

# การพยากรณ์การผิנדชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป

ปณพัฒน์ ประสานบุญเลิศ\*

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

เอกรัฐ รัฐกาญจน์

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

\* Correspondence: panapat.pra@stu.nida.ac.th

## บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อพัฒนาแบบจำลองเพื่อพยากรณ์การผิנדชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป และศึกษาปัจจัยที่มีผลต่อการผิנדชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป จากการศึกษาข้อมูลของผู้กู้จากการกู้ยืมเงินระหว่างบุคคลทั่วไปผ่านแพลตฟอร์มออนไลน์ Bondora (<https://www.bondora.com>) และข้อมูลเศรษฐกิจของสหภาพยุโรป ตั้งแต่เดือนกุมภาพันธ์ 2009 – ธันวาคม 2022 โดยแบบจำลองที่นำมาใช้ในการเปรียบเทียบประสิทธิภาพในการพยากรณ์การผิנדชำระหนี้ ได้แก่ XGBoost, Extra Tree, Random Forest และ Logistic Regression และใช้ SHAP Feature Importance ในการศึกษาปัจจัยที่มีผลต่อการผิנדชำระหนี้ จากผลการวิจัยพบว่า แบบจำลอง XGBoost เป็นแบบจำลองที่มีประสิทธิภาพที่ดีกว่าแบบจำลองชนิดอื่น จากนั้นทางผู้วิจัยได้ใช้เทคนิค Resampling และ GridSearchCV เพื่อพัฒนาแบบจำลองให้มีประสิทธิภาพที่ดียิ่งขึ้น ซึ่งพบว่าแบบจำลอง XGBoost ที่ใช้เทคนิค Resampling และ GridSearchCV ให้ผลลัพธ์ที่ดีขึ้น โดยมีค่า  $AUC = 81.20\%$  และ  $Accuracy = 73.01\%$  ส่วนปัจจัยที่มีผลต่อการผิנדชำระหนี้มากที่สุด 5 อันดับแรก คือ จำนวนเงินผ่อนชำระต่องวด, ระยะเวลากู้ยืม หนี้สินรวม ผลิตภัณฑ์มวลรวมของประเทศ (GDP ทั้งสหภาพยุโรป) และอัตราดอกเบี้ยเงินกู้ยืม ตามลำดับ

**คำสำคัญ:** ผิנדชำระหนี้, กู้ยืมเงินระหว่างบุคคล, แพลตฟอร์มออนไลน์, Bondora, การเรียนรู้ด้วยเครื่องจักร

# Default or Not : Forecasting Credit Default for Online Peer-to-peer Lending in Europe Union

**Panapat Prasanboonlert\***

Faculty of Applied Statistics, National Institute of Development Administration

**Ekarat Rattagan**

Faculty of Applied Statistics, National Institute of Development Administration

\* **Correspondence:** panapat.pra@stu.nida.ac.th

## **Abstract**

This research aims to develop a model to forecast credit default for online peer-to-peer lending (P2P) in Europe Union and to study factors affecting credit default for online peer-to-peer lending (P2P) in Europe Union. Based on the study of information of borrowers who borrow money through online platform “Bondora” (<https://www.bondora.com>) and EU economics data from February 2009 – December 2022. The models used to compare the effectiveness of credit default forecasting were XGBoost, Extra Tree, Random Forest, and Logistic Regression, and SHAP Feature Importance was used to study the factors affecting credit default. According to the research results, the XGBoost model has better performance than any other model. Then, the researcher used Resampling and GridSearchCV techniques to improve the efficiency of the model. It was found that the XGBoost model using Resampling and GridSearchCV gave better results with AUC = 81.19% and Accuracy = 73.01%. The top 5 factors that have the greatest effect on credit default are monthly payment, loan duration, total liabilities, Gross National Product (GDP) and loan interest rates respectively.

**Keywords:** Credit Default, Peer-to-Peer lending, online platform, Bondora, Machine Learning

## 1. บทนำ

### 1.1 ที่มาและความสำคัญ

การให้กู้ยืมแบบ Peer-to-Peer (P2P) ได้รับความสนใจอย่างมากในช่วงไม่กี่ปีที่ผ่านมา และกลายเป็นรูปแบบองค์กรที่เหนือกว่าทางเศรษฐกิจ เมื่อเปรียบเทียบกับธุรกิจธนาคารแบบดั้งเดิม การให้กู้ยืมแบบ P2P เป็นรูปแบบการระดมทุนแบบคราวด์ฟันดิงที่ได้รับความนิยมมากที่สุด โดยที่แพลตฟอร์มจะรวบรวมเงินทุนจำนวนเล็กน้อยจากบุคคลในกลุ่ม เพื่อเป็นเงินทุนให้แก่บุคคลหรือธุรกิจกู้ยืม ซึ่งไม่เหมือนกับธนาคารพาณิชย์ที่เป็นเจ้าหนี้เพียงผู้เดียว ทำให้แพลตฟอร์มนี้จะไม่รับความเสี่ยงโดยตรง ในขณะที่ธนาคารสะสมความเสี่ยงโดยแสดงสถานะในงบดุล แพลตฟอร์มจะกระจายความเสี่ยงโดยกระจายไปยังผู้ใช้งานบนแพลตฟอร์ม (Rainer Lenz, 2016)

ในมุมมองของนักลงทุน การให้กู้ยืมแบบ P2P เป็นวิธีที่น่าสนใจในการกระจายพอร์ตการลงทุนและเพิ่มประสิทธิภาพในระยะยาว โดยเฉพาะอย่างยิ่งผู้เชี่ยวชาญหลายคนเชื่อว่า หุ้นและพันธบัตร มีความเสี่ยงมากในช่วงเวลาที่ผ่านมา และด้วยประโยชน์ของเทคโนโลยีออนไลน์ ทำให้การให้กู้ยืมแบบ P2P ทำงานโดยไม่ต้องมีสาขาและพนักงานของธนาคาร จึงให้บริการถูกกว่าสถาบันการเงินแบบดั้งเดิม ส่งผลให้ผู้กู้ที่มีปัญหาทางการเงินสามารถเข้าถึงการให้บริการได้เช่นกัน ด้วยอัตราดอกเบี้ยต่ำ ขั้นตอนการสมัครที่ง่ายขึ้น และรวดเร็ว ในขณะเดียวกันผู้ให้กู้สามารถได้รับผลตอบแทนข้างต้นในอัตราดอกเบี้ยสูงกว่าเงินฝาก แต่การลงทุนในแบบ P2P ก็มีความเสี่ยงเช่นกัน นั่นก็คือ ความเสี่ยงจากการผิดนัดชำระหนี้ เป็นปัจจัยเสี่ยงที่สำคัญในการทดสอบพฤติกรรมของผู้กู้แบบ P2P โดยสินเชื่อกู้ยืมจะไม่มีหลักประกันและผู้ให้กู้แสวงหาผลตอบแทนที่สูงขึ้นเพื่อชดเชยความเสี่ยงทางการเงินที่พวกเขาได้รับ นอกจากนี้พวกเขายังจำเป็นต้องตัดสินใจภายใต้ความไม่สมดุลของข้อมูลที่เป็นประโยชน์ต่อผู้ให้กู้ในการตัดสินใจอย่างมีเหตุผล ผู้ให้กู้ต้องการลดความเสี่ยงของการผิดนัดชำระหนี้จากการตัดสินใจให้กู้ยืมแต่ละครั้ง และตระหนักถึงผลตอบแทนที่ชดเชยความเสี่ยง (Rainer Lenz, 2016)

โดยแพลตฟอร์มที่จะนำมาศึกษา ชื่อว่า Bondora ก่อตั้งเมื่อปี 2009 ในตลาดนอร์ดิก เป็นแพลตฟอร์มการให้กู้ยืมแห่งแรกในสหภาพยุโรป โดยเป็นแพลตฟอร์มที่ได้รับความนิยมมาก มีการปล่อยเงินกู้มากกว่า 45 ล้านยูโร และมีผู้กู้บนแพลตฟอร์มจำนวนมากกว่า 100,000 คน โดยในปี 2021 มีการลงทุนในแพลตฟอร์มทั้งหมดกว่า 500 ล้านยูโร นักลงทุนได้รับผลตอบแทนรวมมากกว่า 60 ล้านยูโร บัญชี Instagram มีผู้ติดตามมากกว่า 10,000 คน และได้เปิดตัวแอปการลงทุน Go & Grow บนอุปกรณ์สมาร์ตโฟนเพื่ออำนวยความสะดวกสบายให้แก่กลุ่มทุนอีกด้วย (Bondora, 2022)

จากการทบทวนวรรณกรรมพบว่า นักวิจัยหลายท่านยังให้ความสนใจ และดำเนินการศึกษาการผิดนัดชำระหนี้จากการให้กู้ยืมเงินระหว่างบุคคลทั่วไปผ่านแพลตฟอร์มออนไลน์ จากเหตุที่กล่าวมาจึงเป็นเหตุจูงใจให้ผู้วิจัยสนใจศึกษาการพยากรณ์การผิดนัดชำระหนี้จากการให้กู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป โดยผู้วิจัยหวังเป็นอย่างยิ่งว่าจะเป็นประโยชน์ต่อผู้ที่สนใจสามารถนำไปใช้ได้ถูกต้องแม่นยำที่สุด สามารถอธิบายให้ทราบถึงปัจจัยที่มีผลต่อการผิดนัดชำระหนี้จากการให้กู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป เพื่อเพิ่มเติมและขยายองค์ความรู้ รวมถึงผลการวิจัยที่เกี่ยวข้อง

### 1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อพยากรณ์การผิดนัดชำระหนี้จากการให้กู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป

1.2.2 เพื่อศึกษาปัจจัยที่มีผลต่อการผิดนัดชำระหนี้จากการให้กู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 ทฤษฎีที่เกี่ยวข้อง

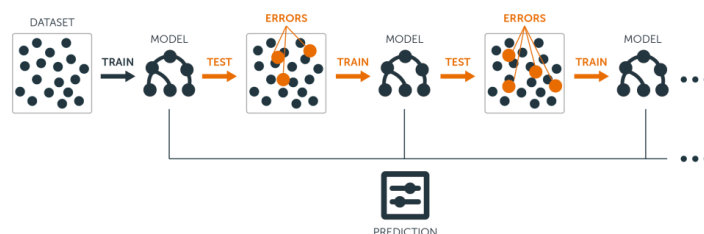
#### 2.1.1 ทฤษฎีการวิเคราะห์สินเชื่อ (5Cs)

เป็นหลักการวิเคราะห์สินเชื่อของธนาคารพาณิชย์ (เนตรชนก สุริยวโรภาส, 2561) เพื่อลดความเสี่ยงที่จะเกิดขึ้นจากการผิดนัดชำระหนี้ ซึ่งคำนึงถึงปัจจัย 5 ประการ ได้แก่ คุณลักษณะ (Character) เงินทุน (Capital) หลักประกัน (Collateral) ความสามารถในการชำระหนี้ (Capacity) และเงื่อนไขการชำระหนี้ (Condition) โดยมีรายละเอียด ดังนี้

- คุณลักษณะ (Character) เป็นการพิจารณาเกี่ยวกับประวัติพฤติกรรม อุปนิสัย ความรับผิดชอบ และความเต็มใจที่จะชำระหนี้ของผู้ขอสินเชื่อ เช่น อายุ การศึกษา หน้าที่การงาน ความตั้งใจชำระหนี้ เป็นต้น
- เงินทุน (Capital) เป็นการพิจารณาเกี่ยวกับฐานะการเงินของผู้ขอสินเชื่อ ได้แก่ ทรัพย์สิน หนี้สิน และเงินทุน ซึ่งใช้ในการสนับสนุนความสามารถในการชำระหนี้
- หลักประกัน (Collateral) เป็นการลดความเสี่ยงให้กับการปล่อยสินเชื่อ หากมีการผิดนัดชำระหนี้สามารถนำหลักประกันมาชำระหนี้แทนได้ ต้องพิจารณาถึงสภาพคล่องและทำเลของหลักประกันด้วย โดยหลักประกันมีทั้งสังหาริมทรัพย์ เช่น เงินฝาก หุ่น พันธบัตร กองทุน และอสังหาริมทรัพย์ เช่น ที่ดิน อาคาร สิ่งปลูกสร้าง เป็นต้น
- ความสามารถในการชำระหนี้ (Capacity) เป็นการพิจารณาความสามารถในการชำระหนี้ ได้แก่ รายได้ รายจ่าย ผลกำไร และฐานะทางการเงิน โดยรายได้หลังหักค่าใช้จ่ายต้องเพียงพอต่อการชำระหนี้
- เงื่อนไขการชำระหนี้ (Condition) เป็นเงื่อนไขในการใช้วงเงินที่กำหนดขึ้นสำหรับผู้ขอสินเชื่อแต่ละราย ให้ปฏิบัติตามเพื่อความเหมาะสม เช่น ระยะเวลากู้ยืม การเบิกใช้วงเงิน อัตราดอกเบี้ย เป็นต้น

#### 2.1.2 XGBoost Model

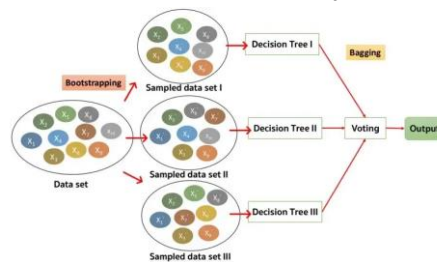
เป็นการเรียนรู้เครื่องจักร ที่ใช้อัลกอริทึมแบบ Boosting ซึ่งสามารถดำเนินการประมวลผลเป็นลำดับแบบขนาน multithreaded parallel โดยแบ่งข้อมูลเป็น train-test แล้วแก้ไขข้อผิดพลาด (errors) และทำซ้ำๆ จนมีความแม่นยำมากที่สุด และสามารถลด overfit ได้อีกด้วย เนื่องจาก model มีการทำ Regularization, Sparse Aware (จัดการ missing value), Parallelization (ทำงานแบบขนานกับจำนวน core ของ CPU) และ Cache Optimization (การใช้ hardware อย่างมีประสิทธิภาพ) (Jun Wang, et al., 2022) ดังรูปภาพที่ 1



รูปภาพที่ 1 XGBoost Model

### 2.1.3 Random Forest Model

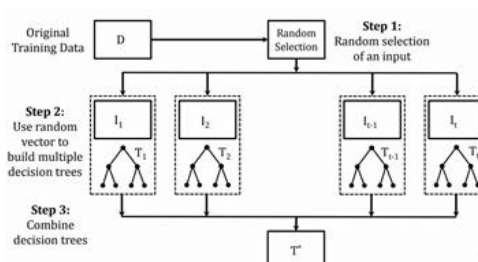
เป็น Ensemble learning โดยการแบ่ง Class ข้อมูลตัวแปร แล้วทำการเรียนรู้หลายๆครั้ง บนชุดข้อมูลเดียวกัน โดยแต่ละครั้งจะเลือกข้อมูลไม่เหมือนกัน เพื่อให้ model ตัดสินใจว่า Class ไหนถูกเลือกและทำนายได้แม่นยำมากที่สุด โดยประโยชน์ของ Random Forest ได้แก่ ทำงานอย่างมีประสิทธิภาพบนฐานข้อมูลขนาดใหญ่ และมีความแม่นยำ สามารถแก้ไขได้ดีกับข้อมูลที่มีความไม่สมดุลของประชากร และสามารถจัดการกับ missing value เพื่อทำนายได้อย่างแม่นยำอีกด้วย (Lin Zhu, 2019) ดังรูปภาพที่ 2



รูปภาพที่ 2 Random Forest Model

### 2.1.4 Extra Tree Model

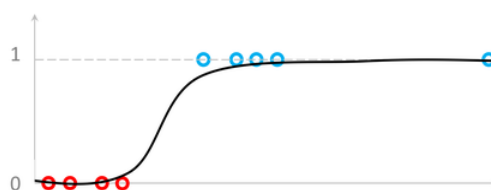
เป็นการเรียนรู้คล้าย random forest ต่างกันที่การเลือกจุด cut-off ในการแบ่งชุดข้อมูลเพื่อ train model ซึ่ง random forest จะใช้ optimum split ส่วน Extra tree จะเลือกแบบสุ่ม (random) แต่ยังคง optimization ซึ่งมีความเร็วในการ train-test model ที่เร็วกว่า random forest (Netty Setiawan, Suharjito and Diana, 2019) ดังรูปภาพที่ 3



รูปภาพที่ 3 Extra Tree Model

### 2.1.5 Logistic Regression Model

เป็นโมเดลที่สร้างขึ้นเพื่อแก้ปัญหการอธิบายการ Classification ของ Linear Regression ซึ่งใช้ Logit function ในการคำนวณ ให้ output ออกมาเป็นค่า 0, 1 เพื่อง่ายต่อการอธิบาย (Ligang Zhou, et al., 2021) ดังรูปภาพที่ 4



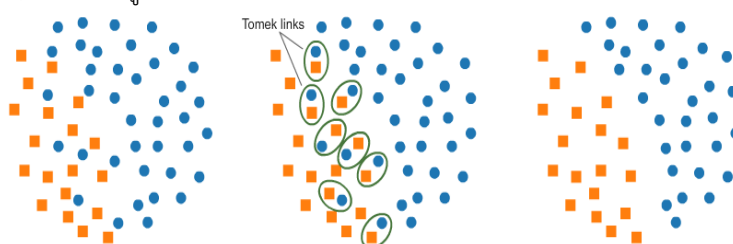
รูปภาพที่ 4 Logistic Regression Model

## 2.1.6 Imbalance

เป็นเทคนิคในการจัดการข้อมูลที่มีจำนวนแตกต่างกันมากในแต่ละกลุ่ม ทำให้การพยากรณ์มีความคลาดเคลื่อนหรือเอนเอียงไปทางกลุ่มที่มีจำนวนมากกว่า ซึ่งเทคนิคดังกล่าว จะทำการลดจำนวนข้อมูลสำหรับกลุ่มที่มีจำนวนมากกว่า ลงไปให้ใกล้เคียงกับกลุ่มที่มีจำนวนน้อยกว่า หรือเพิ่มจำนวนข้อมูลสำหรับกลุ่มที่มีจำนวนน้อยกว่า ให้ใกล้เคียงกับกลุ่มที่มีจำนวนมากกว่า

### 2.1.6.1 Tomek Link

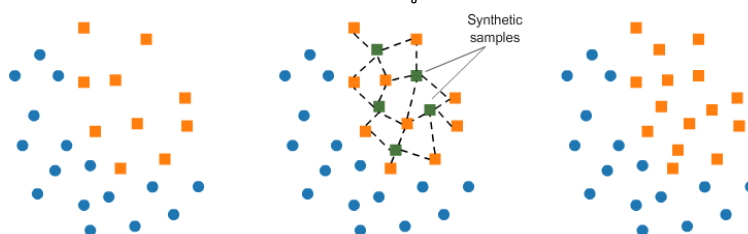
เป็นวิธีการสุ่มตัวอย่างแบบ Under Sampling ที่พัฒนาโดย Tomek เป็นการปรับปรุง Nearest-Neighbor Rule (NNR) คือการลบข้อมูลที่อยู่ในพื้นที่ที่ทับซ้อนกันระหว่าง Class ซึ่ง Class ที่มีจำนวนข้อมูลที่มากกว่าจะถูกลบออก ไม่ให้ข้อมูลถูกทับซ้อนกัน เพื่อให้ง่ายต่อการ Classifier (Elhassan, et al., 2016) ดังรูปภาพที่ 5



รูปภาพที่ 5 Tomek Links

### 2.1.6.2 SMOTE (Synthetic Minority Oversampling Technique)

เป็นเทคนิคการสุ่มตัวอย่างแบบ Over Sampling เพื่อเพิ่มข้อมูลให้กับ Class ที่มีจำนวนที่น้อยกว่า โดยการสร้างตัวอย่างเทียมขึ้นมา เพื่อหลีกเลี่ยงความไม่สมดุลของข้อมูลในแต่ละ Class เพื่อให้ง่ายต่อการ Classifier (Elhassan, et al., 2016) ดังรูปภาพที่ 6



รูปภาพที่ 6 SMOTE

## 2.2 งานวิจัยที่เกี่ยวข้อง

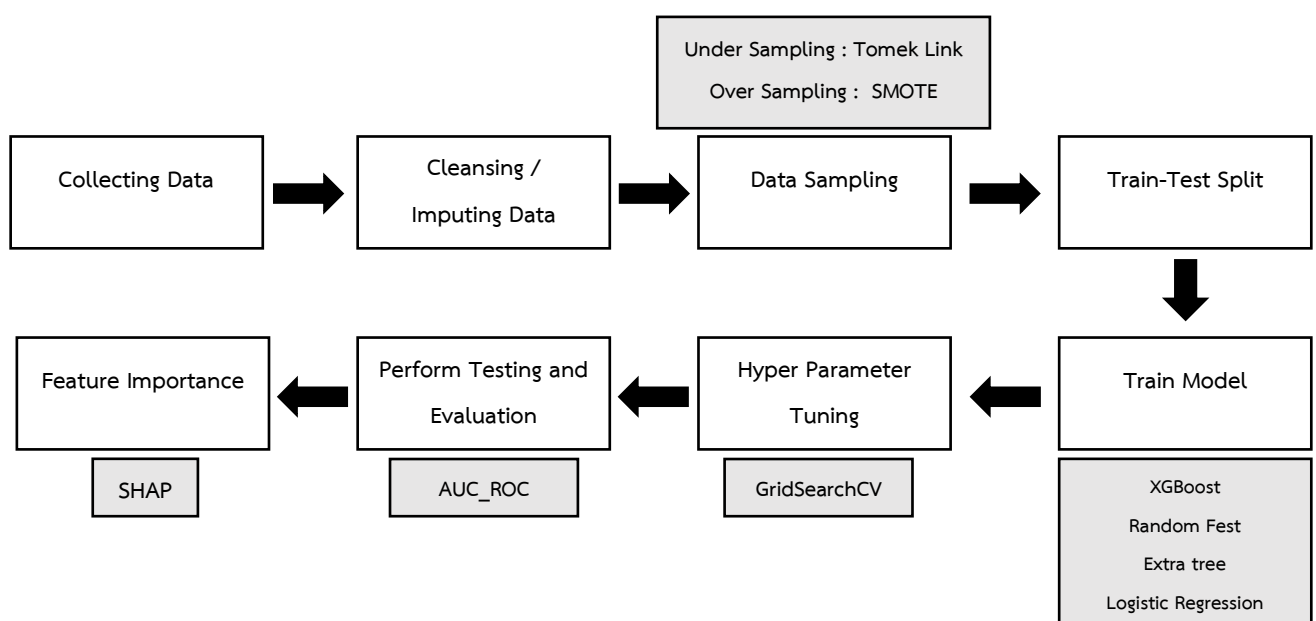
ตารางที่ 1 งานวิจัยที่เกี่ยวข้อง

Papers	Author	Year	Compared Model	Best Model	AUC	ACCURACY
Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting	Ligang Zhou , Hamido Fujita , Hao Ding , Rui Ma	2021	Logistic Regression, Random Forest, Gradient Boosting	Gradient Boosting	83.57%	77.93%
A Comparison of Prediction Methods for Credit Default on Peer to Peer Lending using Machine Learning	Netty Setiawana, Suharjitoa, Dianab	2019	Extra Tree	Extra Tree	-	64%

Papers	Author	Year	Compared Model	Best Model	AUC	ACCURACY
Risk assessment in social lending via random forests	Milad Malekipirbazari, Vural Aksakalli	2015	Random Forest, Neighbor Nearest, Support Vector Machine, Logistic Regression	Random Forest	71%	78%
Determinants of loan funded successful in online P2P Lending	Yuejin Zhanga, Haifeng Lia, Mo Hai, Jiaxuan Lia, Aihua Li	2017	Logistic Regression	Logistic Regression	-	94%
Loan Default Prediction Using Genetic Algorithm: A Study Within Peer-To-Peer Lending Communities	Lee Victor, Mafas Raheem	2021	Logistic Regression, Random Forest, Support Vector Machine	Random Forest	89%	92%
This Paper	Panapat Prasanboonlert	2022	XGBoost, Extra Tree, Random Forest, Logistic Regression	XGBoost	81.20%	73.01%

จากการทบทวนวรรณกรรมพบว่า นักวิจัยหลายท่านยังให้ความสนใจ และดำเนินการศึกษาการผิติดชำระหนี้จากการกู้ยืมเงินระหว่างบุคคลทั่วไปผ่านแพลตฟอร์มออนไลน์ และได้ใช้แบบจำลองต่างๆ ในการช่วยพยากรณ์ ได้แก่ Gradient Boosting (Ligang Zhou, et al., 2021) Extra Tree (Netty Setiawan, Suharjito and Diana, 2019) Logistic Regression, Random Forest และ Linear SVM (Lee Victor and Mafas Raheem, 2021) โดยได้นำตัวแปรอิสระมาศึกษา และแบ่งออกเป็น 3 กลุ่ม ได้แก่ ข้อมูลสินเชื่อ ข้อมูลลักษณะประชากร และประวัติสินเชื่อของผู้กู้ ผลการวิจัยพบว่า ตัวแปรที่มีอิทธิพลต่อการผิติดชำระหนี้มากที่สุด คือ จำนวนเงินที่ต้องชำระต่อเดือน (Ligang Zhou, et al., 2021) อัตราดอกเบี้ยต่อปี ระยะเวลาชำระคืน เพศ คะแนนเครดิต (Netty Setiawan, Suharjito and Diana, 2019) และจำนวนเงินที่ชำระหนี้ทั้งหมด (Lee Victor and Mafas Raheem, 2021)

### 3. วิธีดำเนินการวิจัย



รูปภาพที่ 7 กรอบแนวคิดวิธีดำเนินการวิจัย

### 3.1 แหล่งที่มาของข้อมูล

เลือกใช้ข้อมูลผู้กู้จากเว็บไซต์ <https://www.bondora.com> ซึ่งเป็นข้อมูลที่เปิดเผยต่อสาธารณะ โดยเก็บข้อมูลของผู้กู้ที่เข้ามาใช้บริการตั้งแต่เดือนกุมภาพันธ์ 2009 – ธันวาคม 2022 มีจำนวนผู้กู้ทั้งหมด 178,277 คน และมีจำนวนคอลัมน์ที่เป็นคุณลักษณะของผู้กู้ จำนวน 112 คอลัมน์ ได้แก่ อายุ เพศ การศึกษา สถานภาพสมรส วงเงินกู้ อัตราดอกเบี้ยเงินกู้ ระยะเวลากู้ เป็นต้น และข้อมูลปัจจัยด้านเศรษฐกิจจากเว็บไซต์ <https://sdw.ecb.europa.eu> (เว็บไซต์ธนาคารกลางยุโรป) ซึ่งเป็นข้อมูลที่เปิดเผยต่อสาธารณะ จำนวน 4 คอลัมน์ ได้แก่ อัตราเงินเฟ้อ อัตราดอกเบี้ยกู้ยืม อัตราการว่างงาน และผลิตภัณฑ์มวลรวมของประเทศ (GDP ทั้งสหภาพยุโรป)

### 3.2 การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ทำการเลือกข้อมูลผู้กู้เพื่อนำมาศึกษาสำหรับการพยากรณ์การผิดนัดชำระหนี้จากการกู้ยืมเงินระหว่างบุคคลทั่วไปผ่านแพลตฟอร์มออนไลน์ โดยเลือกตัวแปรที่เกี่ยวข้องกับทฤษฎีการวิเคราะห์สินเชื่อ (5Cs) และงานวิจัยที่เกี่ยวข้องมีทั้งหมด 19 ตัวแปร โดยแสดงรายละเอียดในตารางที่ 2

ตารางที่ 2 รายละเอียดตัวแปร

Variable	Description
Age	The age of the borrower when signing the loan application
Gender	0 Male 1 Woman 2 Undefined
EmploymentDurationCurrentEmployer	Employment time with the current employer
LiabilitiesTotal	Total monthly liabilities
DebtToIncome	Ratio of borrower's monthly gross income that goes toward paying loans
IncomeTotal	Borrower's total income
AmountOfPreviousLoansBeforeLoan	Value of previous loans
ProbabilityOfDefault	Probability of Default, refers to a loan's probability of default within one year horizon.
Rating	Bondora Rating issued by the Rating model
Status	The current status of the loan application
AppliedAmount	The amount borrower applied for originally
Interest	Maximum interest rate accepted in the loan application
LoanDuration	Current loan duration in months
NewCreditCustomer	Did the customer have prior credit history in Bondora 0 Customer had at least 3 months of credit history in Bondora 1 No prior credit history in Bondora
MonthlyPayment	Estimated amount the borrower has to pay every month
unemployment_rate	the percentage of people in the labour force that are unemployed
gdp_growth_rate	The annual average rate of change of the gross domestic product (GDP) at market prices based on constant local currency, for a given national economy, during a specified period of time.
inflation_rate	The rate of increase in prices over a given period of time.
interest_rate	The amount you are charged for borrowing money – a percentage of the total amount of the loan.



### 3.3 การสร้างแบบจำลองการพยากรณ์

#### 3.3.1 การทำความสะอาดข้อมูล

ผู้วิจัยได้ทำการลบ Values ที่เป็นค่า Null เนื่องจากข้อมูลมีจำนวนมาก ทำการ encoding ข้อมูล categorical ให้เป็น numerical ได้แก่ EmploymentDurationCurrentEmployer และ Rating จากนั้น Impute missing value โดยแทนที่ด้วยค่า Median และทำการแบ่งข้อมูลเป็นตัวแปรอิสระ (X) จำนวน 18 ตัวแปร และกลุ่มเป้าหมาย (Y) โดยให้ค่า Repaid เป็น 0 มีจำนวนข้อมูลผู้กู้ 97,849 คน และให้ค่า Late เป็น 1 มีจำนวนข้อมูลผู้กู้ 80,428 คน รวมทั้งหมด 178,277 คน ดังรูปภาพที่ 8 และตารางที่ 3



รูปภาพที่ 8 Exploratory Data Analysis

ตารางที่ 3 Exploratory Data Analysis

Attribute	count	mean	std	Min	0.25	0.5	0.75	max
EmploymentDurationCurrentEmployer	178,277	14.06	3.70	10.00	10.00	15.00	19.00	20.00
Rating	178,277	4.87	2.85	-	3.00	5.00	6.00	20.00
NewCreditCustomer	178,277	0.59	0.49	-	-	1.00	1.00	1.00
AppliedAmount	178,277	2,673.31	2,306.26	10.00	726.00	2,125.00	4,150.00	15,948.00
LoanDuration	178,277	47.04	17.18	1.00	36.00	60.00	60.00	120.00
Interest	178,277	33.85	23.73	2.00	19.90	28.60	39.41	264.31
MonthlyPayment	178,277	112.00	106.46	-	37.23	90.14	142.02	2,368.54
Age	178,277	39.54	12.27	-	30.00	38.00	48.00	77.00
DebtToIncome	178,277	5.68	14.30	-	-	-	-	198.02
LiabilitiesTotal	178,277	494.57	29,418.33	-	40.27	269.00	559.00	12,400,000.00
Gender	178,277	0.47	0.62	-	-	-	1.00	2.00

Attribute	count	mean	std	Min	0.25	0.5	0.75	max
IncomeTotal	178,277	1,905.83	9,479.76	-	904.00	1,300.00	2,000.00	1,012,019.00
AmountOfPreviousLoansBeforeLoan	178,277	2,885.62	4,807.06	-	-	70.30	4,150.00	119,983.00
ProbabilityOfDefault	178,277	0.22	0.14	-	0.11	0.19	0.31	0.99
interest_rate	178,277	0.05	0.00	0.04	0.05	0.05	0.06	0.08
inflation_rate	178,277	0.02	0.02	-0.01	0.01	0.01	0.02	0.11
gdp_growth_rate	178,277	0.02	0.04	-0.14	0.01	0.02	0.03	0.14
unemployment_rate	178,277	0.08	0.01	0.07	0.07	0.08	0.09	0.12

### 3.3.2 Sampling Method

ผู้วิจัยได้ทำการสุ่มข้อมูลให้มีขนาดข้อมูลเท่ากัน เนื่องจากข้อมูลตัวแปรสถานะลูกหนี้ (Status) มีความไม่สมดุล (ค่า 0 มีจำนวน 97,849 คน และค่า 1 มีจำนวน 80,428 คน) โดยใช้เทคนิค resampling 2 วิธี ดังนี้

1. Under Sampling โดยใช้ Tomek Link เพื่อลดจำนวนกลุ่มข้อมูลที่มีจำนวนมากกว่า (ค่า 0) ลงมาให้เท่ากับกลุ่มข้อมูลที่มีจำนวนน้อยกว่า (ค่า 1) ซึ่งจะได้ข้อมูลเท่ากันทั้ง 2 กลุ่ม เท่ากับ 79,257 คน
2. Over Sampling โดยใช้ SMOTE เพื่อเพิ่มจำนวนกลุ่มข้อมูลที่มีจำนวนน้อยกว่า (ค่า 1) ขึ้นมาให้เท่ากับกลุ่มข้อมูลที่มีจำนวนมากกว่า (ค่า 0) ซึ่งจะได้ข้อมูลเท่ากันทั้ง 2 กลุ่ม เท่ากับ 97,849 คน

### 3.3.3 Hyper Parameter Tuning : GridsearchCV

ผู้วิจัยได้ทำการใช้วิธี Cross Validation เพื่อแบ่งชุดข้อมูลในการ train model เพื่อแก้ไขการเลือกข้อมูลมา train แล้วเกิด error ขึ้น และทำการ tuning พารามิเตอร์ที่ต้องการ เพื่อให้ model มีความแม่นยำที่สุด ดังตารางที่ 4 ตารางที่ 4 Hyper Parameter Tuning : GridsearchCV

Classifier	Parameters
XGB Classifier	colsample_bytree: [ 0.3, 0.5 , 0.8 ], reg_alpha: [0, 0.5, 1, 5], reg_lambda: [0, 0.5, 1, 5]
Random Forest Classifier	n_estimators: [100, 500, 1000], max_depth: [1, 5, 10, 25], max_features: [*np.arange(0.1, 1.1, 0.1)]
Extra Trees Classifier	random_state: [0, 1, 2, 3, 4], n_estimators: [320, 340, 360, 380, 400], max_depth: [25, 30, 32, 34, 38, 45]
Logistic Regression	max_iter: [2000], penalty : ['l1', 'l2'], C : np.logspace(-4, 4, 20), solver : ['liblinear']

### 3.3.4 Evaluation

ผู้วิจัยได้ประเมินผลแบบจำลองที่ใช้ในการพยากรณ์การผิดนัดชำระหนี้ โดยใช้ค่า Accuracy, ROC Curve, Precision, Recall, F1-score และพิจารณาปัจจัยที่มีผลต่อการผิดนัดชำระหนี้ ด้วยเทคนิค SHAP Feature Importance (SHapley Additive exPlanations) เป็น Library ที่สามารถอธิบายปัจจัยที่มีผลต่อ Model ได้ดียิ่งขึ้น โดยมี SHAP Values ที่เปรียบเสมือนค่า Based เพื่อรับ Effect จาก Features อื่น ๆ (คล้ายกับ Intercept หรือ Constant)

## 3.4 Experimental Setup

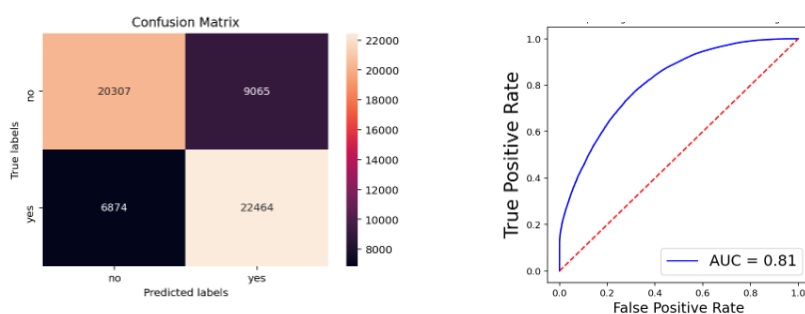
ผู้วิจัยทำการทดลองโดยใช้โปรแกรม Jupyter Notebook ในการเตรียมข้อมูล บนเครื่องคอมพิวเตอร์ Lenovo IdeaPad L340-15IRH Gaming : Intel(R) Core(TM) i5-9300H CPU 2.40 GHz, RAM 8.00 GB

#### 4. ผลการดำเนินงานวิจัย

ตารางที่ 5 ผลการดำเนินงานวิจัย

	Classifier	No GridsearchCV					GridsearchCV				
		AUC	ACCURACY	PRECISION	RECALL	F1-SCORE	AUC	ACCURACY	PRECISION	RECALL	F1-SCORE
Imbalance	XGBoost	77.64%	70.31%	66.24%	69.70%	67.93%	77.94%	70.63%	66.58%	70.07%	68.28%
	Random Forest	77.06%	69.91%	66.68%	66.57%	66.62%	77.70%	70.67%	66.96%	69.03%	67.98%
	Extra Tree	75.28%	68.49%	65.54%	63.56%	64.53%	76.76%	69.71%	66.19%	67.16%	66.67%
	Logistic Regression	67.51%	62.50%	59.16%	54.50%	56.73%	70.19%	64.28%	62.00%	53.74%	57.57%
Under Sampling	XGBoost	80.23%	72.89%	70.99%	78.78%	74.68%	80.27%	72.93%	70.99%	78.90%	74.74%
	Random Forest	79.71%	72.49%	71.22%	76.86%	73.93%	80.31%	72.89%	71.04%	78.65%	74.65%
	Extra Tree	78.16%	71.13%	70.41%	74.37%	72.34%	79.23%	71.98%	70.03%	78.32%	73.94%
	Logistic Regression	70.41%	65.05%	64.02%	71.09%	67.37%	74.03%	67.67%	66.71%	72.47%	69.47%
Over Sampling	XGBoost	81.05%	72.80%	71.24%	76.41%	73.74%	81.20%	73.01%	71.25%	76.57%	73.81%
	Random Forest	80.19%	72.42%	71.35%	74.86%	73.06%	80.49%	72.75%	71.12%	76.57%	73.74%
	Extra Tree	78.99%	71.39%	70.78%	72.79%	71.77%	79.88%	72.24%	70.22%	77.20%	73.54%
	Logistic Regression	68.54%	63.56%	62.09%	69.53%	65.60%	71.80%	65.92%	64.64%	70.19%	67.30%

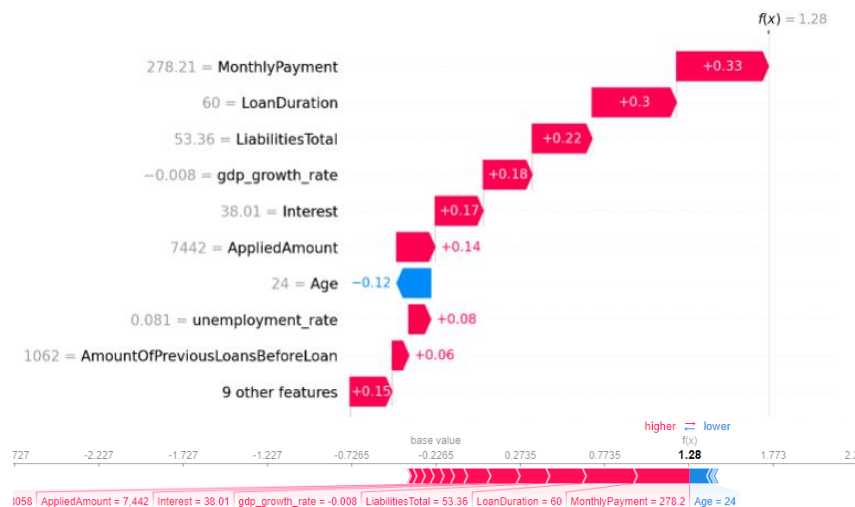
จากผลการดำเนินการวิจัย พบว่า แบบจำลองที่สามารถพยากรณ์การผิดนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปได้ดีที่สุด คือ XGBoost Model ที่ผ่านการ Oversampling : SMOTE โดยมีค่า AUC = 81.20% (สามารถทำนายการผิดนัดชำระหนี้ได้ดี 81.20%) Accuracy = 73.01% (สามารถทำนายการผิดนัดชำระหนี้ได้อย่างแม่นยำ 73.01%) Precision = 71.25%, Recall = 76.57% และ F1-Score = 73.81% ส่วนค่า False Negative มีค่า 11.71% หมายความว่า แบบจำลองทำนายผู้กู้ที่ไม่มีแนวโน้มผิดนัดชำระหนี้ แต่จริงๆ ผู้กู้ผิดนัดชำระหนี้ มีจำนวน 11.71% หรือ 6,874 คน ดังรูปภาพที่ 9



รูปภาพที่ 9 Confusion Matrix และ AUC Curve

ส่วนปัจจัยที่มีผลต่อการผิดนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปมากที่สุด คือ MonthlyPayment (จำนวนเงินที่จะต้องผ่อนชำระต่องวด) มีค่า Trigger point เท่ากับ 278.21 ยูโร กล่าวคือ หากผู้กู้ได้รับข้อเสนอจำนวนเงินที่จะต้องผ่อนชำระต่องวด มากกว่าหรือเท่ากับ 278.21 ยูโร มีโอกาสที่จะเกิดการผิดนัดชำระหนี้ รองลงมาคือ LoanDuration (ระยะเวลากู้ยืม) มีค่า Trigger point เท่ากับ 60 เดือน กล่าวคือ หากผู้กู้ได้รับข้อเสนอระยะเวลา กู้ยืมมากกว่าหรือเท่ากับ 60 เดือน มีโอกาสที่จะเกิดการผิดนัดชำระหนี้ ส่วนปัจจัยที่มีความสัมพันธ์ในทิศทางตรงกันข้ามกับ

การผิณฑชำระหนี้ คือ Age (อายุ) มีค่า Trigger point เท่ากับ 24 ปี กล่าวคือ หากผู้กู้มีอายุน้อยกว่าหรือเท่ากับ 24 ปี มีโอกาสที่จะเกิดการผิณฑชำระหนี้ ดังรูปภาพที่ 10



รูปภาพที่ 10 SHAP Feature Importance

ซึ่งสอดคล้องกับงานวิจัยของ Ligang Zhou, Hamido Fujita, Hao Ding and Rui Ma (2021) ที่ได้ศึกษาเกี่ยวกับการพัฒนาแบบจำลองเพื่อพยากรณ์การผิณฑชำระหนี้จากการกู้ยืมระหว่างบุคคล (peer-to-peer lending) พบว่า ปัจจัยที่ทำให้เกิดการผิณฑชำระหนี้มากที่สุด ได้แก่ MonthlyPayment (จำนวนเงินที่จะต้องผ่อนชำระต่องวด) รองลงมา คือ Loan Duration (ระยะเวลากู้ยืม) ยิ่ง 2 ปัจจัยนี้มีค่ามาก ยิ่งทำให้เกิดการผิณฑชำระหนี้

ส่วนปัจจัยทางเศรษฐกิจที่ส่งผลกระทบต่อ การผิณฑชำระหนี้ คือ gdp\_growth\_rate และ unemployment สอดคล้องกับงานวิจัยของ Maryem Naili และ Younes Lahrichi (2022) ที่ได้ศึกษาปัจจัยที่เกี่ยวกับการผิณฑชำระหนี้ของธนาคาร ซึ่งผลการศึกษาพบว่า GDP และอัตราการว่างงาน ส่งผลต่อการผิณฑชำระหนี้ของธนาคาร เนื่องจากภาวะเศรษฐกิจดีส่งผลให้เกิดเงินเฟ้อ ซึ่งกระทบต่อภาคการเงินของบุคคล และเมื่ออัตราการว่างงานสูงขึ้น ส่งผลต่อการผิณฑชำระหนี้ของธนาคาร เช่นกัน

## 5. สรุปผลการวิจัย

ในการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อพยากรณ์การผิณฑชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป และปัจจัยที่มีผลต่อการผิณฑชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป โดยเลือกใช้ข้อมูลผู้กู้จากเว็บไซต์ <https://www.bondora.com> ซึ่งเป็นข้อมูลที่เปิดเผยต่อสาธารณะ โดยเก็บข้อมูลของผู้กู้ที่เข้ามาใช้บริการตั้งแต่เดือนกุมภาพันธ์ 2009 – ธันวาคม 2022 มีจำนวนผู้กู้ทั้งหมด 178,277 คน

### 5.1 อภิปรายผลการวิจัย

5.1.1 จากผลการดำเนินการวิจัย พบว่า แบบจำลองที่สามารถพยากรณ์การผิณฑชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปได้ดีที่สุด คือ XGBoost Model ที่ผ่านการ Oversampling : SMOTE โดยมีค่า AUC = 81.20% (สามารถทำนายการผิณฑชำระหนี้ได้ดี 81.20%) Accuracy = 73.01% (สามารถทำนายการผิณฑชำระหนี้ได้อย่างแม่นยำ 73.01%) Precision = 71.25%, Recall = 76.57% และ F1-Score = 73.81%

- 5.1.2 ปัจจัยที่มีผลต่อการผิมนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปมากที่สุด 5 อันดับแรก คือ จำนวนเงินที่จะต้องผ่อนชำระต่องวด ระยะเวลากู้ยืม หนี้สินรวม ผลิตภัณฑ์มวลรวมของประเทศ (GDP ทั้งสหภาพยุโรป) และอัตราดอกเบี้ยเงินกู้ยืม ตามลำดับ

## 5.2 ปัญหาและอุปสรรคของการวิจัย

- 5.2.1 ตัวแปรอิสระอาจไม่สามารถอธิบายได้กับลูกหนี้ใหม่เนื่องจากการเริ่มกู้ยืมใหม่ อาจยังไม่ผิมนัดชำระหนี้ในเดือนแรกๆ
- 5.2.2 งานวิจัยนี้เป็นการศึกษาการผิมนัดชำระหนี้ประเภทสินเชื่อส่วนบุคคล อาจจะไม่สะท้อนกับสินเชื่อประเภทอื่นๆ เช่น สินเชื่อบ้าน สินเชื่อรถ เป็นต้น เนื่องจากเป็นสินเชื่อที่ไม่มีหลักประกัน ผู้ที่จะศึกษาควรพิจารณาเพื่อนำไปใช้ในสินเชื่อประเภทอื่นๆ

## 5.3 ประโยชน์ที่ได้จากการวิจัย

- 5.3.1 ได้แบบจำลองที่สามารถพยากรณ์การผิมนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปได้ถูกต้องแม่นยำที่สุด
- 5.3.2 สามารถอธิบายให้ทราบถึงปัจจัยที่มีผลต่อการผิมนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป
- 5.3.3 เพื่อเพิ่มเติมและขยายองค์ความรู้ รวมถึงผลการวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองเพื่อพยากรณ์การผิมนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรป

## 5.4 ข้อเสนอแนะจากการวิจัย

- 5.4.1 สามารถใช้ข้อมูลปัจจัยอื่นๆ เช่น ประเภทอาชีพ, สถานภาพสมรส, พื้นที่ที่อยู่อาศัย, ระยะทางจากที่พักอาศัยไปทำงาน เข้ามาช่วยในการพัฒนาแบบจำลองให้มีความแม่นยำยิ่งขึ้น
- 5.4.2 ผู้ศึกษาสามารถใช้แบบจำลองอื่นๆ หรือ Deep Learning เพื่อนำมาใช้พยากรณ์การผิมนัดชำระหนี้จากการกู้ยืมระหว่างบุคคลผ่านแพลตฟอร์มออนไลน์ในสหภาพยุโรปให้แม่นยำยิ่งขึ้น
- 5.4.3 ศึกษาสามารถใช้ข้อมูลปัจจัยอื่นๆ เช่น Social Network หรือแหล่งข้อมูลที่เกี่ยวข้อง เข้ามาช่วยในการพัฒนาแบบจำลองให้มีความแม่นยำยิ่งขึ้น

## 6. เอกสารอ้างอิง

เนตรชนก สุริยวโรภาส (2561). ปัจจัยที่มีผลต่อการพิจารณาสินเชื่อ SME ของธนาคารกรุงศรีอยุธยา จำกัด (มหาชน).

การค้นคว้าอิสระปริญามหาบัณฑิต, มหาวิทยาลัยรามคำแหง.

Bondora (2022). What is the history of Bondora? <https://support.bondora.com/en/what-is-the-history-of-Bondora>.

Elhassan, et al. (2016). Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method. Global Journal of Technology & Optimization.

European Central Bank (2022). Statistical Data Warehouse. <https://sdw.ecb.europa.eu/>.

Jun Wang, et al. (2022). Credit Debt Default Risk Assessment Based on the XGBoost Algorithm: An Empirical Study from China. Nawab Muhammad Faseeh Qureshi.

- Lee Victor and Mafas Raheem (2021). Loan Default Prediction Using Genetic Algorithm: A Study Within Peer-To-Peer Lending Communities. *International Journal of Innovative Science and Research Technology*, Volume 6, Issue 3, March – 2021.
- Ligang Zhou, et al. (2021). Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting. *Applied Soft Computing* 110.
- Lin Zhu, et al. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162, 503-513.
- Milad Malekipirbazari and Vural Aksakalli (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42, 4621–4631.
- Netty Setiawan, Suharjito and Diana (2019). A Comparison of Prediction Methods for Credit Default on Peer to Peer Lending using Machine Learning. *Procedia Computer Science*, 157, 38–45.
- Yuejin Zhang, et al. (2017). Determinants of loan funded successful in online P2P Lending. *Procedia Computer Science*, 122, 896–901.