ต้องไม่เป็นผู้ประกอบการที่ดำเนินการแพลตฟอร์มเอง นอกจากนี้ยังกำหนดให้ผู้ประกอบการต้องเป็นบริษัทจดทะเบียนในไทย ที่มีทุนจดทะเบียนขั้นต่ำ 5 ล้านบาทและมีสัดส่วนของผู้ถือหุ้นที่เป็นคนไทยไม่น้อยกว่า 75% รวมถึงการคิดอัตราดอกเบี้ยกู้ยืมที่จะต้องไม่เกิน 15% อีกด้วย จากกฎระเบียบการควบคุมดูแลที่เข้มงวดของทั้งสหรัฐอเมริกาและไทย จึงสามารถใช้ข้อมูลจาก Lending Club เป็นแนวทางให้แก่ผู้ประกอบการในไทย ในการสร้างแบบจำลองพยากรณ์สำหรับทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ในแพลตฟอร์ม P2P Lending ภายใต้ทฤษฎีที่เกี่ยวข้อง จากที่กล่าวมาข้างต้นในหัวข้อ 2.1.2 [29]

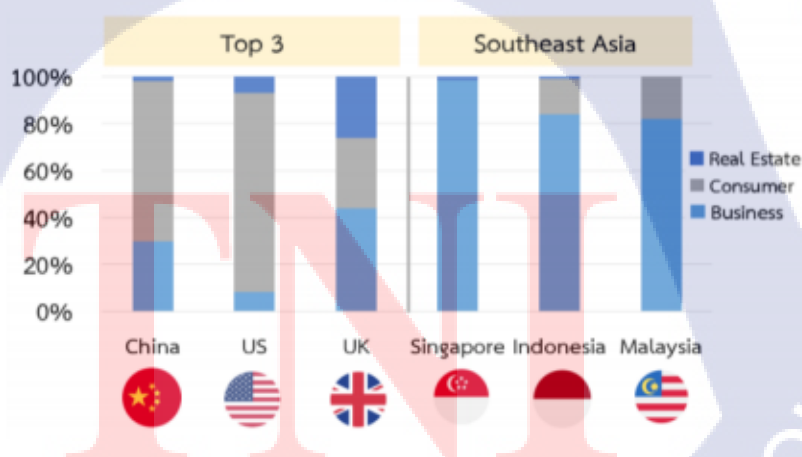
ตารางที่ 2.1 การเปรียบเทียบการควบคุมดูแลแพลตฟอร์ม P2P Lending ในประเทศสหรัฐอเมริกาและไทย

การควบคุมดูแล	สหรัฐอเมริกา	ไทย
ระดับความเข้มงวด	มาก	มาก
ควบคุมดูแลโดย	สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์	ธนาคารแห่งประเทศไทย

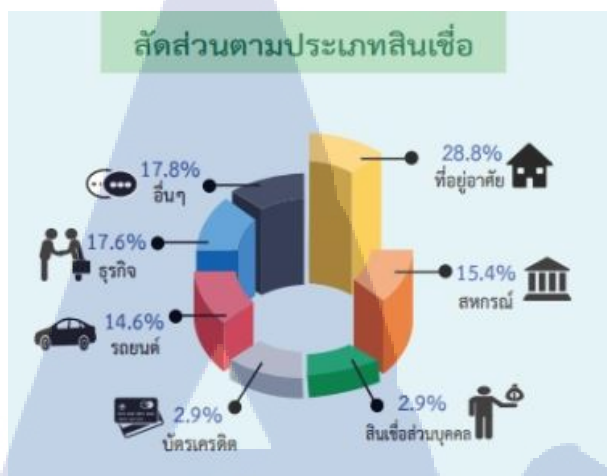
ตารางที่ 2.1 การเปรียบเทียบการควบคุมดูแลแพลตฟอร์ม P2P Lending ในประเทศสหรัฐอเมริกา และไทย (ต่อ)

การควบคุมดูแล	สหรัฐอเมริกา	ไทย
วงเงินในการลงทุนสูงสุด สำหรับนักลงทุนรายย่อย	จำกัดวงเงินลงทุนไม่เกิน 2,500 ดอลลาร์สหรัฐต่อครั้ง หากนักลงทุนมีรายได้ไม่เกิน 70,000 ดอลลาร์สหรัฐต่อปี	จำกัดวงเงินของนักลงทุนราย ย่อยให้ลงทุนไม่เกิน 500,000 บาทต่อปี
วงเงินในการลงทุนสูงสุด สำหรับนักลงทุนรายใหญ่และ สถาบัน	ไม่มี	ไม่มี

นอกจากนี้ ความคล้ายคลึงระหว่างแพลตฟอร์ม P2P Lending ของสหรัฐ และสัดส่วนผู้กู้ในไทยยังได้แก่ สัดส่วนของ P2P Lending ของสหรัฐที่ถูกกระจายอยู่ในส่วนของในกลุ่มประเภทบุคคลธรรมดาเป็นส่วนใหญ่ ขณะที่สัดส่วนของผู้กู้ในไทย ส่วนใหญ่อยู่ในกลุ่มสินเชื่อดังกล่าวเช่นเดียวกัน ตามสถิติจากธนาคารแห่งประเทศไทย [30] ดังรูปด้านล่าง ดังนี้



รูปที่ 2.5 แผนภาพแสดงสัดส่วนของสินเชื่อในแพลตฟอร์ม P2P Lending ในแต่ละประเทศ



รูปที่ 2.6 แผนภาพแสดงสัดส่วนของสินเชื่อในประเทศไทย

2.5 เอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้าง Predictive Model เฉพาะเจาะจงสำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้

เนื่องจากโอกาสในการผิดนัดชำระหนี้ ขึ้นกับความเสี่ยงของผู้กู้ ซึ่งสามารถแบ่งกลุ่มได้จากเกรดของสินเชื่อ (Loan Grade) ที่แบ่งเป็น A ถึง G โดยผู้กู้ที่มีความเสี่ยงต่ำที่สุดจะถูกจัดเป็นสินเชื่อ Grade A ผู้กู้ที่มีความเสี่ยงปานกลางจะถูกจัดเป็นสินเชื่อ Grade B ผู้กู้ที่มีความเสี่ยงสูงจะถูกจัดเป็นสินเชื่อ Grade C และผู้กู้ที่มีความเสี่ยงสูงมากจะถูกจัดเป็นสินเชื่อ Grade D ถึง G ดังนั้นจึงได้มีการสร้าง Predictive Model สำหรับข้อมูลโดยรวม และสำหรับข้อมูลผู้กู้ในแต่ละกลุ่ม Loan Grade เพื่อให้ได้แบบทำนายที่มีความเฉพาะเจาะจงยิ่งขึ้น [31] ทั้งนี้ Loan Grade ที่ใช้จะเป็น Loan Grade มาตรฐานของทั่วโลก ที่จัดกลุ่มจากระดับคะแนนเครดิตฟิโก (Fair, Isaac and Company Score: FICO score) รายงานโดยสำนักงานข้อมูลเครดิตแห่งชาติ (National Credit Bureaus) ของสหรัฐอเมริกา ซึ่งคำนวณจากข้อมูลทางประวัติทางการเงิน 5 ชนิด ได้แก่ ประวัติการชำระเงิน (น้ำหนัก 35%) การหนี้สิน (น้ำหนัก 30%) ประวัติการขออนุมัติสินเชื่อ (น้ำหนัก 15%) ชนิดของสินเชื่อที่เคยกู้ยืม (น้ำหนัก 10%) และการค้นหาล่าสุดของสินเชื่อ (น้ำหนัก 10%) โดย FICO Score ที่มีคะแนนสูงกว่า 750 จะถูกจัดเป็นสินเชื่อ Grade A คะแนนระหว่าง 700-749 จะถูกจัดเป็นสินเชื่อ Grade B คะแนนระหว่าง 650-699 จะถูกจัดเป็นสินเชื่อ Grade C และคะแนนต่ำกว่า 650 จะถูกจัดเป็นสินเชื่อ Grade D ถึง G ดังนั้นในงานวิจัยนี้ จะนำข้อมูลของกลุ่มผู้กู้แต่ละ Loan Grade มาใช้ในการสร้าง Predictive Model เฉพาะเจาะจงสำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้ [32]

แบบจำลองพยากรณ์โดยทั่วไปที่ใช้ในงานวิจัยก่อนหน้านี้ มักสร้างมาจากข้อมูลโดยรวมของผู้กู้ ซึ่งทำให้ความแม่นยำของแบบจำลองแตกต่างกันออกไปในแต่ละกลุ่มผู้กู้ที่มี Loan Grade ต่างกัน

โดยเฉพาะระหว่างกลุ่มผู้กู้ที่มีระดับความเสี่ยงต่ำถึงปานกลาง (Loan Grade A ถึง B) และกลุ่มผู้กู้ที่มีระดับความเสี่ยงสูงถึงสูงมาก (Loan Grade C ถึง G) ดังนั้นในงานวิจัยนี้จึงมีการสร้างแบบจำลองสำหรับผู้กู้แบ่งตามระดับกลุ่มความเสี่ยงของผู้กู้หรือ loan grade เพื่อเพิ่มความเฉพาะเจาะจงสำหรับผู้กู้ในแต่ละกลุ่มความเสี่ยง

นอกจากนี้งานวิจัยยังเพิ่มตัวแปรต้นเชิงระบบทางด้านสภาพเศรษฐกิจ เพื่อรวมตัวแปรจากปัจจัยภายนอกแพลตฟอร์มในการทำนายของแพลตฟอร์ม P2P Lending และทำให้เกิดความครอบคลุมมากขึ้น

อีกทั้งแบบจำลองพยากรณ์แนวโน้มการผิดนัดชำระหนี้ของผู้กู้ในงานวิจัยนี้ ยังได้เพิ่มเทคนิคการจัดการข้อมูล (Data Preparation) ด้วยวิธี SMOTE เพื่อใช้แก้ปัญหาเรื่อง Imbalanced Data และวิธี DBSCAN Clustering เพื่อแก้ปัญหา Outlier ของข้อมูล เพื่อให้แบบจำลองพยากรณ์ด้วยเทคนิคเหมืองข้อมูลมีประสิทธิภาพเพิ่มมากขึ้น

2.6 เอกสารและงานวิจัยที่เกี่ยวข้องกับการออกแบบ User Experience (UX) และ User Interface (UI) เพื่อใช้เป็นแนวทางในการพัฒนาด้านแบบของระบบแพลตฟอร์ม P2P Lending

ทฤษฎีที่เกี่ยวข้องกับการออกแบบ UX และ UI สำหรับแพลตฟอร์ม P2P Lending ไว้ว่าการออกแบบ UX และ UI คือการสร้างประสบการณ์ และความพึงพอใจผ่านทางเว็บไซต์ให้กับผู้ใช้งาน ดังนั้นในการออกแบบจำเป็นต้องมีแนวคิด และมีทิศทางการสื่อสารของสินค้าและบริการที่ชัดเจน โดยควรระมัดระวังข้อขัดแย้งในการออกแบบเดียวกัน สำหรับในแต่ละสถานการณ์ที่แตกต่างกัน ทั้งนี้ขั้นตอนที่ดีในการออกแบบ UX และ UI จะประกอบด้วย 4 ขั้นตอนหลัก คือ

2.6.1 การศึกษาหาความรู้ ซึ่งผลที่ได้รับในขั้นตอนนี้คือ การกำหนดกลุ่มเป้าหมาย และฟังก์ชันการใช้งานของเว็บไซต์ โดยในขั้นตอนนี้จะเป็นการวางแผน (UX Strategy) ที่เป็นการออกแบบบน 4 พื้นฐานหลักคือ กลยุทธ์ทางธุรกิจ (Business Strategy) คุณค่าทางนวัตกรรม (Value Innovation) ตอบโจทย์ของผู้ใช้งานจากการค้นคว้า (Validated User Research) สร้าง UX ที่ก่อให้เกิดประสบการณ์สำหรับผู้ใช้งาน (Killer UX) โดยเริ่มต้นจากการวิเคราะห์เนื้อหาในปัจจุบันเพิ่มเติมและพัฒนาเว็บไซต์ และสร้างแอปพลิเคชันใหม่

2.6.2 การออกแบบโครงสร้างของข้อมูล ซึ่งผลที่ได้รับในขั้นตอนนี้คือ ไดอะแกรมต่างๆ ของงานแต่ละส่วน (Task Flow Diagrams) และการออกแบบเชิงโครงสร้างที่เป็นภาพใหญ่ของเว็บไซต์

2.6.3 การออกแบบรูปภาพและสิ่งที่มองเห็นบนเว็บไซต์ ซึ่งผลที่ได้รับในขั้นตอนนี้คือ ต้นแบบ UX/UI จำลองสำหรับเว็บไซต์ (mockup) รวมถึงการออกแบบกราฟฟิคต่างๆ

2.6.4 การส่งมอบ UX และ UI ซึ่งผลที่ได้รับในขั้นตอนนี้คือ ชิ้นงานเว็บไซต์ และการระบุถึงลักษณะ รวมถึงข้อกำหนดของเว็บไซต์ (Specification)

สำหรับการประเมินผลของการออกแบบ UX/ UI สำหรับแพลตฟอร์ม P2P Lending นั้นสามารถประเมินโดยใช้เทคนิค Human-Computer Interaction (HCI) ที่มีประสิทธิผล โดยใช้แนวทางของการมองเห็นสถานะของระบบ ความสอดคล้องของระบบกับการใช้งานจริง ผู้ใช้งานมีอิสระในการควบคุมการใช้งาน มีความสอดคล้องกันและมีมาตรฐาน มีการป้องกันความผิดพลาดในการใช้งาน มีการสร้างความรับรู้และแจ้งเตือน มากกว่าให้ผู้ใช้งานรับรู้ด้วยตนเอง มีความยืดหยุ่นและมีประสิทธิผลในการใช้งาน มีความสวยงามและใช้การออกแบบที่ไม่มากจนเกินไป ช่วยให้ผู้ใช้งานได้รับรู้และตรวจสอบข้อผิดพลาด รวมถึงแก้ไขให้กลับมาเป็นปกติได้ด้วยตนเอง นอกจากนี้ยังควรมีส่วนของการให้ความช่วยเหลือ ช่องทางการติดต่อเพิ่มเติม หรือคำถามที่พบบ่อย เพื่อช่วยแก้ปัญหาให้กับผู้ใช้งานได้อย่างทันท่วงที [33]

นอกจากนี้ทางผู้วิจัย ยังได้รวบรวมลักษณะ UX และ UI ของแต่ละแพลตฟอร์ม P2P Lending ที่มีชื่อเสียงระดับโลก เพื่อใช้เป็นแนวทางในการพัฒนาด้านแบบของระบบแพลตฟอร์ม P2P Lending ในส่วนของผู้กู้ดังนี้

ตาราง 2.2 การเปรียบเทียบองค์ประกอบด้านการใช้งาน สิ่งที่น่าสนใจ และช่องว่างที่สามารถพัฒนาเพิ่มเติม จากการศึกษาแพลตฟอร์ม P2P Lending ที่มีชื่อเสียงระดับโลก

แพลตฟอร์ม P2P Lending	องค์ประกอบ ด้านการใช้งาน	สิ่งที่น่าสนใจของ แพลตฟอร์ม	ช่องว่างที่สามารถ พัฒนาเพิ่มเติม
Upstart	ส่วนแนะนำเว็บไซต์ สิ่งที่ควรรู้เพื่อขอสินเชื่อ ส่วนแนะนำแพลตฟอร์ม P2P Lending เปรียบเทียบกับการกู้ยืม แบบอดีต ช่องทางการติดต่อ	การเข้าสู่ระบบโดย ตรวจสอบก่อนว่า ผู้ใช้งานเป็นคนหรือไม่ ผ่าน Captcha ก่อนเข้า สู่ระบบ	ไม่มีส่วนของการ ตรวจสอบสถานะ เชื่อมโยงกับ Predictive Model ยังไม่มีวิดีโอที่ Auto- Play สำหรับหน้าแรกที่ใช้
Funding Circle	ส่วนแนะนำเว็บไซต์ สิ่งที่ควรรู้เพื่อขอสินเชื่อ ส่วนแนะนำแพลตฟอร์ม P2P Lending ส่วนของการสมัครเงินกู้ เปรียบเทียบกับการกู้ยืม แบบอดีต ช่องทางการติดต่อ คำถามที่พบบ่อย	การออกแบบที่แสดงให้เห็นข้อมูลแต่ละส่วน อย่างชัดเจน	ใช้ในการสื่อสารกับ ผู้ใช้งานตั้งแต่ครั้งแรก ของการเข้าชมเว็บไซต์ ในส่วนของการสมัคร เงินกู้มีการให้อัพโหลด เอกสารเพื่อยืนยัน ข้อมูล
Prosper Marketplace	ส่วนแนะนำเว็บไซต์ สิ่งที่ควรรู้เพื่อขอสินเชื่อ ส่วนแนะนำแพลตฟอร์ม P2P Lending ส่วนของการสมัครเงินกู้ เปรียบเทียบกับการกู้ยืม แบบอดีต ช่องทางการติดต่อ คำถามที่พบบ่อย	มีการแยกแพลตฟอร์ม สำหรับผู้กู้และผู้ให้กู้ อย่างชัดเจน สามารถตรวจสอบ อัตราดอกเบี้ยกู้ยืมได้	

ตาราง 2.2 การเปรียบเทียบองค์ประกอบด้านการใช้งาน สิ่งที่น่าสนใจ และช่องว่างที่สามารถพัฒนาเพิ่มเติม จากการศึกษาแพลตฟอร์ม P2P Lending ที่มีชื่อเสียงระดับโลก (ต่อ)

แพลตฟอร์ม P2P Lending	องค์ประกอบ ด้านการใช้งาน	สิ่งที่น่าสนใจของ แพลตฟอร์ม	ช่องว่างที่สามารถ พัฒนาเพิ่มเติม
Peerform	ส่วนแนะนำเว็บไซต์ สิ่งที่ควรรู้เพื่อขอสินเชื่อ ส่วนแนะนำแพลตฟอร์ม P2P Lending ส่วนของการสมัครเงินกู้ เปรียบเทียบกับกรู้ยืม แบบอดีต ช่องทางการติดต่อ คำถามที่พบบ่อย	มีการแยกแพลตฟอร์ม สำหรับผู้กู้และผู้ให้กู้ อย่างชัดเจน สามารถตรวจสอบ อัตราดอกเบี้ยกู้ยืมได้ การออกแบบสวยงาม ใช้งานง่าย	ไม่มีส่วนของการ ตรวจสอบสถานะ เชื่อมโยงกับ Predictive Model ยังไม่มีวิดีโอที่ Auto- Play สำหรับหน้าแรกที่ใช้ ในการสื่อสารกับ ผู้ใช้งานตั้งแต่ครั้งแรก ของการเข้าชมเว็บไซต์
Lending Club	ส่วนแนะนำเว็บไซต์ สิ่งที่ควรรู้เพื่อขอสินเชื่อ ส่วนแนะนำแพลตฟอร์ม P2P Lending ส่วนของการสมัครเงินกู้ เปรียบเทียบการกู้ยืม ช่องทางการติดต่อ คำถามที่พบบ่อย	มีการแยกแพลตฟอร์ม สำหรับผู้กู้และผู้ให้กู้ อย่างชัดเจน แยกประเภทของ สินเชื่อที่ให้ผู้กู้ยืม	ในส่วนของการสมัคร เงินกู้มีการให้อัพโหลด เอกสารเพื่อยืนยัน ข้อมูล

2.7 สรุปเอกสารและงานวิจัยที่ใช้

2.7.1 ปัจจัยด้านข้อมูลประวัติทางการเงิน (Hard Information)

ตารางที่ 2.3 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านข้อมูลประวัติทางการเงิน

ปัจจัยที่เกี่ยวข้อง	งานวิจัย	สรุปเนื้อหา
Hard Information	[11]	ข้อมูลด้านฐานะทางการเงิน (Hard Information) เช่น ระดับคะแนนของสินเชื่อ (Credit Grade) อัตราส่วนหนี้สินต่อรายได้ จำนวนบัตรเครดิตของธนาคาร จำนวนสินเชื่อที่ขออนุมัติ ประวัติการขอสินเชื่อ ความเป็นเจ้าของบ้าน เป็นต้น
รายได้ต่อเดือน	[1]	ผู้ที่มีรายได้ต่อเดือนสูง มีแนวโน้มที่จะได้รับเงินทุนจากนักลงทุนมากกว่า รวมถึงยังมีโอกาสในการผิदनชำระหนี้้น้อยกว่า นักลงทุนใช้ Soft Information ของผู้กู้ในการประกอบการตัดสินใจลงทุน P2P Lending เนื่องจาก Hard Information เพียงอย่างเดียว อาจไม่เพียงพอต่อการตัดสินใจ ผู้กู้ที่มีอายุมากกว่า มีการศึกษาที่สูงกว่า สถานภาพแต่งงานแล้ว มีอาชีพ และมีการทำงานที่ยาวนานกว่า มีแนวโน้มที่จะได้รับเงินทุนจากนักลงทุนมากกว่า เนื่องจากนักลงทุนพิจารณาว่าเป็นกลุ่มผู้กู้ที่มีความเสี่ยงต่อการผิदनชำระหนี้้น้อยกว่า
สถานะการมีบ้านเป็นของตนเอง	[1]	ผู้ที่มีบ้านเป็นของตนเอง จะมีแนวโน้มที่ผู้กู้อย่างนั้น จะได้รับการอนุมัติเงินกู้มากกว่า เพราะนักลงทุนพิจารณาว่าจะมีโอกาสในการผิदनชำระหนี้้น้อยกว่า

ตารางที่ 2.3 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านข้อมูลประวัติทางการเงิน (ต่อ)

ปัจจัยที่เกี่ยวข้อง	งานวิจัย	สรุปเนื้อหา
ระดับเครดิตของผู้กู้ (Credit Rating) และระดับความเสี่ยงของผู้กู้ (Loan Grade)	[7] และ [31]	ผู้กู้ที่มี Credit Rating สูง จะมีความเสี่ยงทางสินเชื่อต่ำ มีโอกาสได้รับเงินทุนมากกว่า และมีโอกาสในการผิดนัดชำระหนี้ต่ำกว่า โอกาสในการผิดนัดชำระหนี้ ขึ้นกับความเสี่ยงของผู้กู้ ซึ่งสามารถแบ่งกลุ่มได้จากเกรดของสินเชื่อ (Loan Grade) ที่แบ่งเป็น A ถึง G โดยผู้กู้ที่มีความเสี่ยงต่ำที่สุดจะถูกจัดเป็นสินเชื่อ Grade A และผู้กู้ที่มีความเสี่ยงสูงมากจะถูกจัดเป็นสินเชื่อ Grade D ถึง G
อัตรารายจ่ายต่อรายได้ และอัตราหนี้สินต่อรายได้	[7]	นักลงทุนถือว่า ผู้ที่มีอัตรารายจ่ายต่อรายได้ และอัตราหนี้สินต่อรายได้สูง มีโอกาสได้รับเงินทุนน้อยกว่า และมีโอกาสในการผิดนัดชำระหนี้มากกว่า
วัตถุประสงค์ในการกู้ยืม	[31]	ผู้ที่มีวัตถุประสงค์ในการกู้ยืมเพื่อทำบัตรเครดิตและเพื่อใช้ในการทำธุรกิจขนาดเล็ก จะมีแนวโน้มที่จะเกิดการผิดนัดชำระหนี้มากกว่า ทำให้มีโอกาสดังกล่าวได้รับเงินทุนจากนักลงทุนน้อยกว่า

2.7.2 ปัจจัยด้านข้อมูลส่วนตัว (Soft Information)

ตารางที่ 2.4 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านข้อมูลส่วนตัว

ปัจจัยที่เกี่ยวข้อง	งานวิจัย	สรุปเนื้อหา
Soft Information	[11]	ข้อมูลส่วนตัว (Soft Information) เช่น ข้อมูลด้านประชากร ข้อมูลด้านเครือข่ายทางสังคม รูปภาพ และคำที่ใช้ในแบบคำขออนุมัติเงินกู้ เป็นต้น
อายุ การศึกษา สถานภาพ อาชีพการ ทำงาน	[1]	นักลงทุนใช้ Soft Information ของผู้กู้ในการประกอบการตัดสินใจลงทุน P2P Lending เนื่องจาก Hard Information เพียงอย่างเดียว อาจไม่เพียงพอต่อการตัดสินใจ ผู้กู้ที่มีอายุมากกว่า มีการศึกษาที่สูงกว่า สถานภาพแต่งงานแล้ว มีอาชีพ และมีการทำงานที่ยาวนานกว่า มีแนวโน้มที่จะได้รับเงินทุนจากนักลงทุนมากกว่า เนื่องจากนักลงทุนพิจารณาว่าเป็นกลุ่มผู้กู้ที่มีความเสี่ยงต่อการผิดนัดชำระหนี้ต่ำกว่า
เพศของผู้กู้	[7]	ข้อมูลด้านเพศของผู้กู้ จะไม่ส่งผลต่อโอกาสในการได้รับอนุมัติเงินกู้ และแนวโน้มการผิดนัดชำระหนี้ ในกรณีที่ผู้กู้มี Credit Rating ที่เท่ากัน

2.7.3 ปัจจัยด้านเศรษฐกิจเชิงมหภาพ

ตารางที่ 2.5 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านเศรษฐกิจเชิงมหภาพ

ปัจจัยที่เกี่ยวข้อง	งานวิจัย	สรุปเนื้อหา
ปัจจัยเชิงมหภาค	[12]	การผิติดนัฒซำระหนึ้สามารถเกิดจากปัจจัยทางด้ำนมหภาค เช่น GDP อัตราการว่างงาน การเปลี่ยนแปลงของตลาดหุ้น หรือตลาดอสังหาริมทรัพย์ เป็นต้น
อัตราการว่างงาน	[12]	เมื่อเศรษฐกิจเชิงมหภาคชะลอตัว จะทำให้ผู้ว่างงานมากขึ้น ทำให้มีผู้กู้ที่ต้องการความช่วยเหลือด้านการเงินเพิ่มมากขึ้น รวมถึงธนาคารมีความรัดกุมในการปล่อยสินเชื่อมากขึ้น ส่งผลให้มีการขอกู้ยืมผ่านแพลตฟอร์ม P2P Lending เพิ่มขึ้น และมีแนวโน้มของการผิติดนัฒซำระหนึ้เพิ่มขึ้นขึ้นในที่สุด
ผลตอบแทนของตลาดหลักทรัพย์	[12]	หากตลาดหุ้นมีผลตอบแทนที่ลดลง จะทำให้รายได้ของประชาชนลดลง ส่งผลให้มีความต้องการในการลงทุนในกลุ่มระดับความเสี่ยงสูง ที่ให้ผลตอบแทนดอกเบี้ยสูงเพิ่มขึ้น ส่งผลให้แนวโน้มการผิติดนัฒซำระหนึ้ของผู้กู้เพิ่มสูงขึ้น อย่างไรก็ตาม เนื่องจากแพลตฟอร์ม P2P Lending ถือว่าเป็นช่องทางเลือกสำหรับการลงทุน นอกเหนือจากตลาดหลักทรัพย์ ดังนั้นการที่ตลาดหุ้นให้ผลตอบแทนที่ลดลง จะทำให้นักลงทุนหันมาลงทุนในแพลตฟอร์ม P2P Lending เพิ่มขึ้นได้เช่นเดียวกัน

ตารางที่ 2.5 สรุปเอกสารที่เกี่ยวข้องกับปัจจัยด้านเศรษฐกิจเชิงมหภาพ (ต่อ)

ปัจจัยที่เกี่ยวข้อง	งานวิจัย	สรุปเนื้อหา
ผลตอบแทนของตลาดอสังหาริมทรัพย์	[12]	ราคาของที่อยู่อาศัยลดลง จะทำให้หลักประกันของผู้กู้ในธนาคารมีมูลค่าลดลง ซึ่งเรียกว่าปรากฏการณ์ collateral effect ส่งผลให้ผู้กู้ได้รับการอนุมัติเงินกู้จากธนาคารลดลง การกู้ยืมผ่านแพลตฟอร์ม P2P Lending เพิ่มขึ้น และทำให้แนวโน้มของการผิดนัดชำระหนี้เพิ่มสูงขึ้น
มาตรการในการควบคุมความเสี่ยงในการกู้ยืม	[12]	หากรัฐบาลออกมาตรการในการควบคุมความเสี่ยงในการกู้ยืมผ่านแพลตฟอร์ม P2P Lending จะเป็นการลดความเสี่ยงของการผิดนัดชำระหนี้ในการกู้ยืมผ่าน P2P Lending ส่งผลให้มีการเปิดเผยข้อมูลมีความโปร่งใสเพิ่มขึ้นและลดการเผยแพร่ข้อมูลปลอมบนเว็บไซต์ของ P2P Lending

2.7.4 สรุปเอกสารที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูลและการจัดเตรียมข้อมูล

ตารางที่ 2.6 สรุปเอกสารที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูลและการจัดเตรียมข้อมูล

เทคนิค	งานวิจัย	สรุปเนื้อหา	เหตุผลที่นำมาใช้
Logistic Regression	[26]	ใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวแปร และตัวแปรตามแบบจัดกลุ่ม รวมถึงการพยากรณ์ความน่าจะเป็นที่จะเกิดขึ้นของเหตุการณ์ (Odd)	ใช้ได้กับข้อมูลที่มีขนาดใหญ่ และมีตัวแปรต้นจำนวนมาก เช่นเดียวกับข้อมูลที่ใช้ในงานวิจัยนี้ อีกทั้งยังมีประสิทธิภาพในการจัดการข้อมูลที่ไม่ได้มีการแจกแจงแบบปกติ (Non-Normal Distribution) ซึ่งทำให้วิธีสมการถดถอย (Regression) แบบทั่วไปใช้ไม่ได้ผล
Decision Tree	[27] และ [28]	เทคนิคที่ใช้การแบ่งขั้นตอนการตัดสินใจจนกว่าจะได้ชุดข้อมูลที่สอดคล้องกับผลลัพธ์ที่ต้องการจัดประเภทมากที่สุด โดยใช้แนวคิดของ ถ้าหาก-ดังนั้น (IF-THEN) ซึ่งเส้นทางที่เป็น IF จะประกอบด้วยตัวแปรต้นที่อยู่ตามเส้นทางของต้นไม้ตัดสินใจ ซึ่งเป็นส่วนที่ทำให้เกิดเหตุการณ์ต่างๆตามมา ขณะที่เส้นทางที่เป็น THEN จะประกอบด้วยตัวแปรตามที่เป็นผลลัพธ์ที่เกิดจากตัวแปรต้นต่างๆที่อยู่ในเส้นทางที่เป็น IF	เป็นเทคนิคที่เข้าใจและใช้งานง่าย ทำให้เกิดความรวดเร็วในการคำนวณ อีกทั้งยังเหมาะที่จะนำมาใช้กับข้อมูลที่มีขนาดใหญ่และมี Outlier ดังเช่นข้อมูลนี้

ตารางที่ 2.6 สรุปเอกสารที่เกี่ยวข้องกับเทคนิคการทำเหมืองข้อมูลและการจัดเตรียมข้อมูล (ต่อ)

เทคนิค	งานวิจัย	สรุปเนื้อหา	เหตุผลที่นำมาใช้
Naïve Bayes	[27]	เทคนิคสาย Bayesian Network ที่ ภายใต้สมมติฐานที่ว่าทุกตัวแปรสามารถนำไปสู่การจับประเภท และมีความเกี่ยวเนื่องร่วมกัน เพื่อใช้ทำนายความเป็นไปได้	สามารถประมวลผลได้ในเวลาอันสั้น รวมถึงเพิ่มประสิทธิภาพของการจัดประเภทผ่านการนำตัวแปรที่ไม่เกี่ยวข้องออกไป อีกทั้งยังมีประสิทธิภาพที่ดีในการทำนาย
SMOTE	[16] และ [17]	การสุ่มข้อมูลในกลุ่มรองเพื่อสร้างข้อมูลใหม่ของกลุ่มรองให้มีจำนวนเพิ่มมากขึ้น จนใกล้เคียงหรือเท่ากับจำนวนกลุ่มหลักที่เดิมมีมากกว่า ทำให้เทคนิคนี้สามารถแก้ปัญหาที่เกิดขึ้น และเพิ่มความแม่นยำให้กับเทคนิค Classification แบบต่างๆได้	ใช้ในการจัดกลุ่มข้อมูลที่มีความไม่สมดุล (Imbalanced Data)
DBSCAN Clustering	[18]	เทคนิคแบบ Clustering ที่ใช้ในการแบ่งกลุ่มและแยก Outlier ที่มีในข้อมูลโดยใช้ตัวแปรหลัก 2 ตัวแปร คือ ขนาดรัศมีของข้อมูล และจำนวนเพื่อนบ้านขั้นต่ำภายในรัศมี Eps โดยช่วยเพิ่มความแม่นยำให้กับ Predictive Model หลังจาก Outlier ถูกกำจัดออกไป	ใช้ในการแบ่งกลุ่มเพื่อคัดแยกและกำจัด Outlier ออกไป

บทที่ 3

วิธีการดำเนินงาน

3.1 กรอบแนวคิดการวิจัย

การวิจัยเรื่องแบบจำลองการพยากรณ์ เพื่อช่วยทำนายสถานะการผิन्दชำระหนี้ของผู้กู้ในแต่ละกลุ่มความเสี่ยงของผู้กู้ ผู้วิจัยได้ดำเนินการตามขั้นตอน ดังนี้

3.1.1 การศึกษาข้อมูลรูปแบบของ P2P Lending Model ตัวแปรที่เกี่ยวข้อง และเทคนิคการทำเหมืองข้อมูลที่เหมาะสมสำหรับการออกแบบแบบจำลองการพยากรณ์

3.1.2 การรวบรวมข้อมูลจากกลุ่มประชากร

3.1.3 กระบวนการทำความเข้าใจข้อมูล (Data Understanding)

3.1.4 การจัดเตรียมข้อมูล (Data Preparation)

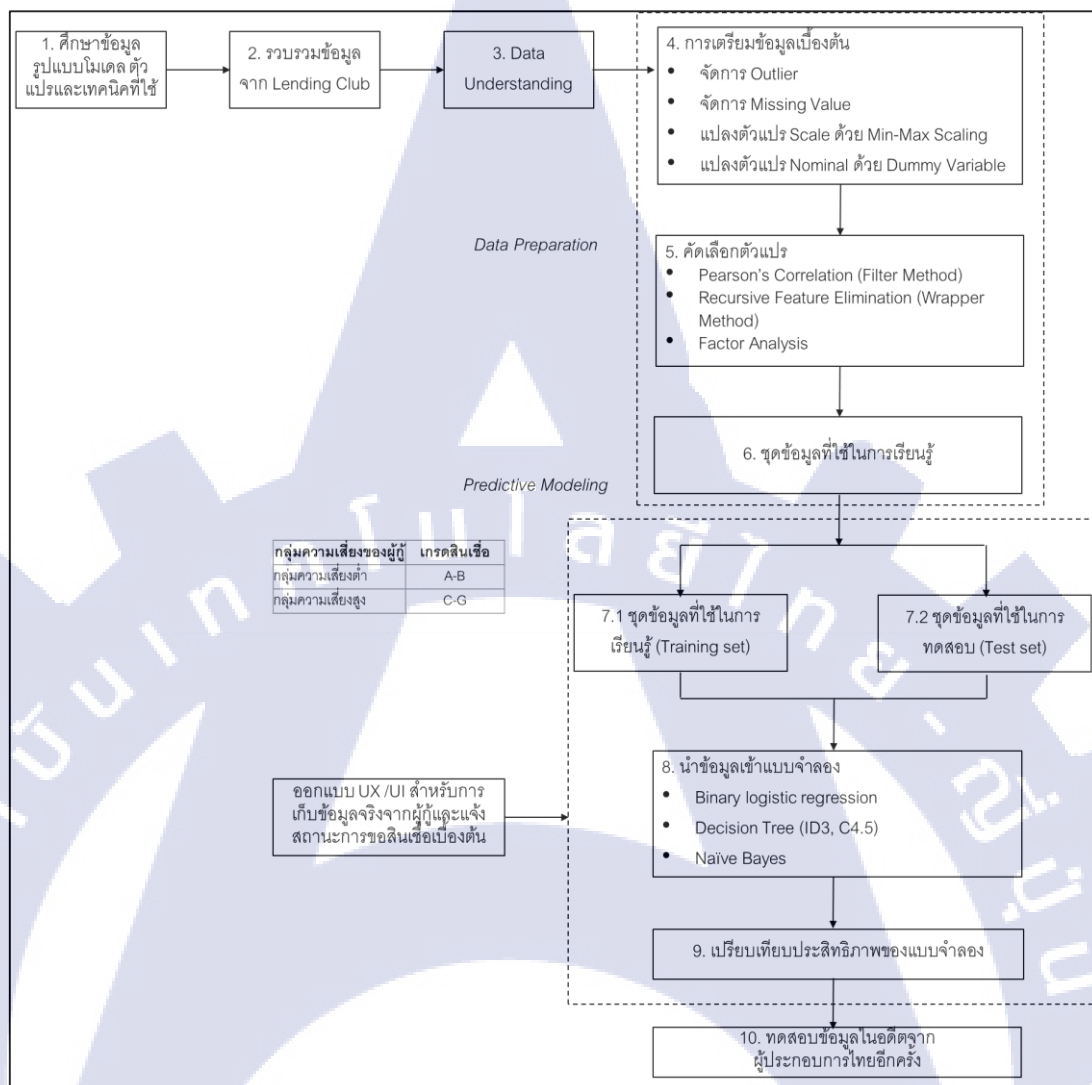
3.1.5 การสร้างแบบจำลองการพยากรณ์

3.1.6 การเปรียบเทียบประสิทธิภาพและการสรุปผลการคัดเลือกแบบจำลอง

3.1.7 การนำข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ

3.1.8 การออกแบบ UX/UI สำหรับการเก็บข้อมูลจริงจากผู้กู้และแจ้งสถานะการขอสินเชื่อ

เบื้องต้น



รูปที่ 3.1 กรอบแนวคิดวิจัยสำหรับแบบจำลองพยากรณ์แนวโน้มการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending

จากกรอบแนวคิดงานวิจัยข้างต้น ได้แสดงถึงกระบวนการในการดำเนินงานวิจัย ซึ่งประกอบไปด้วยขั้นตอนต่างๆ ดังนี้

3.2 การศึกษาข้อมูล สำหรับการออกแบบแบบจำลองการพยากรณ์

3.2.1 การศึกษาลักษณะของ P2P Lending model ในแต่ละรูปแบบ

ทำการศึกษาเพื่อให้ทราบถึงประเภท วิธีการให้ผลตอบแทน และวัตถุประสงค์หลักของ P2P Lending Model ในแต่ละรูปแบบ โดยงานวิจัยนี้จะสร้างแบบจำลองการพยากรณ์สำหรับ P2P

Lending ที่เป็น Lending-Based Model ซึ่งมีผลตอบแทนเป็นดอกเบี้ยเงินกู้ จากการให้กู้แก่ผู้กู้ และมีวัตถุประสงค์สำหรับการลงทุนทั่วไปที่ไม่ได้เฉพาะเจาะจง เนื่องจาก P2P Lending Model ดังกล่าว เป็นรูปแบบที่นิยมใช้โดยทั่วไป เพื่อทำให้งานวิจัยเป็นประโยชน์ต่อนักลงทุน ผู้ประกอบการ และผู้กู้ในวงกว้างที่สุด

3.2.2 การศึกษาตัวแปรที่เกี่ยวข้อง

(1) ตัวแปรต้นด้านผู้กู้และสินเชื่อ โดยปัจจัยด้านผู้กู้และสินเชื่อ ที่จะส่งผลต่อการตัดสินใจว่าผู้กู้แต่ละรายมีแนวโน้มในการผิดนัดชำระหนี้หรือไม่ ได้แก่ ข้อมูลด้านฐานทางการเงิน (Hard Information) และ ข้อมูลส่วนตัว (Soft Information) รวมถึงยังมีการใช้ตัวแปรด้านเศรษฐกิจมหภาค (Macroeconomic Indicator) โดยงานวิจัยนี้จะใช้ตัวแปรต้นทั้งหมด 35 ตัวแปร ดังนี้

ตารางที่ 3.1 สรุปตัวแปรต้นที่เกี่ยวข้องที่ใช้ในงานวิจัย

ตัวแปร	ชนิดของข้อมูล	ความหมาย	ข้อมูลที่น่าไปประมวลผล
Grade	String (Ordinal)	ระดับคะแนนเครดิตของผู้กู้	A ถึง G
Loan_amount	Integer (Scale)	จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ	1,000 ถึง 35,000 ดอลลาร์สหรัฐ
Term	Integer (Interval)	ระยะเวลาของการชำระสินเชื่อ	36 เดือนและ 60 เดือน
Int_rate	Float (Scale)	อัตราดอกเบี้ยของสินเชื่อ	5.32% ถึง 28.99%
Installment	Float (Scale)	งวดชำระต่อเดือนของสินเชื่อ	19.87 ถึง 1,445.46 ดอลลาร์สหรัฐ
Home_ownership	String (Nominal)	สถานะการเป็นเจ้าของบ้าน	มีบ้าน ไม่มีบ้าน ผ่อนเช่า อื่นๆ
Annual_inc	Integer (Scale)	รายได้ต่อปีของผู้กู้	3,800 ถึง 9,000,000 ดอลลาร์สหรัฐ
Issue_d	Year	เวลาที่สินเชื่อได้รับขอการอนุมัติ	ปี 2009 ถึง 2015

ตารางที่ 3.1 สรุปตัวแปรต้นที่เกี่ยวข้องที่ใช้ในงานวิจัย (ต่อ)

ตัวแปร	ชนิดของข้อมูล	ความหมาย	ข้อมูลที่น่าไปประมวลผล
Purpose	String (Nominal)	วัตถุประสงค์ที่ขอสินเชื่อ	เพื่อผ่อนชำระหนี้ทั่วไป เพื่อนำไปชำระหนี้บัตรเครดิต เพื่อซื้อรถยนต์ เพื่อปรับปรุงบ้าน เพื่อการศึกษา เพื่อทำธุรกิจ เป็นต้น
Dti	Float (Scale)	อัตราหนี้สินต่อรายได้ของผู้กู้	0 ถึง 136.97
Delinq_2yrs	Integer (Scale)	จำนวนครั้งที่มีการผิดนัดชำระหนี้เกิน 30 วันจากประวัติของผู้กู้ใน 2 ปีที่ผ่านมา	0 ถึง 29 ครั้ง
Inq_last_6mths	Integer (Scale)	จำนวนการส่งแบบสอบถามสินเชื่อที่ไม่ใช่รถยนต์และบ้านใน 6 เดือนที่ผ่านมา	0 ถึง 10 ครั้ง
Open_acc	Integer (Scale)	จำนวนการเปิดวงเงินสินเชื่อจากประวัติด้านสินเชื่อของผู้กู้	0 ถึง 84 วงเงิน
Pub_rec	Integer (Scale)	จำนวนครั้งที่มีการบันทึกประวัติเสียด้านสินเชื่อ	0 ถึง 86 ครั้ง
Revol_bal	Integer (Scale)	จำนวนสินเชื่อในวงเงินสินเชื่อหมุนเวียน	0 ถึง 2,568,995 ดอลลาร์สหรัฐ
Revol_util	Float (Scale)	อัตราการใช้วงเงินสินเชื่อหมุนเวียน	0 ถึง 892.3 ดอลลาร์สหรัฐ
Total_acc	Integer (Scale)	จำนวนวงเงินล่าสุดจากประวัติด้านสินเชื่อของผู้กู้	2 ถึง 162 วงเงิน

ตารางที่ 3.1 สรุปตัวแปรต้นที่เกี่ยวข้องที่ใช้ในงานวิจัย (ต่อ)

ตัวแปร	ชนิดของข้อมูล	ความหมาย	ข้อมูลที่น่าไปประมวลผล
Initial_list_status	String (Nominal)	สถานการณ์ลงทะเบียนสินเชื่อ	เงินกู้จากนักลงทุนรายย่อยที่มีเงินลงทุนขั้นต่ำ 25 ดอลลาร์สหรัฐต่อสินเชื่อ (Fractional Loan) เงินกู้จากนักลงทุนสถาบัน (Whole Loan)
Funded_amnt_inv	Float (Scale)	สัดส่วนของเงินลงทุนที่ได้รับการลงทุนจริงในอดีต	0% ถึง 100%
Funded_amnt	Float (Scale)	เงินลงทุนที่ได้รับการลงทุนจริงในอดีต	0 ถึง 149,372.86 ดอลลาร์สหรัฐ
Total_pymnt	Float (Scale)	ยอดงวดชำระที่ได้รับล่าสุดจากจำนวนทั้งหมดที่ได้รับอนุมัติเงินกู้	0 ถึง 57,777.58 ดอลลาร์สหรัฐ
Total_pymnt_inv	Float (Scale)	ยอดงวดชำระที่ได้รับล่าสุดจากจำนวนทั้งหมดที่ได้รับการลงทุน	0 ถึง 57,777.58 ดอลลาร์สหรัฐ
Total_rec_prncp	Float (Scale)	จำนวนเงินต้นที่ได้รับล่าสุด	0 ถึง 35,000.03 ดอลลาร์สหรัฐ
Total_rec_int	Float (Scale)	จำนวนดอกเบี้ยที่ได้รับล่าสุด	0 ถึง 24,205.62 ดอลลาร์สหรัฐ
Recoveries	Float (Scale)	จำนวนเงินที่เรียกเก็บได้จากสินเชื่อที่มีแนวโน้มผิดนัดชำระหนี้	0 ถึง 33,520.27 ดอลลาร์สหรัฐ
Last_pymnt_amt	Float (Scale)	จำนวนงวดชำระสุดท้ายที่ได้รับ	0 ถึง 36,234.44 ดอลลาร์สหรัฐ

ตารางที่ 3.1 สรุปตัวแปรต้นที่เกี่ยวข้องที่ใช้ในงานวิจัย (ต่อ)

ตัวแปร	ชนิดของข้อมูล	ความหมาย	ข้อมูลที่น่าไปประมวลผล
Application_type	String (Nominal)	ชนิดการสมัครสินเชื่อ	กู้คนเดียว กู้ร่วม
Emp_title	String (Nominal)	อาชีพของผู้กู้	ครู ผู้จัดการ นางพยาบาล ผู้ตรวจการ เจ้าของธุรกิจ พนักงานขาย คนขับรถ วิศวกร เป็นต้น
Emp_length	Integer (Interval)	ระยะเวลาทำงานของผู้กู้	น้อยกว่า 1 ปี 2 ปี 3 ปี 4 ปี 5 ปี 6 ปี 7 ปี 8 ปี 9 ปี 10 ปี มากกว่า 10 ปี
Addr_state	String (Nominal)	ที่อยู่อาศัยของผู้กู้	รัฐต่างๆในสหรัฐอเมริกา
GDP	Float (Scale)	อัตราการเติบโตของ ผลิตภัณฑ์มวลรวม ภายในประเทศ	US GDP Growth Rate ตามช่วงเวลาที่ยกค่าขอ อนุมัติเงินกู้ของผู้กู้แต่ละราย
Unemp	Float (Scale)	อัตราการว่างงาน	US Unemployment Rate ตามช่วงเวลาที่ยกค่าขอ อนุมัติเงินกู้ของผู้กู้แต่ละราย
Stock	Float (Scale)	ผลตอบแทนของดัชนี หลักทรัพย์	US S&P500 Index Return ตามช่วงเวลาที่ยกค่าขอ อนุมัติเงินกู้ของผู้กู้แต่ละราย
Inflation	Float (Scale)	อัตราเงินเฟ้อ	US Inflation Rate ต่ าม ช่วงเวลาที่ยกค่าขออนุมัติ เงินกู้ของผู้กู้แต่ละราย
Housing_price	Float (Scale)	ดัชนีราคาอสังหาริมทรัพย์	US Housing Price (2005=100) ตามช่วงเวลาที่ยก ค่าขออนุมัติเงินกู้ของผู้กู้ แต่ละราย

ตารางที่ 3.2 สรุปตัวแปรตามที่เกี่ยวข้องที่ใช้ในงานวิจัย

ตัวแปร	ชนิดของข้อมูล	ความหมาย	ข้อมูลที่ปรับสำหรับนำไป ประมวลผล
Loan_status	String (Nominal)	สถานะที่ผู้กู้จะผิดนัดชำระ หนี้หรือไม่	มีแนวโน้มผิดนัดชำระหนี้ ไม่มี แนวโน้มผิดนัดชำระหนี้

3.2.3 การศึกษาเทคนิคการทำเหมืองข้อมูลที่เหมาะสม

การสร้างแบบจำลองพยากรณ์ด้วยวิธีการทำเหมืองข้อมูลควบคู่กับเทคนิคการจัดการข้อมูลด้วย DBSCAN Clustering เพื่อใช้ในการกำจัด Outlier และเทคนิค SMOTE ในการแก้ปัญหา Imbalanced Data ในการทำนายข้อมูลด้านเครดิตสินเชื่อนานาชาติเล็ก ซึ่งเป็นสินเชื่อบริการที่ถูกปล่อยผ่านระบบ P2P Lending

3.3 การรวบรวมข้อมูลจากกลุ่มประชากร

งานวิจัยนี้ใช้กลุ่มตัวอย่างจากข้อมูลสินเชื่อจริงจาก Lending Club ซึ่งเป็นผู้ประกอบการ P2P Lending รายใหญ่ในสหรัฐอเมริกา โดยใช้ข้อมูลจากผู้กู้จำนวน 503,421 ราย ในปี 2009 ถึง 2015 ทั้งนี้ข้อมูลผู้กู้จะถูกแบ่งกลุ่มของระดับความเสี่ยงตามระดับคะแนนของสินเชื่อ (Loan Grade) คือ กลุ่มความเสี่ยงต่ำถึงปานกลาง (A-B) และกลุ่มความเสี่ยงสูงถึงสูงมาก (C-G)

นอกจากนี้ยังมีการนำข้อมูลทางด้านเศรษฐกิจมหภาค ตามช่วงเวลาของผู้กู้ขอสินเชื่อ หนี้ครัวเรือน สำหรับผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) อัตราการว่างงาน อัตราเงินเฟ้อ ดัชนีราคาอสังหาริมทรัพย์ของสหรัฐจาก Organization for Economic Co-operation and Development (OECD Database) และผลตอบแทนจากดัชนีหลักทรัพย์ S&P 500 จาก Morning Star

3.4 กระบวนการทำความเข้าใจข้อมูล (Data Understanding)

การทำความเข้าใจข้อมูลด้วยการดูสัดส่วนของข้อมูลในแต่ละตัวแปร รวมถึงสัดส่วนของข้อมูลที่หายไป (missing value) ผ่าน Python

3.5 การจัดเตรียมข้อมูล (Data Preparation)

3.5.1 เทคนิคการสังเคราะห์ข้อมูลใหม่ (Synthetic Minority Oversampling Technique: SMOTE) ซึ่งเป็นเทคนิคที่ใช้ในการจัดกลุ่มข้อมูลที่มีความไม่สมดุล (Imbalanced

Data) โดยจะใช้การสุ่มข้อมูลในกลุ่มรองเพื่อสร้างข้อมูลใหม่ของกลุ่มรองให้มีจำนวนเพิ่มมากขึ้น จนใกล้เคียงหรือเท่ากับจำนวนกลุ่มหลักที่เดิมมีมากกว่า

3.5.2 เทคนิคการกำจัดข้อมูลผิดปกติ (Outlier) โดยเทคนิคการจัดกลุ่มข้อมูลแบบ DBSCAN (Density-Based Spatial Clustering of Applications with Noise: DBSCAN Clustering)

3.5.3 การลดตัวแปร โดยจะทำการเปรียบเทียบประสิทธิภาพของแบบจำลองผ่านการคัดเลือกตัวแปร 3 วิธี ได้แก่ Pearson's Correlation, Recursive Feature Elimination (RFE) และ Factor Analysis

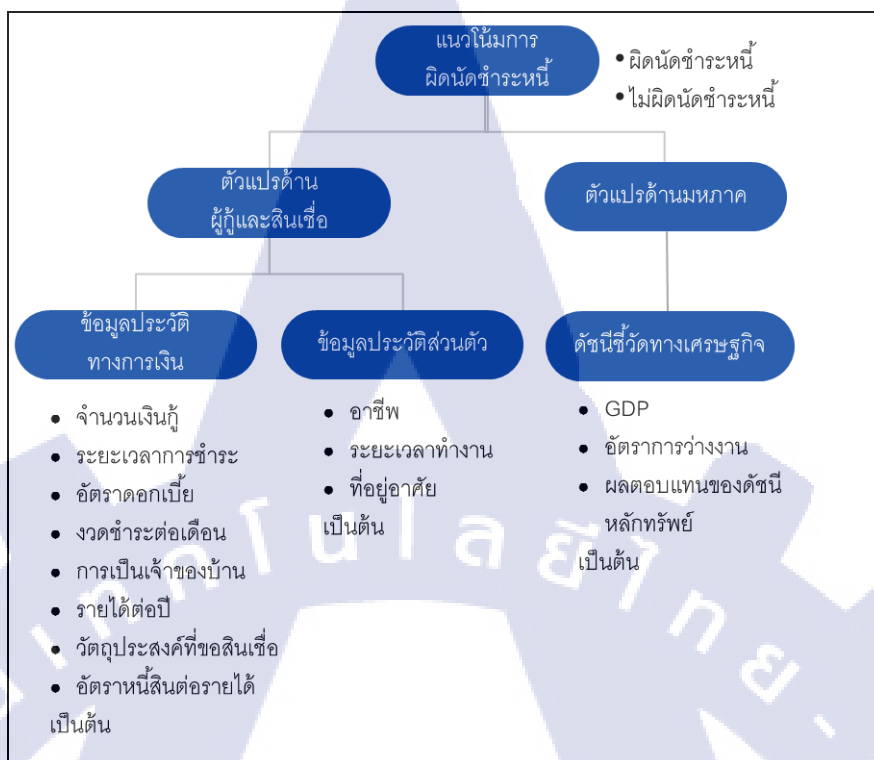
3.5.4 การแก้ปัญหาข้อมูลที่ขาดหาย (Missing Value) โดยการนำแถวที่ไม่มีข้อมูลที่เป็นนัยสำคัญ และมีน้ำหนักของข้อมูลน้อยออกไปจากแบบจำลอง และใช้ค่าเฉลี่ยของข้อมูลกลุ่มที่มีค่าใกล้เคียงกับกลุ่มของข้อมูล เพื่อเติมข้อมูล Missing Value ให้กับข้อมูลในแถวที่เหลือ

3.5.5 การเตรียมข้อมูลเชิงปริมาณ (Scale Variable) ด้วยวิธี Min-Max Scaling โดยการทำให้ข้อมูลเรียงจากค่า 0 ไป 1 เพื่อลดขนาดของข้อมูล และทำให้ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ลดลง และทนต่อผลกระทบจากข้อมูลที่ผิดปกติ (Outlier)

3.5.6 การเตรียมข้อมูลเชิงคุณภาพ (Nominal และ Ordinal Variable) ด้วยการสร้างตัวแปรจำลอง (Dummy หรือ Indicator Variable) เพื่อใช้แทนค่าของตัวแปรแบบจัดกลุ่ม (Nominal Variable) หรือตัวแปรที่สามารถไล่ระดับได้ (Ordinal Variable) ซึ่งวิธีดังกล่าวเป็นการทำให้ตัวแปรเชิงคุณภาพ ถูกแปลงเป็นตัวเลข และสามารถนำมาใช้วิเคราะห์ในสมการถดถอย (Regression) ได้

3.6 การสร้างแบบจำลองการพยากรณ์

ในงานวิจัยนี้จะทำการสร้างแบบจำลองการพยากรณ์ โดยใช้โปรแกรม Python ในการสร้างแบบจำลองด้วยเทคนิคทางเหมืองข้อมูล ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes เพื่อใช้ทำนายแนวโน้มการผิมนัดชำระหนี้ จากข้อมูลโดยรวมแต่ละกลุ่มความเสี่ยงของผู้กู้ โดยมีแผนผังของการสร้างแบบจำลองพยากรณ์จากตัวแปรต่างๆ ดังนี้



กลุ่มความเสี่ยงของผู้กู้	กลุ่มเกรดสินเชื่อ
กลุ่มความเสี่ยงต่ำถึงปานกลาง	A-B
กลุ่มความเสี่ยงสูงถึงสูงมาก	C-G

รูปที่ 3.2 แผนผังของการสร้างแบบจำลองพยากรณ์จากตัวแปรที่ใช้ในงานวิจัย

3.7 การเปรียบเทียบประสิทธิภาพและการสรุปผลการคัดเลือกแบบจำลอง

3.7.1 การเปรียบเทียบประสิทธิภาพกับแบบจำลองที่สร้างจากเทคนิคพื้นฐาน (Benchmark)

ตัวอย่างเช่น การเปรียบเทียบแบบจำลองที่สร้างจากเทคนิคทางเหมืองข้อมูล ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes ซึ่งเป็นเทคนิคพื้นฐานที่ถูกนำไปใช้ในงานวิจัยก่อนหน้านี้ โดยตัวชี้วัดประสิทธิภาพของแบบจำลองจาก Confusion Matrix ได้แก่ ความถูกต้อง (Accuracy) ความแม่นยำ (Precision) และค่าระลึก (Recall) โดย Confusion Matrix คือการแบ่ง Output Class ที่ทำนายได้ออกเป็น 2 กลุ่มย่อยของแต่ละกลุ่มหลัก คือ True กับ False จึงทำให้เกิดผลในการ Classification ได้ทั้งหมด 4 กลุ่ม คือ True Positive (TP) หรือค่าจริงเป็นบวก และทำนายถูกต้องเป็นบวก True Negative (TN) หรือค่าจริงเป็นลบ และ

ทำนายถูกต้องเป็นลบ False Positive (FP) หรือค่าจริงเป็นลบ แต่ทำนายเป็นบวก False Negative (FN) หรือค่าจริงเป็นบวก แต่ทำนายเป็นลบ [34] โดยจะวัดจากค่า

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

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

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

$$\text{F-1 Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

นอกจากนี้ยังสามารถทำการวัดผลได้จาก Receiver Operating Characteristic Curve (ROC Curve) หรือการสร้างกราฟความสัมพันธ์ระหว่าง True Positive Rate และ False Positive Rate หากค่าของพื้นที่ใต้โค้ง ROC Curve ยิ่งเข้าใกล้ 1 มากเท่าใด จะแสดงถึงประสิทธิภาพที่สูงขึ้น

3.8 การนำข้อมูลจริงในอดีตจากผู้ประกอบการในประเทศไทย ด้านการให้กู้ยืมสินเชื่อมาเป็นข้อมูลทดสอบ

หลังจากที่ได้แบบจำลองพยากรณ์ที่มีประสิทธิภาพแล้ว จะนำข้อมูลผู้กู้ส่วนหนึ่ง จากผู้ประกอบการในประเทศไทย ด้านการให้กู้ยืมสินเชื่อในอดีตช่วงเดือน พ.ค. ถึง พ.ย. 2562 โดยไม่ระบุชื่อของผู้กู้ (Anonymous ID) มาใช้เป็นข้อมูลทดสอบ เพื่อประเมินประสิทธิภาพการใช้งานของแบบจำลองที่สร้างกับข้อมูลของผู้กู้ในประเทศไทย โดยกลุ่มผู้กู้ที่จัดตามกลุ่มความเสี่ยงออกเป็นแต่ละ Loan Grade จะได้รับการประเมิน และถูกจัดกลุ่มด้วยมาตรฐานเดียวกับที่ใช้ในข้อมูลจาก Lending Club โดย Loan Grade ถูกใช้เป็นมาตรฐานทั่วโลก ในการวัดระดับความเสี่ยง ตามระดับคะแนนเครดิต FICO score (Fair, Isaac and Company score) ซึ่งรายงานโดยสำนักงานข้อมูลเครดิตแห่งชาติ (National Credit Bureaus) ของสหรัฐอเมริกา ทั้งนี้ Loan Grade คำนวณจากข้อมูลทางประวัติทางการเงิน 5 ชนิด ได้แก่ ประวัติการชำระเงิน (น้ำหนัก 35%) ภาระหนี้สิน (น้ำหนัก 30%) ประวัติการขออนุมัติสินเชื่อ (น้ำหนัก 15%) ชนิดของสินเชื่อที่เคยกู้ยืม (น้ำหนัก 10%) และการขอสินเชื่อใหม่ (น้ำหนัก 10%) โดย FICO Score ที่มีคะแนนสูงกว่า 750 จะถูกจัดเป็นสินเชื่อ Grade A คะแนนระหว่าง 700-749 จะถูกจัดเป็นสินเชื่อ Grade B คะแนนระหว่าง 650-699 จะถูกจัดเป็นสินเชื่อ Grade C และคะแนนต่ำกว่า 650 จะถูกจัดเป็นสินเชื่อ Grade D ถึง G ดังนั้นในงานวิจัยนี้จะ

แบ่งข้อมูลของผู้กู้สำหรับใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ออกเป็น 2 กลุ่มหลักตามระดับความเสี่ยงของผู้กู้ คือ กลุ่มความเสี่ยงต่ำถึงปานกลาง (สินเชื่อ Grade A ถึง B) และกลุ่มความเสี่ยงสูงถึงสูงมาก (สินเชื่อ Grade C ถึง G) [32] โดย myFICO ได้ประเมิน FICO score ด้วยคำถามดังต่อไปนี้ สำหรับการประเมินคะแนนและเปรียบเทียบเป็นเกรดตามหลักการ FICO score

- (1) จำนวนบัตรเครดิตที่ผู้กู้ถืออยู่
- (2) ระยะเวลาที่ถือบัตรเครดิตใบแรก
- (3) จำนวนสินเชื่อหรือสินเชื่อบัตรเครดิตที่ผู้กู้ได้สมัครในปีก่อนหน้า
- (4) ระยะเวลาจนถึงปัจจุบันที่ผู้กู้ได้ขอสินเชื่อหรือสินเชื่อบัตรเครดิตล่าสุด
- (5) วงเงินสินเชื่อหรือสินเชื่อบัตรเครดิตของผู้กู้ในปัจจุบัน
- (6) วงเงินของสินเชื่อทั้งหมดที่นอกเหนือจากสินเชื่อจำนอง
- (7) ระยะเวลาล่าสุดที่ผู้กู้ผิดนัดชำระสินเชื่อหรือสินเชื่อบัตรเครดิต
- (8) จำนวนสินเชื่อหรือสินเชื่อบัตรเครดิตที่เกินกำหนดการชำระหนี้
- (9) สัดส่วน % ของวงเงินจำกัดต่อวงเงินทั้งหมดสำหรับสินเชื่อบัตรเครดิต
- (10) ผู้กู้มีประวัติการล้มละลาย หรือมีรายชื่ออยู่ในบัญชีผู้ผิดนัดชำระหนี้ [35]

3.9 การออกแบบ UX/UI สำหรับการเก็บข้อมูลจริงจากผู้กู้และแจ้งสถานะการขอสินเชื่อเบื้องต้น

หลังจากการนำข้อมูลจริงในอดีตจากผู้ประกอบการในประเทศไทย ด้านการให้กู้ยืมสินเชื่อในอดีตช่วงเดือน พฤษภาคม ถึง พฤศจิกายน พ.ศ. 2560 ไปทดสอบในแบบจำลอง เพื่อยืนยันประสิทธิภาพของการพยากรณ์แนวโน้มการผิดนัดชำระหนี้ ว่าแบบจำลองสามารถนำมาปรับใช้กับข้อมูลของประกอบการในไทยได้แล้ว จะมีการออกแบบต้นแบบ (Prototype) ของระบบแพลตฟอร์มโดยการออกแบบ User Experience (UX) และ User Interface (UI) สำหรับการเก็บข้อมูลของผู้กู้โดยใช้ตัวแปรที่ทำการคัดเลือกมาจากแบบจำลอง เป็นคำถามหลักที่ผู้กู้จำเป็นต้องกรอกข้อมูล เพื่อใช้ในการขอสินเชื่อ รวมถึงใช้คำถามมาตรฐานที่ใช้สำหรับ FICO Score ในการคัดกรองว่าผู้กู้นั้นมีระดับคะแนนเครดิตอยู่ที่เท่าใด เพื่อคัดแยกว่าจะนำข้อมูลของผู้กู้อย่างดังกล่าว เข้าไปที่แบบจำลองสำหรับผู้ที่มีความเสี่ยงต่ำ (เกรดของเครดิตสินเชื่อบัตรระดับ A-B) หรือความเสี่ยงสูง (เกรดของเครดิตสินเชื่อบัตรระดับ C-G) นอกจากนี้ยังมีข้อมูลสำหรับผู้กู้และผู้ให้กู้ในการสอบถาม จากคำถามที่พบบ่อยในแพลตฟอร์มที่ได้รวบรวมมา ดังต่อไปนี้

ตารางที่ 3.3 คำถามที่พบบ่อยจากแพลตฟอร์ม P2P Lending

แพลตฟอร์ม P2P Lending	ประเภทคำถามสำหรับผู้กู้
Upstart	วิธีการสมัคร จำนวนเงินที่กู้ยืมได้ วิธีการชำระเงินกู้ อัตราดอกเบี้ยสำหรับเงินกู้ยืม
Funding Circle	วิธีการสมัคร ค่าใช้จ่ายและค่าธรรมเนียมเพิ่มเติม เกณฑ์ในการกู้ยืม วิธีการชำระเงินกู้
Prosper Marketplace	วิธีการสมัคร เกณฑ์ในการกู้ยืม ระยะเวลาและอัตราดอกเบี้ย
CircleBack Lending	วิธีการสมัคร เกณฑ์ในการกู้ยืม วิธีการชำระเงินกู้ ค่าใช้จ่ายเพิ่มเติม ข้อปฏิบัติหากเกิดการผิดนัดชำระหนี้
Peerform	วิธีการสมัคร วิธีการคัดเลือกผู้กู้ ขั้นตอนการอนุมัติเงินกู้
Lending Club	วิธีการสมัคร เกณฑ์ในการกู้ยืม วิธีการชำระเงินกู้ ระยะเวลาและอัตราดอกเบี้ย

นอกจากการเก็บรวบรวมข้อมูลจากผู้กู้แล้ว งานวิจัยนี้จะออกแบบ UX/ UI ให้สามารถแสดงสถานะจากการทำนายสถานะการขออนุมัติสินเชื่อของผู้กู้เบื้องต้น รวมถึงการตอบคำถามที่พบบ่อย เพื่ออำนวยความสะดวกให้ผู้กู้ในการสืบค้นข้อมูลที่เป็น

นอกจากนี้ยังมีวิธีการตรวจสอบสถานะของสินเชื่อด้วยวิธีการสัมภาษณ์นอกแพลตฟอร์ม (Offline) นอกเหนือจากการทำนายผ่านแบบจำลองพยากรณ์ทางออนไลน์ (Online) เพื่อช่วยให้ผู้ให้กู้และผู้กู้มีความเชื่อมั่นในแพลตฟอร์มเพิ่มมากขึ้น รวมถึงมีจำนวนของผู้ให้กู้และผู้กู้ที่เข้าร่วมมากขึ้น ส่งผลให้แพลตฟอร์มมีโอกาสในการประสบความสำเร็จมากขึ้น [1] สำหรับขั้นตอนการดำเนินการของแพลตฟอร์ม P2P Lending จะประเมินผู้กู้ จากข้อมูลทางการเงินและข้อมูลส่วนตัวของผู้กู้ จากนั้นผู้ให้กู้จะมีการกำหนดจำนวนที่จะให้กู้ รวมถึงอัตราดอกเบี้ยที่ต้องการได้รับ และเงื่อนไขต่างๆ เพื่อทำการคัดเลือกผู้กู้ ตามเงื่อนไขที่ผู้ให้กู้ต้องการ ดังนั้นในการประยุกต์ใช้ในงานวิจัยนี้ จะเป็นการใช้แบบจำลองพยากรณ์ทำนายแนวโน้มสถานะของผู้กู้แบบเบื้องต้น ก่อนที่จะทำการสัมภาษณ์นอกแพลตฟอร์ม เพื่อตรวจสอบเอกสารและประวัติการขอสินเชื่อของผู้กู้อย่างละเอียด ก่อนที่ตัดสินใจจะทำการอนุมัติเงินกู้ให้ผู้กู้หรือไม่ [36]

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยเรื่องแบบจำลองการพยากรณ์เพื่อช่วยทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ในแต่ละกลุ่มความเสี่ยงของผู้กู้ โดยเปรียบเทียบเทคนิคการคัดเลือกตัวแปรและเหมืองข้อมูลแต่ละชนิด เพื่อคัดเลือกตัวแปรและแบบจำลองที่มีประสิทธิภาพสำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้สามารถแบ่งผลการศึกษาวิจัยและการวิเคราะห์ข้อมูลได้ ดังนี้

4.1 การเปรียบเทียบประสิทธิภาพของแบบจำลอง จากวิธีคัดเลือกตัวแปรและเทคนิคเหมืองข้อมูลแบบต่างๆ สำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

4.2 การประเมินค่าความสำคัญของตัวแปรที่ใช้ ในแบบจำลองที่มีประสิทธิภาพมากที่สุดในการทำนายสถานะการผิดนัดชำระหนี้ สำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

4.3 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ สำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

4.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง จากวิธีคัดเลือกตัวแปรและเทคนิคเหมืองข้อมูลแบบต่างๆ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

4.5 การประเมินค่าความสำคัญของตัวแปรที่ใช้ ในแบบจำลองที่มีประสิทธิภาพมากที่สุดในการทำนายสถานะการผิดนัดชำระหนี้ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

4.6 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

4.7 การออกแบบ User Experience (UX) และ User Interface (UI) เพื่อใช้สำหรับการเก็บข้อมูลของผู้กู้ การแจ้งสถานะการขออนุมัติเงินกู้ การรวบรวมคำถามที่พบบ่อยสำหรับผู้กู้ รวมถึงข้อมูลการติดต่อของแพลตฟอร์ม P2P Lending

4.1 การเปรียบเทียบประสิทธิภาพของแบบจำลอง จากวิธีคัดเลือกตัวแปรและเทคนิคเหมืองข้อมูลแบบต่างๆ สำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

ผู้ทำวิจัยได้ใช้โปรแกรม Python ในการวิเคราะห์และเปรียบเทียบแบบจำลองจากการคัดเลือกตัวแปรด้วยวิธีที่แตกต่างกัน ได้แก่ ได้แก่ Pearson's Correlation Recursive Feature

Elimination (RFE) และ Factor Analysis ร่วมกับเทคนิคการทำเหมืองข้อมูลที่แตกต่างกัน ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes รวมถึงการเปรียบเทียบระหว่างการสร้างแบบจำลองที่ใช้ และไม่ใช้การจัดการข้อมูลด้วยวิธี SMOTE เพื่อสร้างข้อมูลใหม่ของกลุ่มรองให้มีจำนวนเพิ่มมากขึ้น จึงทำให้แบบจำลองที่สร้างเพื่อเปรียบเทียบ สำหรับแต่ละกลุ่มความเสี่ยงของผู้กู้ สามารถจัดเป็นประเภทต่างๆได้ ดังนี้

4.1.1 แบบจำลองเต็มรูปแบบที่ไม่ได้คัดเลือกตัวแปร และไม่ได้จัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.2 แบบจำลองเต็มรูปแบบที่ไม่ได้คัดเลือกตัวแปร แต่ใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.3 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Pearson's Correlation แต่ไม่ได้จัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression, Decision Tree ID3, Decision Tree C4.5 และ Naïve Bayes

4.1.4 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Pearson's Correlation และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.5 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Recursive Feature Elimination (RFE) แต่ไม่ได้จัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.6 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Recursive Feature Elimination (RFE) และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.7 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis แต่ไม่ได้จัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

4.1.8 แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิคเหมืองข้อมูลต่างๆ ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes

โดย Feature Selection แบบ I II และ III คือ Pearson's Correlation Recursive Feature Elimination (RFE) และ Factor Analysis ตามลำดับ

ตารางที่ 4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้มีความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) ด้วยเกณฑ์ Accuracy Score

Model Type	Decision Tree (ID3)	Decision Tree (C4.5)	Logistic Regression	Naïve Bayes
Full Model	0.8465	0.8878	0.8836	0.8822
Full Model with SMOTE	0.8008	0.8846	0.8806	0.8798
Feature Selection I	0.8670	0.8840	0.8810	0.8792
Feature Selection I + SMOTE	0.7554	0.8840	0.8806	0.8806
Feature Selection II	0.8506	0.8831	0.8800	0.7682
Feature Selection II + SMOTE	0.7516	0.8827	0.8806	0.8806
Feature Selection III	0.8778	0.8826	0.8798	0.8767
Feature Selection III + SMOTE	0.6792	0.8843	0.8806	0.5861

ตารางที่ 4.2 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้มีความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) ด้วยเกณฑ์ F-1 Score

Model Type	Decision Tree (ID3)	Decision Tree (C4.5)	Logistic Regression	Naïve Bayes
Full Model	0.9136	0.9403	0.9382	0.9374
Full Model with SMOTE	0.8815	0.9385	0.9365	0.9360
Feature Selection I	0.9270	0.9382	0.9368	0.9357
Feature Selection I + SMOTE	0.8486	0.9380	0.9365	0.9365
Feature Selection II	0.9167	0.9377	0.9362	0.8640
Feature Selection II + SMOTE	0.8470	0.9375	0.9365	0.9365
Feature Selection III	0.9346	0.9374	0.9360	0.9342
Feature Selection III + SMOTE	0.7904	0.9383	0.9365	0.7114

จากตารางที่ 4.1 และ 4.2 จะเห็นว่าแบบจำลองที่มีประสิทธิภาพสูงที่สุดที่ใกล้เคียงกับแบบจำลองเต็มรูปแบบ (Full Model) วัดจากค่า Accuracy Rate และ F-1 Score สำหรับกลุ่มผู้มีความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) คือแบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) โดยมีค่า Accuracy Rate และ F-1 Score อยู่ที่ 0.8843 และ 0.9383 ตามลำดับ ซึ่งมีประสิทธิภาพใกล้เคียงกับแบบจำลองเต็มรูปแบบที่มีค่า Accuracy Rate และ F-1 Score อยู่ที่ 0.8878 และ 0.9403 ตามลำดับ สะท้อนให้เห็นว่าแบบจำลองที่คัดเลือกมา มีประสิทธิภาพทั้งในด้านการทำนายแนวโน้มสถานะการผิดนัดชำระหนี้ของผู้กู้ รวมถึงประสิทธิภาพในด้านความเร็วในการประมวลผล เมื่อเทียบกับแบบจำลองเต็มรูปแบบ นอกจากนี้ ยังพบว่าการใช้เทคนิคการจัดการข้อมูลแบบ SMOTE ร่วมกับวิธีการคัดเลือกตัวแปรแบบ Factor Analysis ในแบบจำลองที่ใช้เทคนิค Decision Tree (C4.5) ยังทำให้แบบจำลองมีประสิทธิภาพเพิ่มมากขึ้น อย่างไรก็ตาม ผลการวิจัยพบว่าการใช้เทคนิคการจัดการข้อมูลแบบ SMOTE ร่วมกับวิธีการคัดเลือกตัวแปรแบบ Factor Analysis ในแบบจำลองที่ใช้เทคนิค Decision Tree (ID3) และ Naïve Bayes มีประสิทธิภาพลดลง โดยมีค่า Accuracy Score ลดลงจาก 0.8778 เป็น 0.6792 และลดลงจาก 0.8767 เป็น 0.5861 ตามลำดับเมื่อเทียบกับก่อนการใช้เทคนิค SMOTE นอกจากนี้ยังมีค่า F-1 Score ลดลงจาก 0.9346 เป็น 0.7904 และลดลงจาก 0.9342 เป็น

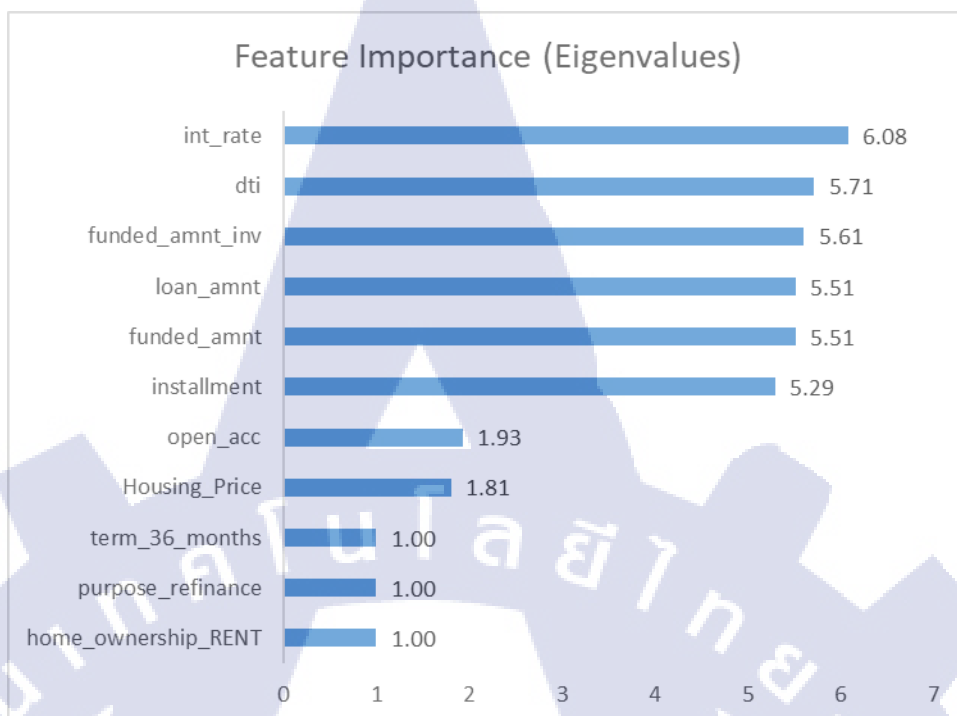
0.7114 ตามลำดับ ซึ่งสอดคล้องกับผลการทดลองที่ค้นพบว่าถึงแม้วิธีการจัดการข้อมูลแบบ SMOTE จะเป็นที่ยอมรับสำหรับการจัดการข้อมูลที่ไม่สมดุล เนื่องจากเป็นวิธีที่ง่าย แต่อาจทำให้ค่า Accuracy Score ลดลงได้ เนื่องจากไม่สามารถทำให้ข้อมูลกระจายตัวได้อย่างเท่าเทียมได้ โดยเฉพาะข้อมูลที่เป็นข้อมูลที่ได้รับทางออนไลน์ เช่นเดียวกับข้อมูลที่ใช้ในงานวิจัยนี้ รวมถึงได้มีการแนะนำให้ใช้วิธี SMOTE ร่วมกับเทคนิคอื่นๆ เช่น เทคนิค Fuzzy-Rough Selection เทคนิค Kernel Trick เป็นต้น [37]

4.2 การประเมินค่าความสำคัญของตัวแปรที่ใช้ ในแบบจำลองที่มีประสิทธิภาพมากที่สุดในการทำนายสถานะการผิคนัดชำระหนี้ สำหรับกลุ่มความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

จากการศึกษาวิจัยข้างต้น ได้ทำการคัดเลือกแบบจำลอง คือ แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) ดังนั้นการคัดเลือกตัวแปรที่มีความสำคัญได้ทำการจัดลำดับด้วยค่าเส้นทแยงมุมของเมทริกซ์ของค่าไอเกน (Eigenvalue) ซึ่งเป็นค่าความแปรปรวนของตัวแปรเดิมทั้งหมดในแบบจำลองที่ถูกอธิบายด้วยแต่ละตัวแปร คำนวณจากผลรวมกำลังสองของค่า Factor Loading ของแต่ละตัวแปร โดยค่า Factor Loading คือ ค่าความเกี่ยวเนื่องกันระหว่างตัวแปรกับองค์ประกอบที่ใช้ในแบบจำลอง โดยวัดค่าจากตาราง Component Matrix สำหรับตัวแปรใดที่มีค่าน้ำหนักมาก ก็จะถูกคัดเลือกเข้ามาใช้ในแบบจำลอง ทั้งนี้ eigenvalue เป็นค่าที่นิยมใช้สำหรับการบ่งบอกถึงความสำคัญของแต่ละตัวแปรที่ใช้ในการคัดเลือกตัวแปรด้วยวิธี Factor Analysis โดยตัวแปรที่มีความสำคัญและนำมาใช้ควรมีค่ามากกว่า 1 ขึ้นไป ซึ่งแสดงถึงความสามารถในการอธิบายค่าความแปรปรวนของกลุ่มตัวแปรได้มากกว่า 1 ตัวแปร [38] ซึ่งค่าความสำคัญของตัวแปรที่คัดเลือกเป็นดังนี้

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY



รูปที่ 4.1 กราฟแสดงค่าความสำคัญของตัวแปรที่คัดเลือกสำหรับแบบจำลอง

จากรูปที่ 4.1 ข้างต้น แสดงถึงตัวแปรที่มีค่าความสำคัญที่ใช้ในแบบจำลองสำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) ซึ่งวัดจากค่า Eigenvalue ที่มากกว่า 1 ได้แก่ Int_rate (อัตราดอกเบี้ย) Dti (อัตราหนี้สินต่อรายได้ของผู้กู้) Funded_amnt_inv (สัดส่วนของเงินที่เคยได้รับ) Loan_amnt (จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ) Funded_amnt (เงินที่เคยได้รับ) Installment (งวดชำระต่อเดือนของสินเชื่อ) Open_acc (จำนวนการเปิดวงเงินสินเชื่อจากประวัติด้านสินเชื่อของผู้กู้) Housing_Price (ดัชนีราคาอสังหาริมทรัพย์) Term_36_months (ระยะเวลาของการชำระสินเชื่อ 36 เดือน) Purpose_refinance (วัตถุประสงค์ที่ขอสินเชื่อ สำหรับการ refinance) และ Home_ownership_RENT (สถานะการเป็นเจ้าของบ้านแบบเช่า)

4.3 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ สำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B)

จากการทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือกสำหรับกลุ่มผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทย พบว่าแบบจำลองที่คัดเลือกมีประสิทธิภาพอยู่ในเกณฑ์ยอมรับได้ วัดจาก Accuracy Score ที่อยู่ระดับ 0.8541

4.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง จากวิธีคัดเลือกตัวแปรและเทคนิคเหมืองข้อมูลแบบต่างๆ สำหรับกลุ่มผู้ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

ตารางที่ 4.3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) ด้วยเกณฑ์ Accuracy Score

Model Type	Decision Tree (ID3)	Decision Tree (C4.5)	Logistic Regression	Naïve Bayes
Full Model	0.7824	0.7865	0.7239	0.5609
Full Model with SMOTE	0.7689	0.7867	0.7188	0.5134
Feature Selection I	0.7776	0.7300	0.7227	0.6190
Feature Selection I + SMOTE	0.7675	0.7810	0.7190	0.5033
Feature Selection II	0.7543	0.7248	0.7244	0.5589
Feature Selection II + SMOTE	0.6998	0.7699	0.7190	0.5017
Feature Selection III	0.7588	0.7332	0.7297	0.7225
Feature Selection III + SMOTE	0.6905	0.7778	0.7188	0.5569

ตารางที่ 4.4 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองสำหรับกลุ่มผู้ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) ด้วยเกณฑ์ F-1 Score

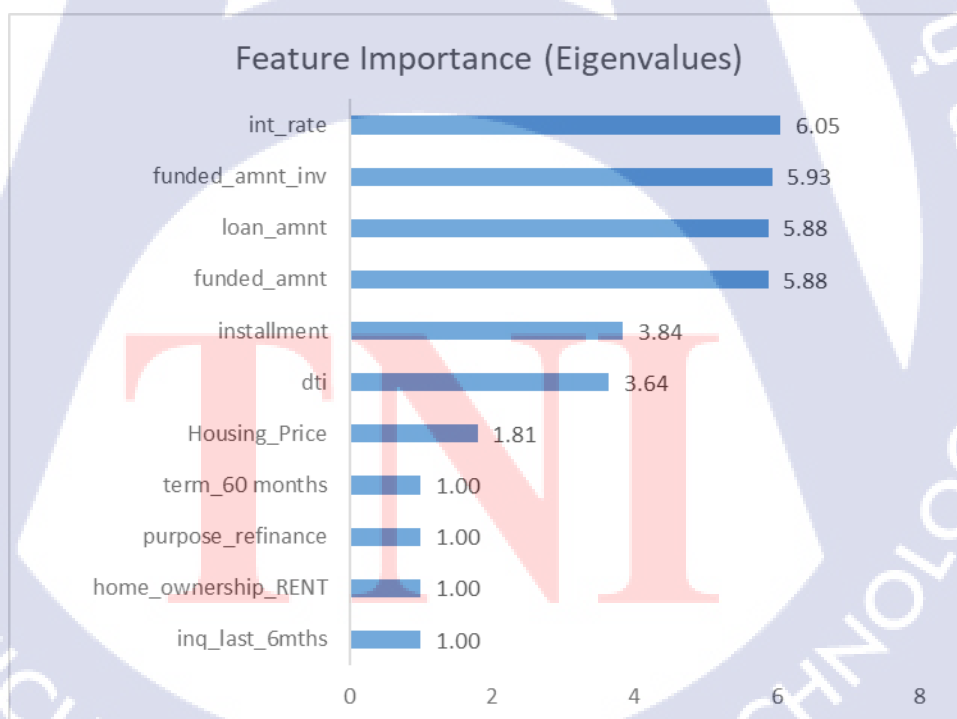
Model Type	Decision Tree (ID3)	Decision Tree (C4.5)	Logistic Regression	Naïve Bayes
Full Model	0.8503	0.8641	0.8399	0.6290
Full Model with SMOTE	0.8345	0.8693	0.8364	0.5556
Feature Selection I	0.8614	0.8424	0.8390	0.7008
Feature Selection I + SMOTE	0.8329	0.8616	0.8365	0.5403
Feature Selection II	0.8336	0.8403	0.8402	0.6239
Feature Selection II + SMOTE	0.7788	0.8553	0.8365	0.5360
Feature Selection III	0.8456	0.8450	0.8437	0.8148
Feature Selection III + SMOTE	0.7734	0.8621	0.8364	0.6300

จากตารางที่ 4.3 และ 4.4 จะเห็นว่าแบบจำลองที่มีประสิทธิภาพสูงที่สุดที่ใกล้เคียงกับแบบจำลองเต็มรูปแบบ (Full Model) วัดจากค่า Accuracy Rate และ F-1 Score สำหรับกลุ่มผู้ที่มีความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) คือแบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีการจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) เช่นเดียวกันโดยมีค่า F-1 Score อยู่ที่ 0.8621 และมี Accuracy Score ที่อยู่ในเกณฑ์สูงและยอมรับได้ที่ 0.7778 ซึ่งมีประสิทธิภาพใกล้เคียงกับแบบจำลองเต็มรูปแบบที่มีค่า Accuracy Rate และ F-1 Score อยู่ที่ 0.7865 และ 0.8641 ตามลำดับ สะท้อนให้เห็นว่าแบบจำลองที่คัดเลือกมา มีประสิทธิภาพทั้งในด้านการทำนายแนวโน้มสถานะการผิดนัดชำระหนี้ของผู้กู้ รวมถึงประสิทธิภาพในด้านความเร็วในการประมวลผล เมื่อเทียบกับแบบจำลองเต็มรูปแบบ นอกจากนี้ ยังพบว่าการใช้เทคนิคการจัดการข้อมูลแบบ SMOTE ร่วมกับวิธีการคัดเลือกตัวแปรแบบ Factor Analysis ในแบบจำลองที่ใช้เทคนิค Decision Tree (C4.5) ยังทำให้แบบจำลองมีประสิทธิภาพเพิ่มมากขึ้น อย่างไรก็ตาม ผลการวิจัยพบว่าการใช้เทคนิคการจัดการข้อมูลแบบ SMOTE ร่วมกับวิธีการคัดเลือกตัวแปรแบบ Factor

Analysis ในแบบจำลองที่ใช้เทคนิค Decision Tree (ID3) Logistic Regression และ Naïve Bayes มีประสิทธิภาพลดลง โดยมีค่า Accuracy Score ลดลงจาก 0.7588 0.7297 และ 0.7225 เป็น 0.6905 0.7188 และ 0.5569 ตามลำดับเมื่อเทียบกับก่อนการใช้เทคนิค SMOTE นอกจากนี้ยังมีค่า F-1 Score ลดลงจาก 0.8456 0.8437 และ 0.8148 เป็น 0.7734 0.8364 และ 0.6300 ตามลำดับ ซึ่งสอดคล้องกับผลการทดลองที่ว่าวิธีการจัดการข้อมูลแบบ SMOTE อาจไม่ทำให้ข้อมูลกระจายตัวได้อย่างเท่าเทียมได้ สำหรับข้อมูลที่ได้รับทางออนไลน์ จึงอาจทำการศึกษาการใช้เทคนิค SMOTE ร่วมกับเทคนิคอื่นๆ ในการศึกษาครั้งต่อไป [37]

4.5 การประเมินค่าความสำคัญของตัวแปรที่ใช้ ในแบบจำลองที่มีประสิทธิภาพมากที่สุดในการทำนายสถานะการผิดนัดชำระหนี้ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

จากการศึกษาวิจัยข้างต้น ได้ทำการคัดเลือกแบบจำลอง คือ แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) ดังนั้นการคัดเลือกตัวแปรที่มีความสำคัญ ได้ทำการจัดลำดับด้วยค่าไอเกน (Eigenvalue) ที่มีค่ามากกว่า 1 ขึ้นไปเช่นเดียวกัน ซึ่งค่าความสำคัญของตัวแปรที่คัดเลือกเป็นดังนี้



รูปที่ 4.2 กราฟแสดงค่าความสำคัญของตัวแปรที่คัดเลือกสำหรับแบบจำลอง

จากรูปที่ 4.2 ข้างต้น แสดงถึงตัวแปรที่มีค่าความสำคัญที่ใช้ในแบบจำลองสำหรับกลุ่มผู้กู้ ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) ซึ่งวัดจากค่า Eigenvalue ที่มากกว่า 1 ได้แก่ Int_rate (อัตราดอกเบี้ย) Funded_amnt_inv (สัดส่วนของเงินที่เคยได้รับ) Loan_amnt (จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ) Funded_amnt (เงินที่เคยได้รับ) Installment (งวดชำระต่อเดือนของสินเชื่อ) Dti (อัตราหนี้สินต่อรายได้ของผู้กู้) Housing_Price (ดัชนีราคาอสังหาริมทรัพย์) Term_60_months (ระยะเวลาของการชำระสินเชื่อ 60 เดือน) Purpose_refinance (วัตถุประสงค์ที่ขอสินเชื่อ สำหรับการ Refinance) Home_ownership_RENT (สถานะการเป็นเจ้าของบ้านแบบเช่า) และ Inq_last_6mths (จำนวนการส่งแบบสอบถามสินเชื่อที่ไม่ใช่รถยนต์และบ้านใน 6 เดือนที่ผ่านมา)

4.6 การทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G)

จากการทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือกสำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) กับข้อมูลจริงในอดีตจากผู้ประกอบการในไทย พบว่าแบบจำลองที่คัดเลือกมีประสิทธิภาพอยู่ในเกณฑ์ยอมรับได้ วัดจาก Accuracy Score ที่อยู่ระดับ 0.6956

4.7 การออกแบบ User Experience (UX) และ User Interface (UI)

การออกแบบ UX/ UI สำหรับการพัฒนาด้านแบบ (Prototype) ของระบบการทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ด้วยวิธีการทำเหมืองข้อมูล มีวัตถุประสงค์ดังต่อไปนี้

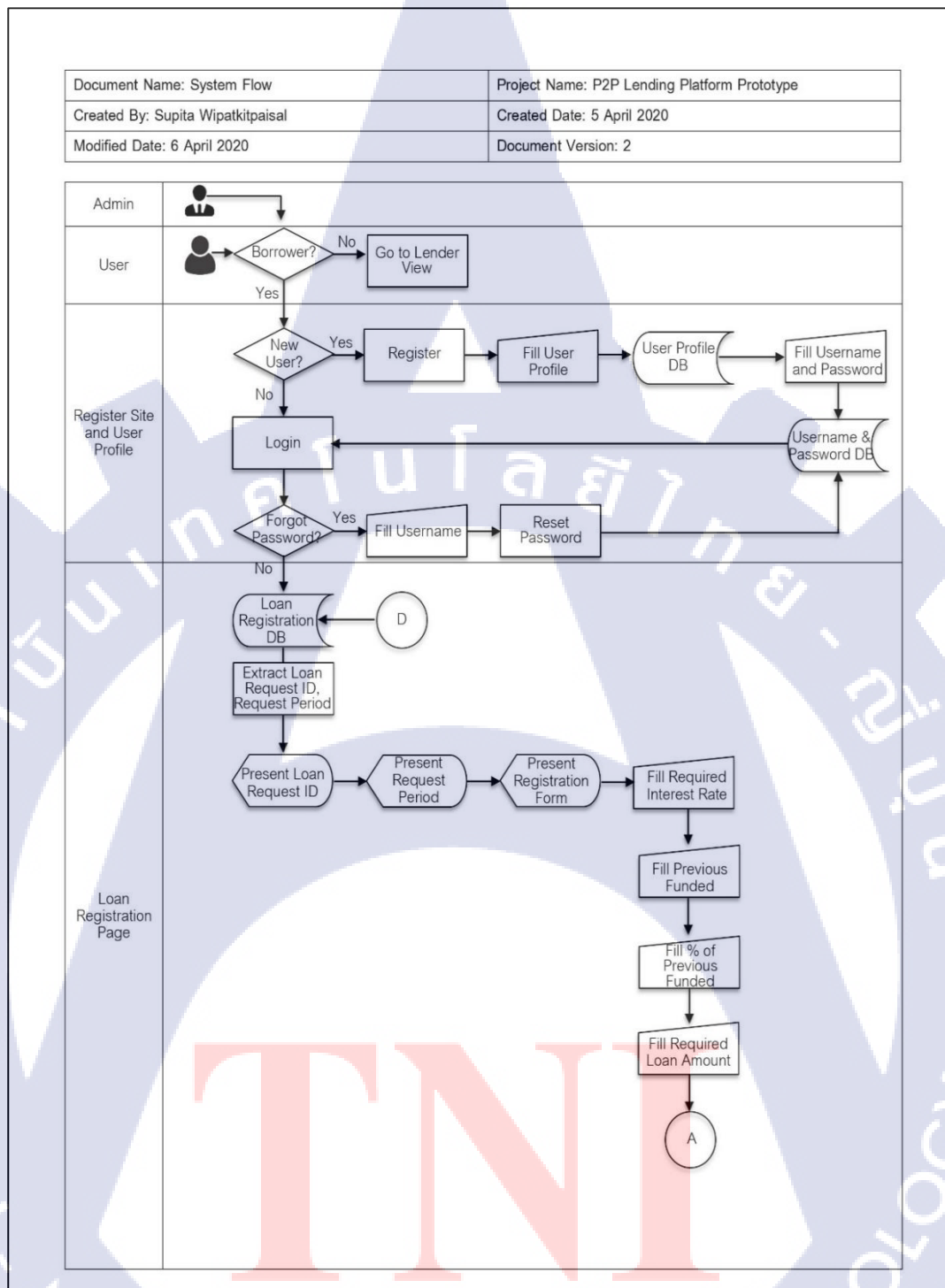
4.7.1 เพื่อใช้เป็นแบบฟอร์มในการเก็บข้อมูลของผู้กู้ เพื่อใช้ในการสมัครขออนุมัติเงินกู้

4.7.2 เพื่อใช้ในการแจ้งสถานะและขั้นตอนในการขออนุมัติเงินกู้

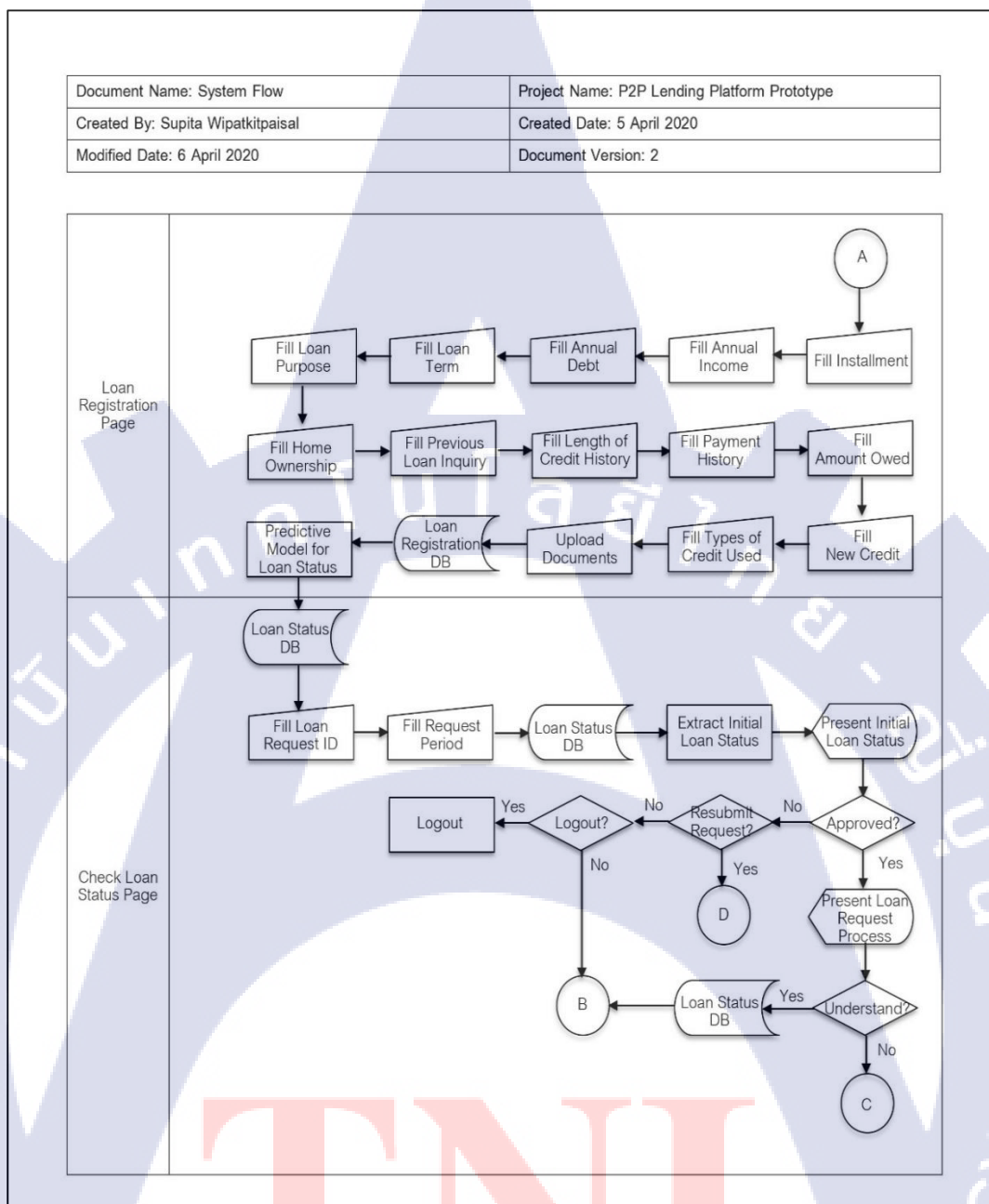
4.7.3 เพื่อใช้ในการรวบรวมคำถามที่พบบ่อยสำหรับผู้กู้

4.7.4 เพื่อใช้ระบุข้อมูลการติดต่อของแพลตฟอร์ม P2P Lending

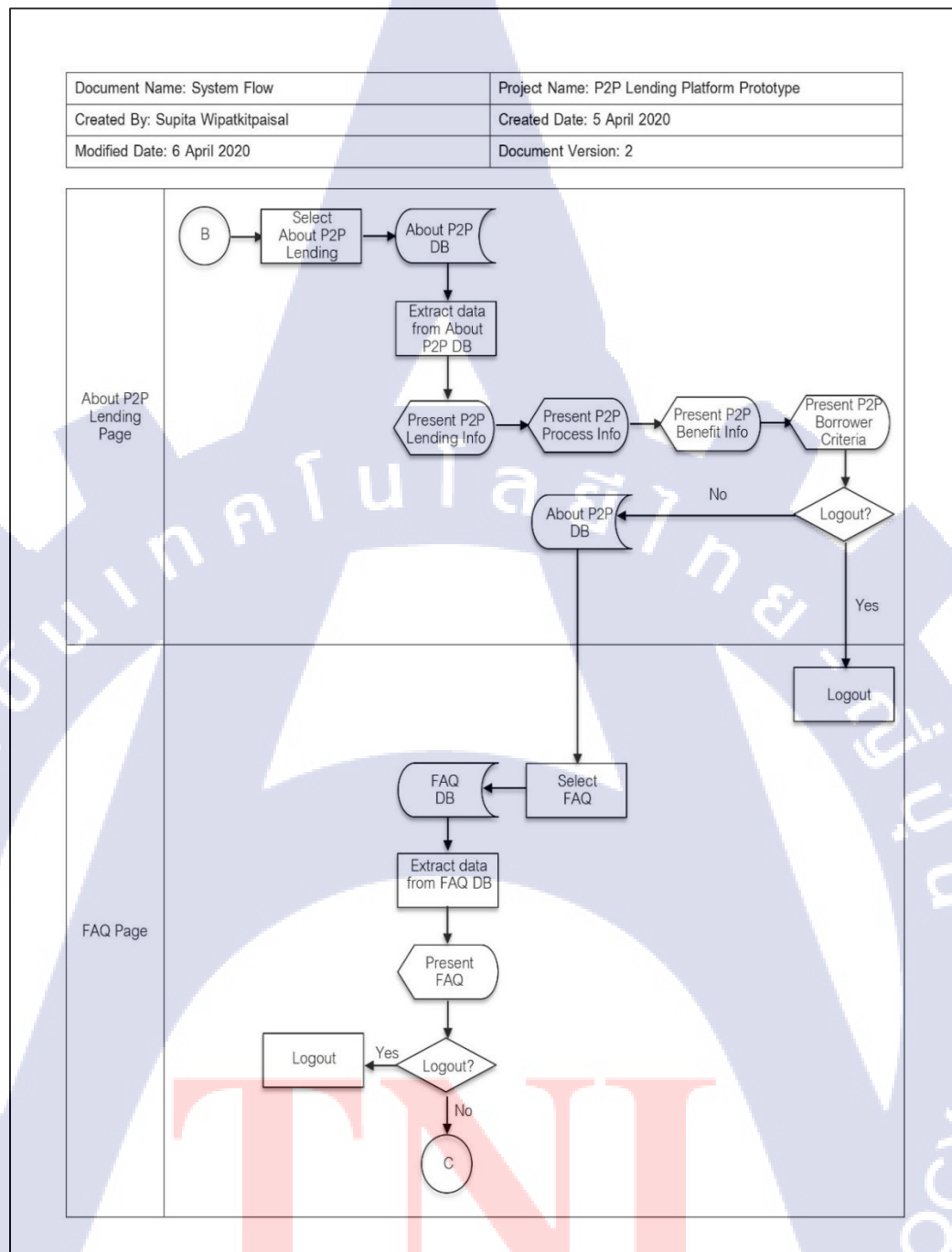
โดยต้นแบบ (Prototype) ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending มี System Flow ดังนี้



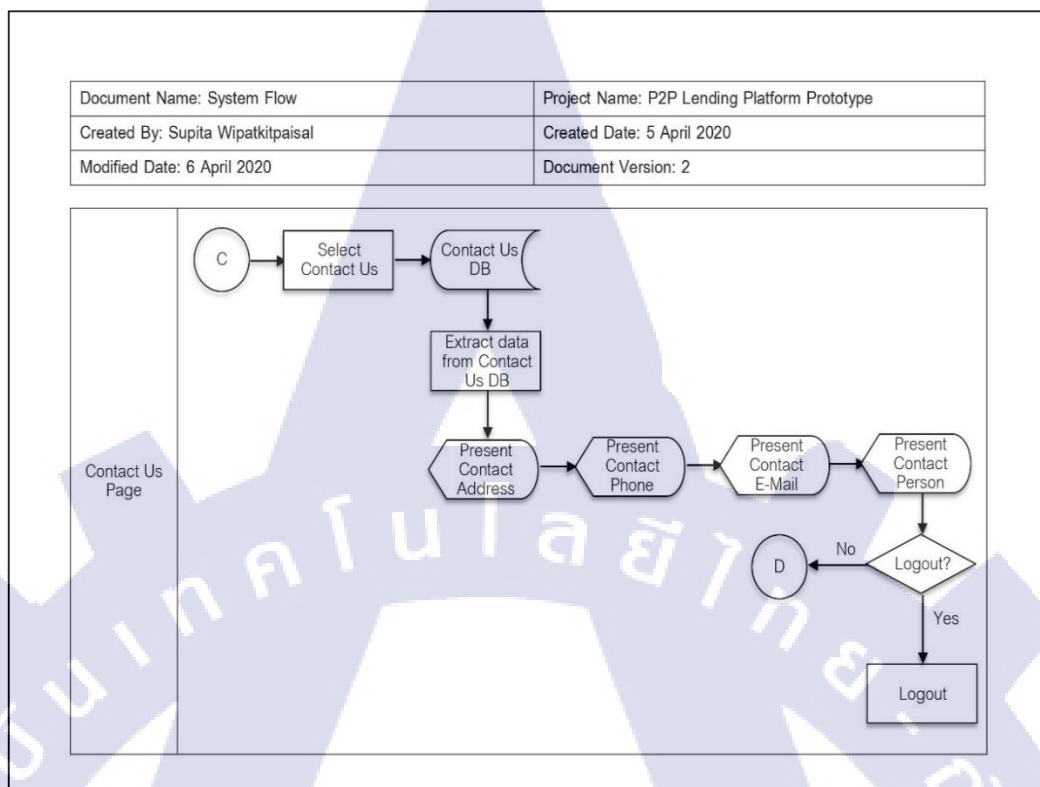
รูปที่ 4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (a)



รูปที่ 4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (b)



รูปที่ 4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (c)



รูปที่ 4.3 System Flow แสดง Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending (d)

จาก System Flow ข้างต้น ได้ทำการออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending ตามรูปด้านล่าง ดังนี้

TNI



รูปที่ 4.4 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิณฑ์ชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Home

Home Page เป็นหน้าแสดงเริ่มต้นสำหรับการเริ่มใช้งาน โดยระบบจะแสดง Logo คำอธิบายเกี่ยวกับบริษัท P2P Better Life Co., Ltd ซึ่งเป็นชื่อของ Platform เบื้องต้น พร้อมทั้งแสดงรายละเอียดเบื้องต้นเกี่ยวกับการดำเนินงานของ Platform ในการจับคู่ระหว่างผู้กู้และผู้ให้กู้ผ่านระบบแพลตฟอร์ม P2P Lending นอกจากนี้ยังแสดงวิดีโอที่เล่นแบบอัตโนมัติ ซึ่งจะใช้อธิบายรายละเอียดของความหมายของ P2P Lending รวมถึงความแตกต่างระหว่างการกู้ยืมผ่าน P2P Lending และการกู้ยืมผ่านธนาคารแบบในอดีต นอกจากนี้ในวิดีโอจะมีการแสดงรายละเอียดเกี่ยวกับข้อดีในการกู้ยืมผ่าน P2P Lending ในแง่ผู้กู้ ผู้ให้กู้ และเศรษฐกิจของประเทศ เพื่อให้ผู้ใช้งานเกิดความเข้าใจเกี่ยวกับ P2P Lending เพิ่มมากขึ้น อีกทั้งระบบจะแสดงปุ่มเพื่อให้ผู้ใช้งานเลือกกดว่าเป็นผู้กู้หรือผู้ให้กู้ โดยสามารถกดเลือก 'You are Borrower' หากผู้ใช้งานเป็นผู้กู้ ซึ่งหลังจากกด ระบบจะนำผู้ใช้งานไปสู่หน้าเว็บไซต์ของผู้กู้ หรือกดเลือก 'You are Lender' หากผู้ใช้งานเป็น

ผู้ให้กู้ ซึ่งหลังจากกด ระบบจะนำผู้ใช้งานไปสู่หน้าเว็บไซต์ของผู้ให้กู้ สำหรับงานวิจัยนี้จะเป็นการออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending จึงขออนุญาตทำการเน้นไปที่หน้าเว็บไซต์ของผู้กู้เป็นหลัก



รูปที่ 4.5 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Login

หน้า Login จะเป็นหน้าเริ่มต้น หลังจากที่ถูกผู้ใช้งานที่เป็นผู้กู้กดปุ่ม ‘You are Borrower’ ในหน้า Home Page เพื่อเข้าสู่การใช้งานเว็บไซต์สำหรับผู้กู้ โดยจะให้ผู้กู้ทำการเข้าสู่ระบบ ซึ่งสามารถเข้าสู่ระบบ โดยการระบุ Username (ที่เป็นอีเมลเท่านั้น) และ Password หลังจากกรอกรระบบจะทำการตรวจสอบว่า Username และ Password ถูกต้องตามที่ฐานข้อมูล Username & Password DB เก็บข้อมูลไว้หรือไม่ หาก Password ไม่ตรงตามข้อมูลที่ระบุบันทึกไว้ในฐานข้อมูล แต่มี Username ที่อยู่ในฐานข้อมูลแล้ว ระบบนำผู้ใช้งานเข้าสู่หน้า Forgot Password หรือหาก

ผู้ใช้งานลืมรหัสผ่าน สามารถกดที่คำว่า 'Forgot Password' ระบบนำผู้ใช้งานเข้าสู่หน้า Forgot Password เช่นกัน ทั้งนี้หาก Username ที่ระบุ (ที่เป็นอีเมลเท่านั้น) ไม่ตรงกับ Username ที่อยู่ในฐานข้อมูล ระบบนำผู้ใช้งานเข้าสู่หน้า Register หรือ หากผู้ใช้งานยังไม่เคยทำการสมัครบัญชีในเว็บไซต์ ผู้ใช้งานสามารถกดที่คำว่า 'Register' ระบบนำผู้ใช้งานเข้าสู่หน้า Register เช่นกัน

รูปที่ 4.6 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Register

หน้า Register เป็นหน้าที่จะปรากฏหลังจากที่ผู้ใช้งานเลือกคำว่า 'Register' ในหน้า Login หรือผู้ใช้งานระบุ Username ที่ยังไม่เคยถูกบันทึกในฐานข้อมูล Username and Password DB โดยหลังจากที่ระบบนำผู้ใช้งานเข้าสู่หน้า Register ระบบจะแสดงช่องให้ผู้ใช้งานระบุ ได้แก่ ชื่อ นามสกุล มือถือ Username Password ยืนยัน Password เพศ ช่วงอายุ ที่อยู่ โดยหลังจากที่ผู้ใช้งานระบุข้อมูลเสร็จเรียบร้อยแล้ว สามารถกดปุ่ม 'Register' ข้อมูลจะถูกบันทึกลงไปที่ฐานข้อมูล User

Profile และ Username and Password DB ทั้งนี้ระบบจะทำการตรวจสอบว่าข้อมูลที่ถูกบันทึกนั้นเป็นไปตามกฎและรูปแบบของข้อมูลดังต่อไปนี้หรือไม่ หากไม่ตรงตามกฎหรือรูปแบบ ระบบจะขอให้ผู้ใช้งานกรอกข้อมูลอีกครั้ง หากผู้ใช้งานต้องการย้อนกลับจากหน้านี้ สามารถกดปุ่ม 'Back' เพื่อย้อนกลับไปที่หน้าหลักอีกครั้ง

ตารางที่ 4.5 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Register

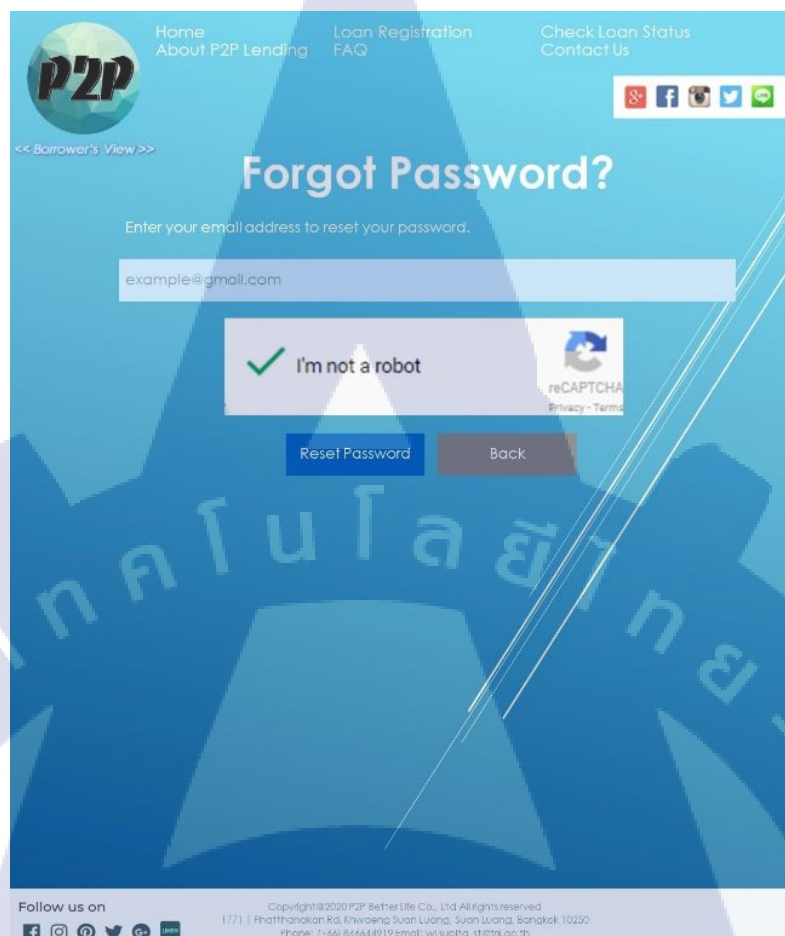
ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Name	ชื่อ	Varchar (50)	ต้องเป็นตัวอักษรเท่านั้น และมีขนาดไม่เกิน 50 ตัวอักษร
Surname	นามสกุล	Varchar (50)	ต้องเป็นตัวอักษรเท่านั้น มีขนาดไม่เกิน 50 ตัวอักษร และไม่ซ้ำกับชื่อ
Mobile	เบอร์มือถือ	Varchar (10)	ต้องเป็นตัวเลขที่เป็นตัวเลขเท่านั้น มีขนาดไม่เกิน 10 ตัวเลข และไม่ซ้ำกับหมายเลขที่มีอยู่ในระบบแล้ว
Username	ชื่อผู้ใช้งาน	Varchar (50)	ต้องเป็นบัญชี Email ของผู้ใช้งาน โดยที่ต้องประกอบด้วย @ และมีตัวอักษรไม่เกิน 50 ตัวอักษร

ตารางที่ 4.5 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Register (ต่อ)

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Password	รหัสผ่าน	Varchar (32)	ต้องมีทั้งตัวอักษร (A-Z) ตัวเลข (0-9) และสัญลักษณ์พิเศษ (! @ # \$ % ^ & * () - _ = + \ [] { } ; : / ? . > <) 8-32 ตัวอักษร โดยตัวอักษรมีทั้งตัวเล็กและตัวใหญ่ ต้องไม่มีตัวเลขเรียงกัน ไม่มีช่องว่าง และ Password ต้องต่างจาก Username นอกจากนี้ต้องไม่มีชื่อหรือนามสกุลของผู้ใช้งาน ประกอบอยู่ใน Password โดยขณะที่ผู้ใช้งานระบุ จะปรากฏเป็นเครื่องหมาย * เพื่อความปลอดภัย
Confirm Password	ยืนยันรหัสผ่าน	Varchar (32)	ต้องเหมือนกับ Password ทุกตัวอักษร โดยขณะที่ผู้ใช้งานระบุ จะปรากฏเป็นเครื่องหมาย * เพื่อความปลอดภัย
Gender	เพศ	Varchar (6)	ผู้ใช้งานต้องเลือกเฉพาะ Female (เพศหญิง) หรือ Male (เพศชาย) เท่านั้น

ตารางที่ 4.5 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Register (ต่อ)

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Age	ช่วงอายุ	Varchar (15)	ผู้ใช้งานต้องเลือกเฉพาะช่วงอายุที่ตามกำหนดเท่านั้น
Address	ที่อยู่	Varchar (50)	ผู้ใช้งานต้องเลือกเฉพาะที่อยู่ (ประเทศ จังหวัด เขต/อำเภอ แขวง/ตำบล ถนน และ บ้านเลขที่) ตามที่กำหนดเท่านั้น โดยเมื่อผู้ใช้งานเลือกข้อมูลที่อยู่ในระดับใหญ่กว่าไปแล้ว ระบบจะแสดงข้อมูลที่อยู่ในระดับเล็กกว่า ให้สอดคล้องกับข้อมูลระดับที่ใหญ่กว่า เช่น เมื่อผู้ใช้งานเลือกประเทศไทย ระบบจะแสดงข้อมูล 76 จังหวัดของประเทศไทยให้เลือก โดยดึงจากฐานข้อมูล User Profile DB



รูปที่ 4.7 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Forgot Password

หน้า Forgot Password เป็นหน้าที่จะปรากฏ หลังจากที่ผู้ใช้งานคลิก ‘Forgot Password’ หรือผู้ใช้งานระบุ Password ผิดในหน้า Login แต่ได้ระบุ Username ที่เคยถูกบันทึกใน Username and Password DB เรียบร้อยแล้ว ซึ่งหลังจากระบบนำผู้ใช้งานมาสู่หน้า Forgot Password แล้ว ระบบจะแสดงช่องให้ผู้ใช้งานกรอก Username หรืออีเมลที่ได้เคยทำการสมัครไปแล้ว และกดยืนยันว่าผู้ใช้งานไม่ใช่หุ่นยนต์ (I’m not a robot) รวมถึง กดปุ่ม ‘Reset Password’ เพื่อให้ระบบส่งรหัสผ่านใหม่ ไปทางอีเมลของผู้ใช้งานที่ระบุ เพื่อให้ผู้ใช้งานตั้งรหัสใหม่ในการเข้าสู่ระบบครั้งถัดไป โดยรหัสใหม่นี้ จะถูกบันทึกแบบออฟเพคเข้าที่ฐานข้อมูล Username and Password DB เพื่อให้ระบบใช้ในการตรวจสอบ Username และ Password จากฐานข้อมูลอีกครั้งในการเข้าสู่ระบบครั้งถัดไป แต่หากผู้ใช้งานต้องการย้อนกลับจากหน้านี้ สามารถกดปุ่ม ‘Back’ เพื่อย้อนกลับไปหน้าจอหลักเข้าสู่ระบบ (Login Page) ได้อีกครั้ง

Home About P2P Lending Loan Registration FAQ Check Loan Status Contact Us

<< Borrower's View >>

User Profile

Name: Supiit

Surname: Wilp

Gender: Female

Age (Years): 25 - 30 Years

Mobile: 0909844844

Username (Email): Supiit@gmail.com

Borrower/ Lender: Borrower

Full Address: 573 Mahachai Road, Samranraj Sub-District, Pranakorn District, Bangkok 10200

Edit Profile Save Back

Follow us on

Copyright © 2020 P2P Better Life Co., Ltd. All rights reserved.
1771 1 Phatthanakan Rd, Khwaeng Suan Luang, Suan Luang, Bangkok 10250
Phone: (+66) 846644919 Email: wlsuplita_sl@int.ac.th

รูปที่ 4.8 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า User Profile

หน้า User Profile จะปรากฏสู่ผู้ใช้งาน หลังจาก que ผู้ใช้งานทำการเข้าสู่ระบบจากหน้า Login หรือสมัครโดยกรอกรายละเอียดจากหน้า Register เรียบร้อย โดยระบบจะดึงข้อมูลจากฐานข้อมูล Username and Password DB และ User Profile DB เพื่อแสดง ชื่อ นามสกุล เพศ ช่วงอายุ Username (ที่เป็นอีเมล) สถานะของผู้ใช้งานที่เป็นผู้กู้ (Borrower) และ ที่อยู่แบบเต็มให้แก่ผู้ใช้งาน เพื่อให้ผู้ใช้งานทำการตรวจสอบข้อมูลที่ตนเองได้บันทึกไว้อีกครั้ง ทั้งนี้หากผู้ใช้งานต้องการแก้ไขข้อมูลที่เคยบันทึกไว้แล้ว ให้กด 'Edit Profile' เพื่อทำการแก้ไขข้อมูล หลังจาก que ผู้ใช้งานได้รับข้อมูลใหม่เรียบร้อยแล้ว ผู้ใช้งานสามารถบันทึกผลการแก้ไขได้ โดยการกด 'Save' หลังจากนั้นระบบจะทำการบันทึกข้อมูลใหม่แบบอัปเดตของผู้ใช้งานลงในฐานข้อมูล Username and Password DB และ User Profile DB เพื่อใช้แสดงข้อมูลใหม่สำหรับการเข้าใช้งานหน้านี้ครั้งถัดไป หากผู้ใช้งานต้องการย้อนกลับจากหน้านี้ สามารถกดปุ่ม 'Back' เพื่อย้อนกลับไป que หน้าหลักอีกครั้ง

Home About P2P Lending Loan Registration FAQ Check Loan Status Contact Us

<< Borrower's View >>

Loan Registration

Loan Request ID: 6655
Loan Request Period: 23/3/2020

1 2 3

* Interest Rate: 1% 5% 12%

* Previous Funded (THB): 1,000,000

* % of Previous Funded: 0% 60% 100%

* Loan Amount (THB): 1,000,000

* Installment (per month): 10,000

* Annual Income (THB): 1,000,000

* Annual Debt (THB): 10,000

* Loan Term (months): 12 months 24 months 36 months 48 months 60 months

* Loan Purpose: Credit Loan Car Loan Home Loan Refinance Business Loan Study Loan Other, specify

* Home Ownership: Own Install Rent No

* Loan Inquiry in last 6 months: 0 1 10

* Checksum

Edit Save Back

Follow us on

Copyright © 2020 P2P BetterLife Co., Ltd All rights reserved
1771 1 Phatthanakan Rd, Khwaeng Suan Luang, Suan Luang, Bangkok 10250
Phone: (+66) 846644919 Email: wlsuplta_31@lnt.ac.th

รูปที่ 4.9 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้า ที่ 1

หน้า Loan Registration คือหน้าที่จะปรากฏสู่ผู้ใช้งาน หากผู้ใช้งานได้ทำการเข้าสู่ผู้ใช้งาน หลังจากผู้ใช้งานได้คลิกที่แท็บ Loan Registration ด้านบน โดยระบบจะแสดง Loan ID ที่เป็นลำดับเลขอัตโนมัติ (Running Number) ของสินเชื่อที่ผู้กู้ขออนุมัติแต่ละสินเชื่อ นอกจากนี้ยังแสดง Loan Request Period เป็นวันที่ที่ผู้ใช้งานเข้ามาที่หน้า Loan Registration นี้ จากนั้นระบบจะแสดงช่องที่ให้ผู้ใช้งานกรอกรายละเอียดที่ใช้ในการสมัครเพื่อขออนุมัติสินเชื่อ โดยที่รายละเอียดทั้งหมดนี้มาจากตัวแปรที่สำคัญที่ถูกคัดเลือกมาจากแบบจำลองพยากรณ์เพื่อทำนายสถานะของผู้กู้ของทั้งสองกลุ่มความเสี่ยง ที่นอกเหนือจากข้อมูลตัวแปรทางด้านดัชนีเศรษฐกิจมหภาค เช่น Housing Price หรือ ดัชนีราคาอสังหาริมทรัพย์ ที่จะดึงข้อมูลระบบจะดึงข้อมูลจาก World Bank API โดยอัตโนมัติตามช่วงเวลาที่คุณทำการเสนอขออนุมัติเงินกู้แล้ว หลังจากผู้ใช้งานได้กรอกข้อมูลเรียบร้อยแล้ว ผู้ใช้งานสามารถกดปุ่ม 'Save' เพื่อบันทึกข้อมูล ซึ่งระบบจะบันทึกข้อมูลลงไปใน Loan Registration

DB หากผู้ใช้งานต้องการแก้ไข สามารถกดปุ่ม ‘Edit’ หากผู้ใช้งานต้องการกลับสู่หน้าหลัก สามารถกดปุ่ม ‘Back’ โดยผู้ใช้งานจะต้องระบุข้อมูลสำหรับหน้า Loan Registration หน้าที 1 ดังต่อไปนี้

ตารางที่ 4.6 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที 1

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Interest Rate	อัตราดอกเบี้ย	Decimal (10)	ต้องเป็นไปตาม % ที่ให้ ผู้ใช้งานเลือก ได้แก่ 1% ถึง 12%
Previous Fund	สินเชื่อที่เคยได้รับ	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น
% of Previous Fund	สัดส่วนของสินเชื่อที่เคยได้รับ	Decimal (10)	ต้องเป็นไปตาม % ที่ให้ ผู้ใช้งานเลือก ได้แก่ 0% ถึง 100%
Loan Amount	จำนวนสินเชื่อที่ต้องการ	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และ ไม่เกิน 5 เท่าของรายได้ และไม่เกิน 50 ล้านบาทต่อ ปี (คำนวณจากการขอ สินเชื่อสะสมจากผู้ใช้งาน คนเดียวกันต่อปี)
Installment	จำนวนเงินที่ชำระต่อเดือน	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และ ไม่ต่ำกว่า 5,000 บาทต่อ เดือน

ตารางที่ 4.6 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที่ 1 (ต่อ)

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Annual Income	รายได้ต่อปี	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และ ไม่ต่ำกว่า 80,000 บาท สำหรับรายได้ต่อปี
Annual Debt	หนี้สินต่อปี	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น ต้อง ไม่เกิน 50,000,000 บาท สำหรับสินเชื่ต่อปี
Loan Term	จำนวนงวด (เดือน)	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และ เป็นไปตามจำนวนเดือนที่ให้ ผู้ใช้งานเลือก ได้แก่ 12 24 36 48 60 เดือน
Loan Purpose	วัตถุประสงค์ในการขอ สินเชื่อ	Varchar (20)	ต้องเป็นไปตามที่ให้ผู้ใช้งาน เลือก ได้แก่ เพื่อนำไปชำระ หนี้บัตรเครดิต เพื่อซื้อ รถยนต์ เพื่อปรับปรุงบ้าน เพื่อรีไฟแนนซ์ เพื่อ การศึกษา เพื่อทำธุรกิจ
Home Ownership	สถานะการเป็น เจ้าของบ้าน	Varchar (20)	ต้องเป็นไปตามที่ให้ผู้ใช้งาน เลือก ได้แก่ มีบ้าน ผ่อน เช่า หรือไม่มีบ้าน
Loan Inquiry	จำนวนการส่ง แบบสอบถามสินเชื่อ ใน 6 เดือนที่ผ่านมา	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และ เป็นไปตามที่ให้ผู้ใช้งาน เลือก ได้แก่ 1 ถึง 10 ครั้ง ในช่วง 6 เดือนที่ผ่านมา

Home About P2P Lending Loan Registration FAQ Check Loan Status Contact Us

<< Borrower's View >>

Loan Registration

Loan Request ID: 6655
Loan Request Period: 23/3/2020

1 2 3

- * Numbers of Credit Cards
- * First Credit Card Receipt until Now (Years)
- * Numbers of Applied Loans In Last Year
- * New Loan Request until Now Period (Years)
- * Current Loan Balance (THB)
- * Non-Mortgage Loan Balance (THB)
- * Last Default Payment Time (Years)
- * Numbers of Default Loan
- * Total Credit Card Limits
- * Bankruptcy History In Last 10 Years

* Default

Edit Save Back

Follow us on

Copyright © 2020 P2P BetterLife Co., Ltd. All rights reserved.
1771 1 Phatthanakan Rd, Khwaeng Suan Luang, Suan Luang, Bangkok 10250
Phone: (+66) 846644919 Email: wlsupfla_31@int.ac.th

รูปที่ 4.10 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิดนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้าที่ 2

นอกจากนี้ ผู้ใช้งานยังต้องกรอกข้อมูลเพื่อใช้ในการประเมิน FICO Score โดยใช้คำถามใน 5 หมวดหมู่ดังที่ได้กล่าวไว้แล้วในบทที่ 2 ที่ผ่านมา ได้แก่ หมวดประวัติระยะเวลาของการขอสินเชื่อ หมวดประวัติการชำระสินเชื่อ หมวดจำนวนสินเชื่อที่ติดค้าง หมวดการขอสินเชื่อใหม่ และหมวดชนิดของสินเชื่อที่ใช้ หลังจากผู้ใช้งานได้กรอกข้อมูลเรียบร้อยแล้ว ผู้ใช้งานสามารถกดปุ่ม 'Save' เพื่อบันทึกข้อมูล ซึ่งระบบจะบันทึกข้อมูลลงไปใน Loan Registration DB หากผู้ใช้งานต้องการแก้ไข สามารถกดปุ่ม 'Edit' หากผู้ใช้งานต้องกลับสู่หน้าที่ 1 สามารถกดปุ่ม 'Back' โดยผู้ใช้งานจะต้องระบุข้อมูลสำหรับหน้า Loan Registration หน้าที่ 2 ดังต่อไปนี้

ตารางที่ 4.7 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที่ 2

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Numbers of Credit Cards	จำนวนบัตรเครดิตที่ผู้กู้ถืออยู่	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50
First Credit Card Receipt until Now Period	ระยะเวลา (ปี) ที่ถือบัตรเครดิตใบแรก	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50
Numbers of Applied Loans in Last Year	จำนวนสินเชื่อหรือสินเชื่อบัตรเครดิตที่ผู้กู้ได้สมัครในปีก่อนหน้า	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50
New Loan Request until Now Period	ระยะเวลา (ปี) จนถึงปัจจุบันที่ผู้กู้ได้ขอสินเชื่อหรือสินเชื่อบัตรเครดิตล่าสุด	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50
Current Loan Balance	วงเงินสินเชื่อหรือสินเชื่อบัตรเครดิตของผู้กู้ในปัจจุบัน	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และต้องเท่ากับตัวเลขที่ระบุใน Current Debt
Non-Mortgage Loan Balance	วงเงินของสินเชื่อทั้งหมดที่นอกเหนือจากสินเชื่อจำนอง	Bigint (20)	ต้องเป็นตัวเลขเท่านั้น และต้องน้อยกว่า Current Loan Balance
Last Default Payment Time	ระยะเวลา (ปี) ล่าสุดที่ผู้กู้ผิดนัดชำระสินเชื่อหรือสินเชื่อบัตรเครดิต	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50

ตารางที่ 4.7 แสดงชนิดและรูปแบบของข้อมูลที่ระบุลงในหน้า Loan Registration หน้าที่ 2 (ต่อ)

ข้อมูล	ความหมาย	ชนิดของข้อมูล	กฎของข้อมูลที่ระบบตรวจสอบ
Numbers of Default Loan	จำนวนสินเชื่อหรือสินเชื่อบัตรเครดิตที่เกินกำหนดการชำระหนี้	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50
Total Credit Card Limits	สัดส่วน % ของวงเงินจำกัดต่อวงเงินทั้งหมดสำหรับสินเชื่อบัตรเครดิต	Decimal (10)	ต้องเป็นไปตาม % ที่ให้ผู้ใช้งานเลือก ได้แก่ 0% ถึง 100%
Bankruptcy History in Last 10 Years	จำนวนครั้งการล้มละลายในช่วง 10 ปีที่ผ่านมา	Bigint (10)	ต้องเป็นตัวเลขเท่านั้น และเป็นไปตามที่ให้ผู้ใช้งานเลือก ได้แก่ 0 ถึง 50

Home About P2P Lending Loan Registration FAQ Check Loan Status Contact Us

<< Borrower's View >>

Loan Registration

Loan Request ID: 6655
Loan Request Period: 23/3/2020

1 2 3

- * ID Card or Passport
- * House Registration
- * Book Bank
- * Salary Slips or Bank Cash Flow in Previous 6 Months
- * First Credit Card
- * Past Loan Request in Previous 5 Years

* Acceptable Upload Files are .pdf, .jpg, .png only with size not exceeded than 10 MB per file.

Edit Save Back Submit

Follow us on

Copyright © 2020 P2P Better Life Co., Ltd. All rights reserved.
1771 1 Phatthanakan Rd, Khwaeng Suan Luang, Suan Luang, Bangkok 10250
Phone: (+66) 846644919 Email: wlsuplta_st@int.ac.th

รูปที่ 4.11 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิมนัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Loan Registration หน้า ที่ 3

นอกจากนี้ ผู้ใช้งานยังต้องทำการอัปโหลดเอกสารที่เกี่ยวข้อง ได้แก่ เอกสารสำเนาบัตรประชาชนหรือพาสปอร์ตที่ยังมีอายุการใช้งาน เอกสารสำเนาทะเบียนบ้าน เอกสารหน้าบัญชีธนาคาร (สำหรับการรับสินเชื่อกู้ยืมได้รับการอนุมัติ หรือการชำระดอกเบี้ยหรือเงินต้น เมื่อครบกำหนดแต่ละงวด) เอกสารสลิปเงินเดือนหรือบัญชีธนาคารที่แสดงกระแสเงินสดหมุนเวียนในช่วง 6 เดือนที่ผ่านมา เอกสารบัตรเครดิตใบแรก เอกสารการขอสินเชื่อที่ผ่านมาในช่วง 5 ปีที่ผ่านมา โดยระบบจะทำการตรวจสอบว่าไฟล์ที่อัปโหลดนั้นมีนามสกุล .pdf หรือ .jpg หรือ .png หรือไม่ และต้องมีขนาดไม่เกิน 10 MB ต่อเอกสารหรือไม่ หากเกินระบบจะแจ้งเตือนให้ผู้ใช้งานทำการอัปโหลดเอกสารใหม่อีกครั้ง ผู้ใช้งานสามารถกดปุ่ม 'Save' เพื่อบันทึกข้อมูล ซึ่งระบบจะบันทึกข้อมูลลงไปใน Loan Registration DB หากผู้ใช้งานต้องการแก้ไข สามารถกดปุ่ม 'Edit' หากผู้ใช้งานต้องการกลับสู่หน้าที่ 2 สามารถกดปุ่ม

‘Back’ หากผู้ใช้งานสามารถกดปุ่ม ‘Submit’ หากผู้ใช้งานต้องการส่งคำขออนุมัติสินเชื่อ จากนั้นระบบจะนำผู้ใช้งานเข้าสู่เมนูหลัก

The screenshot displays the 'Check Loan Status' interface of a P2P Lending system. At the top, there are navigation links: Home, About P2P Lending, Loan Registration, FAQ, Check Loan Status, and Contact Us. The user is logged in as 'Hi Supit' and can click 'Logout'. The main section shows the 'Loan Request ID' as 6655 and the 'Loan Request Period' as 23/3/2020. The 'Loan Status' is '1st Round: Initial Approved'. Below this, an 'Overall Loan Request Process' flowchart is shown with four steps: 1st Round: Initial Loan Review (to do: Check for initial loan status within 1 hour), 2nd Round: Online Interview* (to do: Interview online with our officials within 1 working day), 3rd Round: Final Loan Review (to do: Final loan status (after interview)), and 4th Round: Sign Contract** (to do: Sign Contract (after loan maturing)). At the bottom, there are buttons for 'I Understand' and 'Contact Official'. A footer section includes social media links, copyright information for P2P Lender Ltd, and contact details.

รูปที่ 4.12 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิदनัดชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Check Loan Status

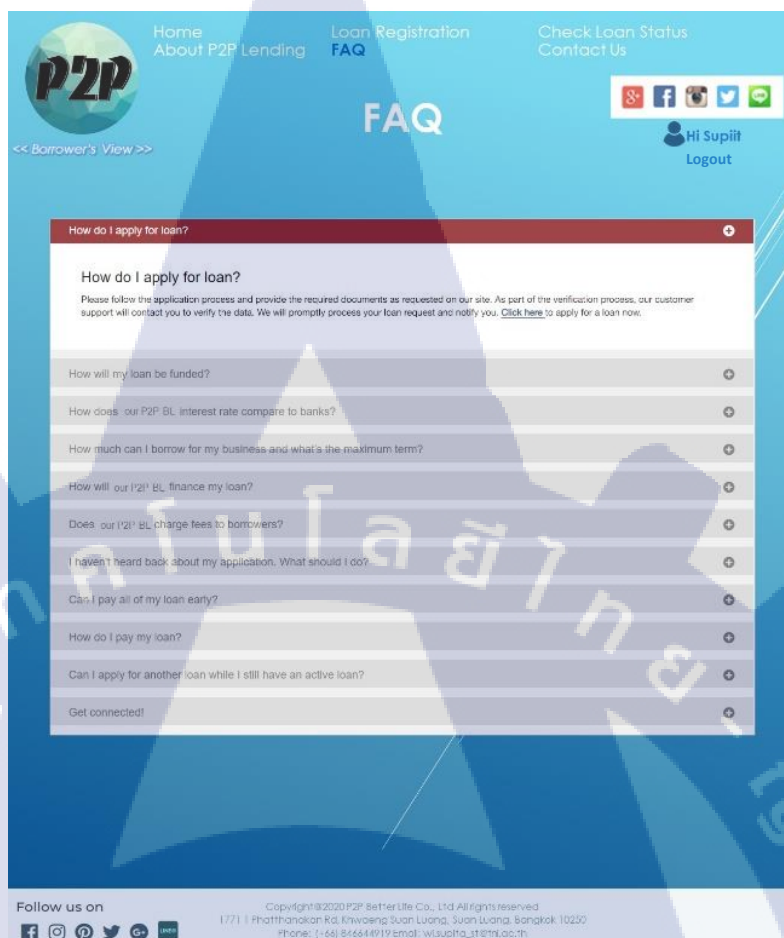
หน้า Check Loan Status จะปรากฏสู่ผู้ใช้งานเมื่อผู้ใช้งานคลิกแท็บ Check Loan Status โดยในหน้านี้ ผู้ใช้งานจะต้องเลือก Loan Request ID ซึ่งได้เคยแสดงไว้ในหน้า Loan Registration โดยข้อมูลที่ให้เลือกจะเป็น Loan Request ID ระบบจะดึงข้อมูลจากฐานข้อมูล Loan Registration DB เพื่อแสดงเฉพาะ Loan Request ID ที่ผู้กู้เคยทำการขออนุมัติสินเชื่อไว้ นอกจากนี้ผู้ใช้งานยังต้องเลือก Loan Request Period ที่เป็นวันที่ผู้ใช้งานได้ทำการขออนุมัติสินเชื่อเรียบร้อยแล้ว โดยระบบจะทำการตรวจสอบจากฐานข้อมูลฐานข้อมูล Loan Registration DB ว่าวันที่ที่ผู้ใช้งานเลือกดังกล่าวสอดคล้องกับวันที่สินเชื่อที่ระบุจาก Loan Request ID ที่ผู้ใช้งานได้ทำการขออนุมัติสินเชื่อหรือไม่ หากไม่ตรงระบบจะแจ้งเตือนให้ผู้ใช้งานเลือก Loan Request ID อีกครั้ง ทั้งนี้หลังจากที่ผู้ใช้งานได้ส่ง

คำขออนุมัติสินเชื่อแล้ว ข้อมูลที่ถูกบันทึกใน Loan Registration DB จะถูกนำเข้าไปสู่กระบวนการเตรียมข้อมูล (Transform) และวิเคราะห์ในแบบจำลองพยากรณ์เพื่อทำนายสถานะสินเชื่อของผู้กู้ เพื่อทำนายว่าผู้กู้นั้นได้รับการอนุมัติสินเชื่อเบื้องต้นหรือไม่ ซึ่งจะบันทึกลงในฐานข้อมูล Loan Status DB จากนั้นระบบจะดึงข้อมูลสถานะสินเชื่อที่ได้รับการวิเคราะห์จากแบบจำลอง เพื่อแสดงในส่วนของ Loan Status ของหน้านี้ โดยจะแสดงเป็นสถานะสินเชื่อรอบแรก (1st Round) ว่าผู้กู้นั้นได้รับการอนุมัติเบื้องต้น (Initial Approved) หรือไม่ได้รับการอนุมัติ (Not Approved) หากผู้กู้นั้นได้รับการอนุมัติเงินกู้ ระบบจะแสดงกระบวนการของการอนุมัติสินเชื่อขั้นตอนต่อไป ได้แก่ การสัมภาษณ์ทางออนไลน์อีกครั้ง เพื่อตรวจสอบข้อมูลที่ถูกได้ให้มา (2nd Round: Online Interview) หลังจากนั้น 1 วันทำการจะทำการแจ้งสถานะการขออนุมัติสินเชื่อครั้งสุดท้าย (3rd Round: Final Loan Review) และเมื่อผู้กู้ได้รับการเลือกจับคู่กับผู้ให้กู้ที่สนใจจะลงทุนให้สินเชื่อแก่ผู้กู้ ระบบจะส่งสัญญาให้กับผู้กู้ทางอีเมลอีกครั้ง เพื่อให้ผู้กู้ทำการเซ็นสัญญาและส่งกลับให้กับทางแพลตฟอร์ม (4th Round: Sign Contract) อย่างไรก็ตาม หากผู้กู้นั้นไม่ได้รับการอนุมัติการขอสินเชื่อเบื้องต้น ตั้งแต่รอบแรกจากการทำนายของแบบจำลอง ระบบจะแสดง Loan Status ในหน้านี้ว่า 1st Round: Not Approved และแจ้งเตือนว่าผู้กู้ต้องการทำการสมัครสินเชื่อใหม่อีกครั้งหรือไม่ หากผู้กู้ต้องการโดยการกดยืนยัน ระบบจะนำผู้ใช้งานไปสู่หน้า Loan Registration หน้าที่ 1 แต่หากผู้กู้ไม่ต้องการ ระบบจะนำผู้ใช้งานไปสู่หน้า Logout หากผู้ใช้งานกดออกจากระบบ ระบบจะนำผู้ใช้งานออกจากระบบ แต่หากไม่ได้กดออกจากระบบ ระบบจะนำผู้ใช้งานสู่เมนูหลัก



รูปที่ 4.13 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิณฑ์ชำระหนี้
ของผู้กู้ใน P2P Lending – หน้า About P2P Lending

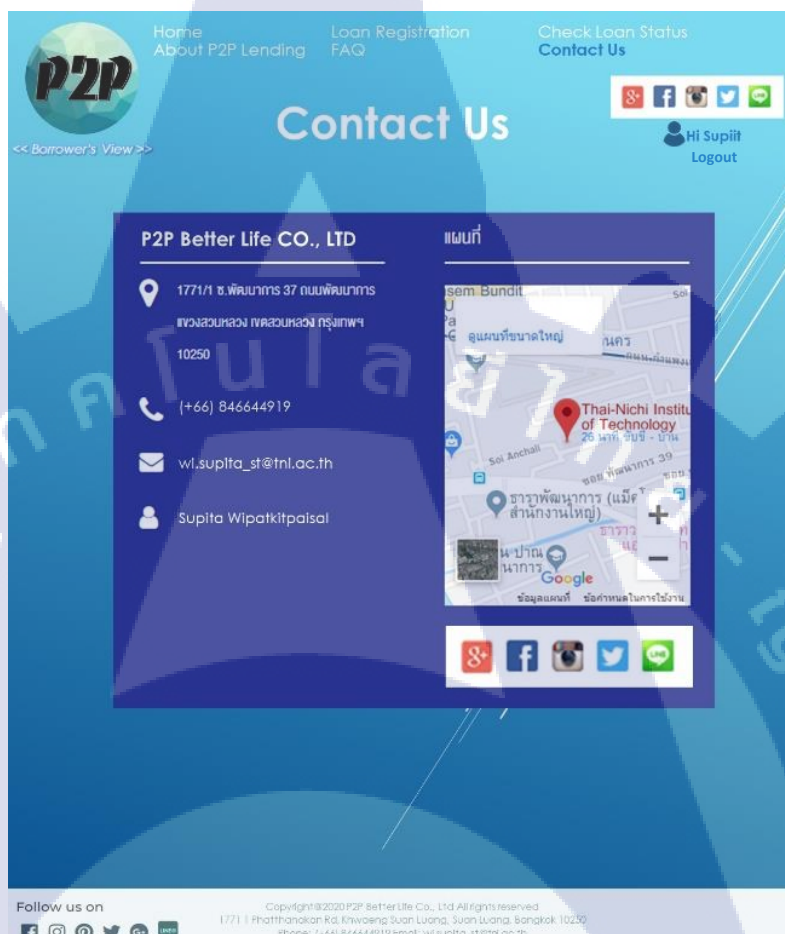
หน้า About P2P Lending จะปรากฏสู่ผู้ใช้งาน เมื่อผู้ใช้งานคลิกแท็บ About P2P Lending ด้านบน โดยระบบจะดึงข้อมูลจากฐานข้อมูล About P2P DB เพื่อแสดงข้อมูลทั่วไปเกี่ยวกับแพลตฟอร์ม P2P Lending ได้แก่ การเปรียบเทียบระหว่างการกู้ยืมตัวกลางทางการเงินรูปแบบเดิม กับการกู้ยืมในรูปแบบ P2P Lending กระบวนการขอสินเชื่อในแพลตฟอร์ม P2P Lending ข้อดีของ P2P Lending และ เกณฑ์ของผู้กู้และเงินกู้ที่สามารถขอสินเชื่อจากแพลตฟอร์ม P2P Lending ได้ เมื่อแสดงข้อมูลทั้งหมดเรียบร้อยแล้ว ระบบจะทำการบันทึกเวลาที่ผู้ใช้งานเข้าใช้งานในหน้านี้อยู่ในฐานข้อมูล About P2P DB หลังจากหากผู้ใช้งานกดปุ่ม Logout เพื่อออกจากระบบ ระบบจะนำผู้ใช้งานสู่หน้า Logout หากผู้ใช้งานกดยืนยันการออกจากระบบ ระบบจะนำผู้ใช้งานออกจากกระบบ แต่หากไม่ได้กดออกจากระบบ ระบบจะนำผู้ใช้งานสู่เมนูหลัก



รูปที่ 4.14 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิन्छชำระหนี้
ของผู้กู้ใน P2P Lending – หน้า FAQ

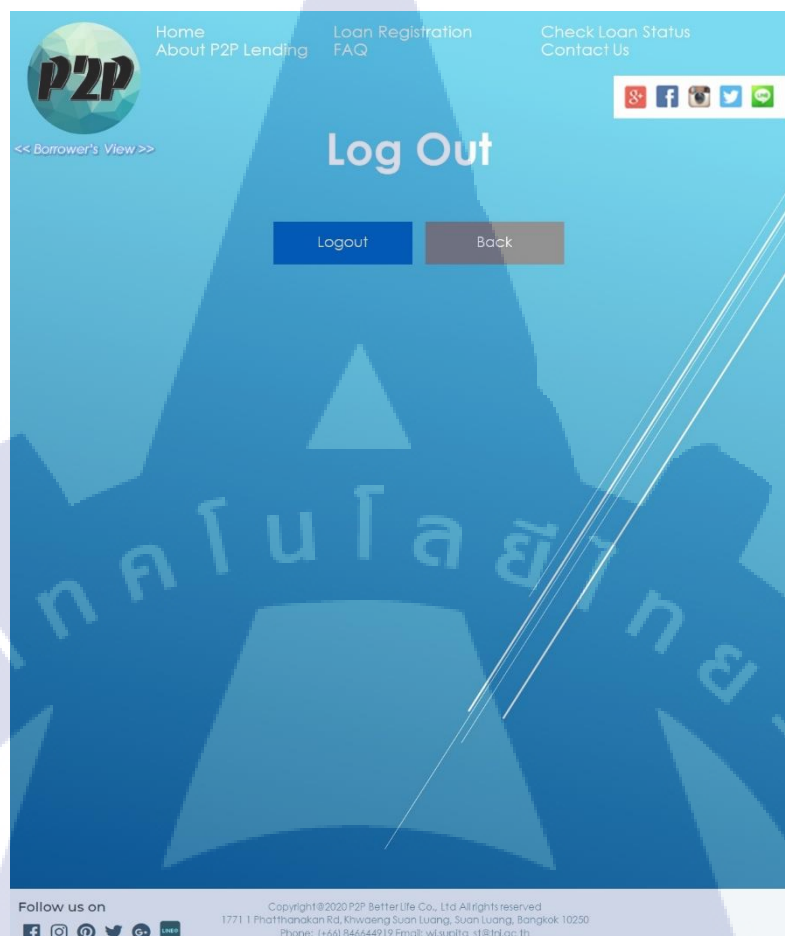
หน้า FAQ จะปรากฏสู่ผู้ใช้งาน เมื่อผู้ใช้งานคลิกแท็บ FAQ ด้านบน โดยระบบจะดึงข้อมูลจากฐานข้อมูล FAQ DB เพื่อแสดงข้อมูลทั่วไปเกี่ยวกับคำถามที่พบบ่อย ซึ่งได้รวบรวมมาจากแพลตฟอร์มต่างๆทั่วโลก ตามที่ได้อธิบายไว้ในบทที่ 3 ที่ผ่านมา โดยระบบจะดึงข้อมูลจากฐานข้อมูล FAQ DB เพื่อแสดงข้อมูลเกี่ยวกับคำถามที่พบบ่อยต่างๆ ได้แก่ วิธีการสมัครขออนุมัติสินเชื่อ (ระบบจะนำผู้ใช้งานไปสู่หน้า Loan Registration หากผู้ใช้งานคลิกคำว่า Click Here) วิธีการที่สินเชื่อจะได้รับการลงทุนหรือให้สินเชื่อ อัตราดอกเบี้ยของแพลตฟอร์มเมื่อเทียบกับการกู้ยืมผ่านธนาคาร ค่าธรรมเนียมในการกู้ยืมในแพลตฟอร์ม P2P Lending การติดตามสถานะการขออนุมัติสินเชื่อ การชำระดอกเบี้ยหรือเงินต้นก่อนครบกำหนด การสมัครขอสินเชื่อเพิ่มเติม หากมีสินเชื่อเดิมที่ได้สมัครในแพลตฟอร์มอยู่แล้ว ช่องทางการติดต่อแพลตฟอร์ม (ระบบจะนำผู้ใช้งานไปสู่หน้า FAQ หากผู้ใช้งานคลิกคำว่า Click Here) หลังจากหากผู้ใช้งานกดปุ่ม Logout เพื่อออกจากระบบ ระบบจะนำผู้ใช้งาน

สู่หน้า Logout หากผู้ใช้งานกดยืนยันการออกจากระบบ ระบบจะนำผู้ใช้งานออกจากระบบ แต่หากไม่ได้กดออกจากระบบ ระบบจะนำผู้ใช้งานสู่เมนูหลัก



รูปที่ 4.15 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิวดนชำระหนี้ของผู้กู้ใน P2P Lending – หน้า Contact Us

หน้า Contact Us จะปรากฏสู่ผู้ใช้งาน เมื่อผู้ใช้งานคลิกแท็บ Contact Us ด้านบน โดยระบบจะดึงข้อมูลจากฐานข้อมูล Contact Us DB เพื่อแสดงข้อมูลทั่วไปเกี่ยวกับช่องทางการติดต่อแพลตฟอร์ม ได้แก่ ที่อยู่ เบอร์โทรศัพท์ติดต่อ อีเมลติดต่อ ผู้ติดต่อหลัก รวมถึงแผนที่ที่ตั้งของบริษัท ที่ดึงข้อมูลจาก Google Maps หลังจากหากผู้ใช้งานกดปุ่ม Logout เพื่อออกจากระบบ ระบบจะนำผู้ใช้งานสู่หน้า Logout หากผู้ใช้งานกดยืนยันการออกจากระบบ ระบบจะนำผู้ใช้งานออกจากระบบ แต่หากไม่ได้กดออกจากระบบ ระบบจะนำผู้ใช้งานสู่เมนูหลัก



รูปที่ 4.16 การออกแบบ UX/ UI สำหรับ Prototype ของระบบทำนายสถานะการผิมนัดชำระหนี้
ของผู้กู้ใน P2P Lending – หน้า Logout

หน้า Logout จะปรากฏสู่ผู้ใช้งาน เมื่อผู้ใช้งานคลิกปุ่ม Logout เพื่อออกจากระบบ หากผู้ใช้งานกดยืนยันการออกจากระบบ ระบบจะนำผู้ใช้งานออกจากระบบ แต่หากไม่ได้กดออกจากระบบ ระบบจะนำผู้ใช้งานสู่เมนูหลัก

โดยในขั้นตอนต่อไป หลังจากการนำ Prototype ของระบบทำนายสถานะการผิมนัดชำระหนี้ของผู้กู้ใน P2P Lending ที่ได้ออกแบบ UX/UI ไปพัฒนาต่อ จะทำการสำรวจความพึงพอใจจากผู้ใช้งานในช่วงทดลองระบบ เพื่อนำมาปรับปรุงระบบ ก่อนที่จะนำมาใช้งานจริง

บทที่ 5

บทสรุป ข้อเสนอแนะ บทวิจารณ์

จากผลการศึกษาแบบจำลองที่ใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ ในแพลตฟอร์ม P2P Lending โดยใช้เทคนิคเหมืองข้อมูล ซึ่งการวิจัยนี้มีวัตถุประสงค์ตามบทที่ 1 ได้นำเสนอข้อสรุปตามลำดับดังนี้

5.1 สรุปผลและอภิปรายผลการวิจัย

5.2 ปัญหาและอุปสรรคในการทำวิจัย

5.3 ข้อเสนอแนะงานวิจัย

5.1 สรุปและอภิปรายผลการวิจัย

5.1.1 ผลการศึกษาปัจจัยที่ใช้ทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้

(1) ปัจจัยด้านผู้กู้และสินเชื่อ ทั้งด้านฐานะทางการเงิน เช่น ระดับคะแนนเครดิต หนี้สินต่อรายได้ วัตถุประสงค์ของการกู้ เป็นต้น และด้านข้อมูลส่วนตัว เช่น อายุ การศึกษา อาชีพ สถานภาพ เป็นต้น

(2) ตัวแปรทางเศรษฐศาสตร์มหภาค เช่น GDP อัตราการว่างงาน อัตราเงินเฟ้อ, ผลตอบแทนของดัชนีหลักทรัพย์ ดัชนีราคาอสังหาริมทรัพย์ เป็นต้น

5.1.2 ผลจากการเปรียบเทียบแบบจำลองพยากรณ์แนวโน้มการผิดนัดชำระหนี้ของผู้กู้

จากการใช้โปรแกรม Python ในการวิเคราะห์และเปรียบเทียบแบบจำลองจากการคัดเลือกตัวแปรด้วยวิธีที่ต่างกัน ได้แก่ ได้แก่ Pearson's Correlation Recursive Feature Elimination (RFE) และ Factor Analysis ร่วมกับเทคนิคการทำเหมืองข้อมูลที่ต่างกัน ได้แก่ Logistic Regression Decision Tree ID3 Decision Tree C4.5 และ Naïve Bayes รวมถึงการเปรียบเทียบระหว่างการสร้างแบบจำลองที่ใช้ และไม่ใช้การจัดการข้อมูลด้วยวิธี SMOTE พบว่า

(1) แบบจำลองที่มีประสิทธิภาพ สำหรับผู้กู้ความเสี่ยงต่ำถึงปานกลาง (กลุ่มเกรดสินเชื่อ A-B) คือ แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) ซึ่งมีค่า Accuracy Score และ F-1 Score อยู่ที่ 0.8843 และ 0.9383 ตามลำดับโดยมีตัวแปรที่ถูกคัดเลือก ได้แก่ Int_rate (อัตราดอกเบี้ย) Dti (อัตราหนี้สินต่อรายได้ของผู้กู้) Funded_amnt_inv (สัดส่วนของเงินที่เคยได้รับ) Loan_amnt (จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ) Funded_amnt (เงินที่เคยได้รับ) Installment (งวดชำระต่อเดือนของสินเชื่อ)

Open_acc (จำนวนการเปิดวงเงินสินเชื่อจากประวัติด้านสินเชื่อของผู้กู้) Housing_Price (ดัชนีราคาอสังหาริมทรัพย์) Term_36_months (ระยะเวลาของการชำระสินเชื่อ 36 เดือน) Purpose_refinance (วัตถุประสงค์ที่ขอสินเชื่อ สำหรับการ Refinance) และ Home_ownership_RENT (สถานะการเป็นเจ้าของบ้านแบบเช่า)

(2) แบบจำลองที่มีประสิทธิภาพ สำหรับกลุ่มผู้กู้ความเสี่ยงสูง (กลุ่มเกรดสินเชื่อ C-G) คือ แบบจำลองที่คัดเลือกตัวแปรด้วยวิธี Factor Analysis และใช้วิธีจัดการข้อมูลด้วยวิธี SMOTE ผ่านเทคนิค Decision Tree (C4.5) เช่นเดียวกัน ซึ่งมีค่า Accuracy Score และ F-1 Score อยู่ที่ 0.7778 และ 0.8621 ตามลำดับ แต่มีตัวแปรที่ถูกคัดเลือกแตกต่างกัน ได้แก่ Int_rate (อัตราดอกเบี้ย) Funded_amnt_inv (สัดส่วนของเงินที่เคยได้รับ) Loan_amnt (จำนวนเงินกู้ที่ผู้กู้ขออนุมัติ) Funded_amnt (เงินที่เคยได้รับ) Installment (งวดชำระต่อเดือนของสินเชื่อ) Dti (อัตราหนี้สินต่อรายได้ของผู้กู้) Housing_Price (ดัชนีราคาอสังหาริมทรัพย์) Term_60_months (ระยะเวลาของการชำระสินเชื่อ 60 เดือน) Purpose_refinance (วัตถุประสงค์ที่ขอสินเชื่อ สำหรับการ Refinance) Home_ownership_RENT (สถานะการเป็นเจ้าของบ้านแบบเช่า) และ Inq_last_6mths (จำนวนการส่งแบบสอบถามสินเชื่อที่ไม่ใช่รถยนต์และบ้านใน 6 เดือนที่ผ่านมา)

(3) จากการนำข้อมูลจริงในอดีตจากผู้ประกอบการในไทยมาเป็นข้อมูลทดสอบประสิทธิภาพของแบบจำลองที่คัดเลือก พบว่าค่า Accuracy Score อยู่ในเกณฑ์ยอมรับได้ สำหรับแบบจำลองของกลุ่มผู้กู้ทั้งสองกลุ่มความเสี่ยง โดยมีค่า Accuracy Score อยู่ที่ระดับ 0.8541 และ 0.6956 สำหรับผู้กู้กลุ่มเกรดสินเชื่อ A-B และ C-G ตามลำดับ ซึ่งให้เห็นว่าข้อมูลที่นำมาใช้ในการทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ในงานวิจัยนี้ สามารถใช้เป็นแนวทางให้กับผู้ประกอบการไทยในการใช้วิเคราะห์และทำนายสถานะของผู้กู้ในแพลตฟอร์ม P2P Lending ได้

5.1.3 ประโยชน์ที่ได้รับจากผลการศึกษา

(1) ทำให้ทราบถึงปัจจัยที่มีผลต่อแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ สำหรับแพลตฟอร์ม P2P Lending

(2) ทำให้ทราบถึงแนวทางของเทคนิคการทำเหมืองข้อมูลที่เหมาะสมกับการสร้าง predictive model เพื่อใช้ในการทำนายแนวโน้มการผิดนัดชำระหนี้ของผู้กู้

(3) ผู้ประกอบการสามารถใช้เป็นแนวทางในการรวบรวมข้อมูลของผู้กู้และสินเชื่อให้เพียงพอต่อการตัดสินใจของนักลงทุน

(4) ผู้ประกอบการสามารถใช้เป็นแนวทางในคัดกรองผู้กู้ที่มีแนวโน้มในการผิดนัดชำระหนี้ เพื่อการพัฒนาระบบแพลตฟอร์มให้มีประสิทธิภาพต่อไป

(5) นักลงทุนสามารถใช้เป็นแนวทางประกอบการตัดสินใจในการลงทุน P2P Lending

(6) ผู้กู้สามารถใช้เป็นแนวทางในการเปิดเผยข้อมูลที่เกี่ยวข้องต่อการตัดสินใจของนักลงทุนในคำขออนุมัติเงินกู้ P2P Lending

5.2 ปัญหาและอุปสรรคในการทำงานวิจัย

5.2.1 การรวบรวมข้อมูลที่จะนำมาใช้วิเคราะห์เพื่อสร้างแบบจำลองทำนายสถานะแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ในแต่ละความเสี่ยง ที่มาจากแหล่งข้อมูลที่น่าเชื่อถือ มีจำนวนเพียงพอต่อการวิเคราะห์ และข้อมูลมาจากแพลตฟอร์มที่มีบริบทใกล้เคียงกับแพลตฟอร์ม P2P Lending ในประเทศไทย เพื่อนำมาใช้เป็นแนวทางสำหรับผู้ประกอบการไทยในการใช้พัฒนาแพลตฟอร์มต่อไป

5.2.2 การศึกษาและคัดเลือกตัวแปรที่เกี่ยวข้องและส่งผลกระทบต่อแนวโน้มการผิดนัดชำระหนี้ของผู้กู้

5.2.3 การศึกษาและเลือกใช้เทคนิคการจัดการข้อมูล เพื่อเตรียมข้อมูลก่อนการสร้างแบบจำลอง รวมถึงการช่วยเพิ่มประสิทธิภาพของแบบจำลอง

5.2.4 การศึกษาและเปรียบเทียบเทคนิคการคัดเลือกตัวแปรแบบต่างๆ เพื่อนำมาใช้ในการคัดเลือกจากตัวแปรทั้งหมดที่มีจำนวนมาก เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพ

5.2.5 การศึกษาและเลือกใช้เทคนิคการทำเหมืองข้อมูลให้สอดคล้องกับชนิดของข้อมูลที่นำมาศึกษา เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพสำหรับการทำนายสถานะแนวโน้มการผิดนัดชำระหนี้ของผู้กู้ในแต่ละกลุ่มความเสี่ยง

5.3 ข้อเสนอแนะงานวิจัย

ผู้วิจัยมีข้อเสนอแนะสำหรับการศึกษาวิจัยในอนาคต ดังนี้

5.3.1 ศึกษาตัวแปรด้านกฎหมายและการควบคุมการซื้อขาย P2P Lending ของหน่วยงานภาครัฐ เช่น การกำหนดให้นักลงทุนรายย่อยสามารถลงทุนได้ไม่เกิน 5 แสนบาทต่อปี หรือการกำหนดให้ผู้ประกอบการสามารถคิดดอกเบี้ยได้ไม่เกิน 15% ต่อปี เป็นต้นส่งผลกระทบต่อตัดสินใจของนักลงทุนใน P2P Lending ในประเทศไทยหรือไม่ อย่างไร

5.3.2 ศึกษาเทคนิคการทำเหมืองข้อมูล ในรูปแบบเทคนิคผสม (hybrid model) เพื่อให้ได้แนวทางในการสร้างแบบจำลองการพยากรณ์ที่มีประสิทธิภาพเพิ่มมากขึ้น

5.3.3 การใช้เทคนิคการจัดการข้อมูลที่ไม่สมดุลด้วยวิธี SMOTE ร่วมกับเทคนิคอื่นๆ เช่น เทคนิค Fuzzy-Rough Selection, เทคนิค Kernel Trick เป็นต้น

5.3.4 ศึกษารูปแบบการนำเทคโนโลยี face recognition มาช่วยประเมินความน่าเชื่อถือทางเครดิตของผู้กู้ในแพลตฟอร์ม P2P Lending เพื่อใช้สังเกตว่าผู้กู้มีการแสดงความรู้สึกเมื่อมีการพูดเท็จหรือไม่ โดยนำมาเปรียบเทียบกับฐานข้อมูลที่ได้เก็บมา

5.3.5 การพัฒนาแพลตฟอร์มเต็มรูปแบบสำหรับ P2P Lending เพื่อใช้ในการจับคู่ระหว่างคำขออนุมัติเงินกู้ของผู้กู้และความต้องการลงทุนของผู้ให้กู้

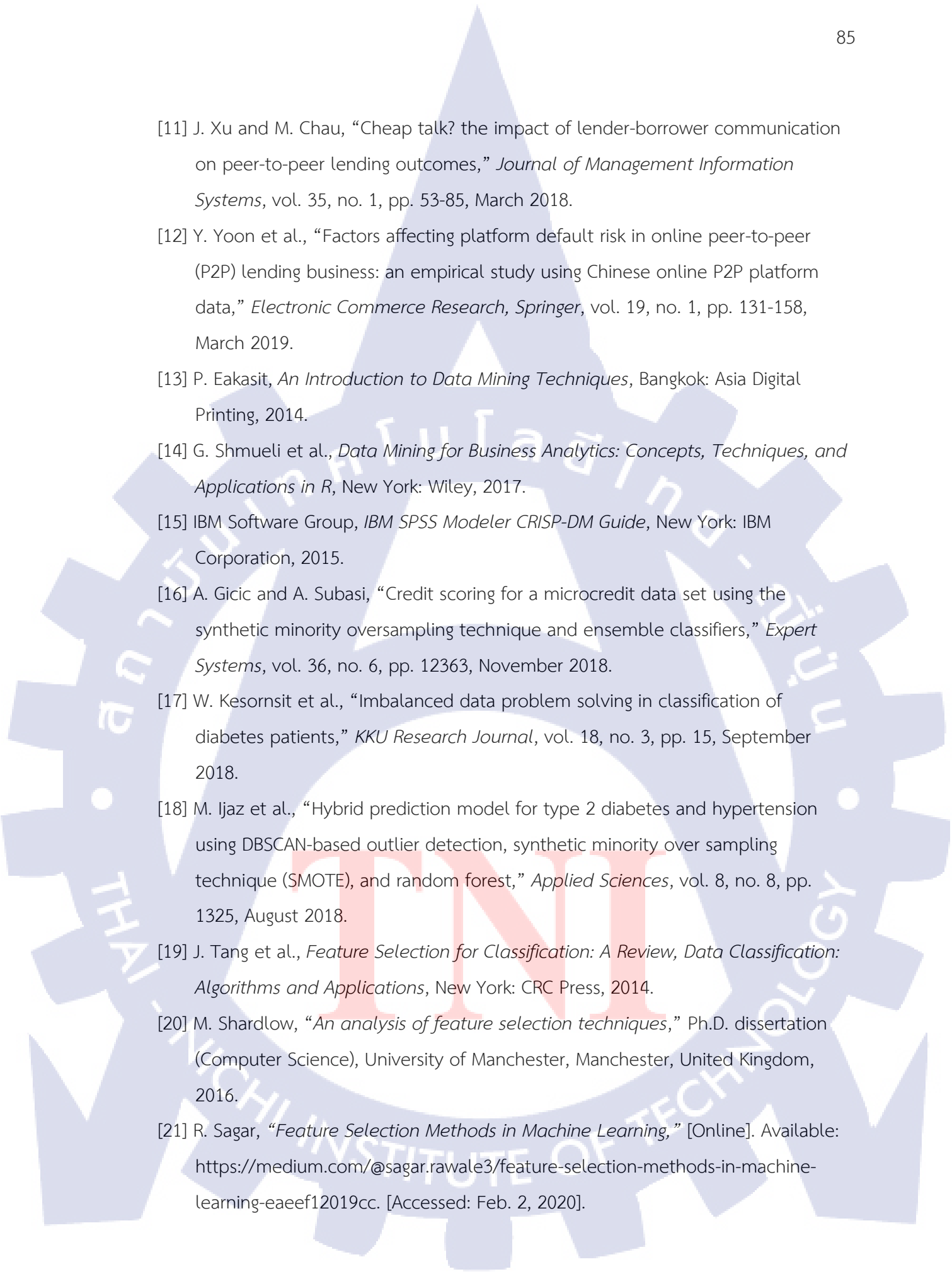


บรรณานุกรม

TNI

บรรณานุกรม

- [1] Q. Tao et al., "Who can get money? evidence from the Chinese peer-to-peer lending platform," *Information Systems Frontiers*, vol. 19, no. 3, pp. 425-441, March 2017.
- [2] X. Chen et al., "Are investors rational or perceptual in P2P lending?" *Information Systems and E-Business Management*, vol. 14, no. 4, pp. 921-944, February 2016.
- [3] X. Zeng et al., "A decision support model for investment on P2P lending platform," *PloS One*, vol. 12, no. 9, pp. e0184242, September 2017.
- [4] H. Yekta, "Essays on matching markets: course allocation, team formation, and P2P lending markets," Ph.D. dissertation (Business Administration), Connecticut University, Connecticut, USA, 2018.
- [5] Transparency Market Research, "Peer-to-Peer Lending Market: Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2016 – 2024," [Online]. Available: <https://www.transparencymarketresearch.com/peer-to-peer-lending-market.html>. [Accessed: Jan. 5, 2019].
- [6] C. Lu and L. Zhang, "Research on risk factors identification of P2P lending platforms," *American Journal of Industrial and Business Management*, vol. 8, no. 5, pp. 1344, May 2018.
- [7] C. Canfield, "Determinants of default in P2P lending: the Mexican case," *Independent Journal of Management & Production*, vol. 9, no. 1, pp. 1-24, March 2018.
- [8] Q. Wei and Q. Zhang, "P2P lending risk contagion analysis based on a complex network model," *Discrete Dynamics in Nature and Society*, vol. 2016, no. 5013954, pp. 1-8, June 2016.
- [9] H. Zhao et al., "P2P lending survey: platforms, recent advances and prospects," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 6, pp. 72:1-72:5, July 2017.
- [10] J. Kim and S. Cho, "Towards repayment prediction in peer-to-peer social lending using deep learning," *Mathematics*, vol. 7, no. 11, pp. 1041, November 2019.

- 
- [11] J. Xu and M. Chau, "Cheap talk? the impact of lender-borrower communication on peer-to-peer lending outcomes," *Journal of Management Information Systems*, vol. 35, no. 1, pp. 53-85, March 2018.
- [12] Y. Yoon et al., "Factors affecting platform default risk in online peer-to-peer (P2P) lending business: an empirical study using Chinese online P2P platform data," *Electronic Commerce Research, Springer*, vol. 19, no. 1, pp. 131-158, March 2019.
- [13] P. Eakasit, *An Introduction to Data Mining Techniques*, Bangkok: Asia Digital Printing, 2014.
- [14] G. Shmueli et al., *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*, New York: Wiley, 2017.
- [15] IBM Software Group, *IBM SPSS Modeler CRISP-DM Guide*, New York: IBM Corporation, 2015.
- [16] A. Gicic and A. Subasi, "Credit scoring for a microcredit data set using the synthetic minority oversampling technique and ensemble classifiers," *Expert Systems*, vol. 36, no. 6, pp. 12363, November 2018.
- [17] W. Kesornsit et al., "Imbalanced data problem solving in classification of diabetes patients," *KKU Research Journal*, vol. 18, no. 3, pp. 15, September 2018.
- [18] M. Ijaz et al., "Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest," *Applied Sciences*, vol. 8, no. 8, pp. 1325, August 2018.
- [19] J. Tang et al., *Feature Selection for Classification: A Review, Data Classification: Algorithms and Applications*, New York: CRC Press, 2014.
- [20] M. Shardlow, "An analysis of feature selection techniques," Ph.D. dissertation (Computer Science), University of Manchester, Manchester, United Kingdom, 2016.
- [21] R. Sagar, "Feature Selection Methods in Machine Learning," [Online]. Available: <https://medium.com/@sagar.rawale3/feature-selection-methods-in-machine-learning-eaeef12019cc>. [Accessed: Feb. 2, 2020].

- [22] S. Abhini, “Feature Selection with Sklearn and Pandas,” [Online]. Available: <https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b>. [Accessed: Mar, 12, 2020].
- [23] N. Avinash, “Introduction to Factor Analysis in Python,” [Online]. Available: <https://www.datacamp.com/community/tutorials/introduction-factor-analysis>. [Accessed: Mar. 12, 2020].
- [24] M. Kishan, “5 Ways to Handle Missing Values in Machine Learning Datasets,” [Online]. Available: <https://analyticsindiamag.com/5-ways-handle-missing-values-machine-learning-datasets>. [Accessed: Mar. 12, 2020].
- [25] R. Sebastian, “MinMax Scaling,” [Online]. Available: http://rasbt.github.io/mlxtend/user_guide/preprocessing/minmax_scaling. [Accessed: Mar. 12, 2020].
- [26] ยุทธ ไกยวรรณ, “หลักการและการใช้การวิเคราะห์การถดถอยโลจิสติกสำหรับการวิจัย,” *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย*, ปีที่ 4, ฉบับที่ 1: 1-12, หน้า 3-11, มกราคม, 2555.
- [27] S. Jadhav and H. Channe, “Comparative study of k-nn, naive bayes and decision tree classification techniques,” *International Journal of Science and Research*, vol. 5, no. 1, pp. 1842-1845, January 2016.
- [28] S. Singh and P. Gupta, “Comparative study ID3, cart and C4.5 decision tree algorithm: a survey,” *International Journal of Advanced Information Science and Technology (IJAIST)*, vol. 27, no. 27, pp. 97-103, July 2014.
- [29] Y. Hang, “Status quo of the US and UK’s P2P lending regulation,” *Global Journal of Economics and Business Administration*, vol. 1, no. 1: 0032-0036, January 2016.
- [30] พรชนก เทพขาม., “เศรษฐกิจแบ่งปัน: บทเรียนการกู้ยืมผ่านช่องทางอิเล็กทรอนิกส์ (Peer-to-peer Lending),” [Online]. Available: https://www.bot.or.th/Thai/MonetaryPolicy/ArticleAndResearch/FAQ/FAQ_148.pdf. [Accessed: 12 มีนาคม 2563].
- [31] M. Polena and T. Regner, “Determinants of borrowers’ default in P2P lending under consideration of the loan risk class,” *MDPI*, vol. 9, no. 4, pp. 1-12, October 2018.

- [32] M. Polena, “*Performance analysis of credit scoring models on lending club data*,” M.S. Thesis (Social Sciences), Charles University, Prague, Czechia, 2017.
- [33] P. Khast, “*UX/UI design process for a peer to peer financial platform*,” M.S. Thesis (Information Technology), University of Applied Sciences, Hamburg, Germany, 2017.
- [34] A. Arora, “Evaluation measures for data mining tasks. winter school on data mining techniques and tools for knowledge discovery in agricultural datasets,” *Indian Agricultural Statistics Research Institute*, vol. 7, no. 10, pp. 145-152, November 2011.
- [35] MyFICO, “*Fico Score Estimator*,” [Online]. Available: <https://www.myfico.com/fico-credit-score-estimator/estimator>. [Accessed: Feb. 2, 2020].
- [36] A. Bachmann et al., “Online peer-to-peer lending-a literature review,” *Journal of Internet Banking and Commerce*, vol. 16, no. 2, pp. 1, August 2011.
- [37] C. Gong and L. Gu, “A novel SMOTE-based classification approach to online data imbalance problem,” *Mathematical Problems in Engineering*, vol. 2016, no. 5685970, pp. 1-14, May 2016.
- [38] J. Brown, “What is an eigenvalue?,” *JALT Testing & Evaluation SIG Newsletter*, vol. 5, no. 1, pp. 15-19, April 2001.

ภาคผนวก

TNI

ภาคผนวก ก.
การตีพิมพ์ผลงานวิจัย

TNI

ได้เข้าร่วมงานวิจัยในงานประชุมวิชาการ TNIAC 2019

