

การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล

นายนิธิคุณ สุนทรานุรักษ์

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาระบบสารสนเทศเพื่อการจัดการ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2568
ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล

นายนิธิคุณ สุนทรานุรักษ์

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาระบบสารสนเทศเพื่อการจัดการ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2568

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



ใบรับรองการค้นคว้าอิสระ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล

โดย นายนิธิคุณ สุนทรานุรักษ์

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบสารสนเทศเพื่อการจัดการ

คณบดีคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

(ผู้ช่วยศาสตราจารย์ ดร.สุนันฑา สดสี)

คณะกรรมการสอบการค้นคว้าอิสระ

ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.เดช ธรรมศิริ)

อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล)

กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.มณเฑียร รัตนศิริวงศ์วุฒิ)

ชื่อ : นายนิธิคุณ สุนทรานุรักษ์
ชื่อการค้นคว้าอิสระ : การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล
สาขาวิชา : ระบบสารสนเทศเพื่อการจัดการ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก : ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล
ปีการศึกษา : 2568

บทคัดย่อ

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลในการพยากรณ์ผู้กู้สินเชื่อบุคคลที่มีแนวโน้มผิดนัดชำระหนี้ โดยชุดข้อมูลที่ใช้ในการวิจัยประกอบด้วยข้อมูลจำนวน 20,000 ระเบียน และมีตัวแปรทั้งหมด 11 ตัวแปร ข้อมูลทั้งหมดได้รับการทำความสะอาด การแปลงรูปแบบข้อมูลและการปรับสมดุลข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกินจำนวนกลุ่มน้อยสังเคราะห์ ข้อมูลครอบคลุมรายละเอียดสินเชื่อบุคคลในช่วงเดือนมกราคม 2565 ถึงเดือนกรกฎาคม 2566 และมีการติดตามพฤติกรรมการชำระหนี้เพิ่มเติมอีก 18 เดือน จนถึงเดือนกุมภาพันธ์ 2568 ผู้กู้ถูกจำแนกออกเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ไม่ผิดนัดชำระหนี้ และกลุ่มที่ผิดนัดชำระหนี้ งานวิจัยนี้ได้ทำการเปรียบเทียบอัลกอริทึม 3 วิธี ได้แก่ โครงข่ายประสาทเทียมกราฟคอนโวลูชัน เอ็กตรีมกราเดี้ยนบูทติ้ง และ โมเดลแบบผสมผสานระหว่างโครงข่ายประสาทเทียมกราฟคอนโวลูชันและเอ็กตรีมกราเดี้ยนบูทติ้ง ประสิทธิภาพของโมเดลถูกประเมินด้วยตัวชี้วัด ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก และ ค่าการถ่วงดุล ผลการศึกษาพบว่า แบบจำลองเอ็กซ์ตรีมกราเดียนต์บูสติ้ง (XGBoost) มีประสิทธิภาพดีที่สุดทั้งในขั้นตอนการฝึกแบบจำลองและการทดสอบ โดยในขั้นตอนการฝึกแบบจำลองสามารถให้ค่าความถูกต้องเท่ากับ 92.44% ค่าความแม่นยำเท่ากับ 0.93 และมีค่าระลึกและค่าการถ่วงดุลเท่ากับ 0.92 นอกจากนี้เมื่อทดสอบแบบจำลองด้วยข้อมูลที่ไม่เคยใช้ในการฝึกแบบจำลองมาก่อน พบว่าสามารถให้ค่าความถูกต้องเท่ากับ 86.8% และมีค่าความแม่นยำ ค่าระลึก และค่าการถ่วงดุล เท่ากับ 0.87

(การค้นคว้าอิสระมีจำนวนทั้งสิ้น 89 หน้า)

คำสำคัญ : โครงข่ายประสาทเทียมกราฟคอนโวลูชัน เอ็กตรีมกราเดี้ยนบูทติ้ง การผิดนัดชำระหนี้สินเชื่อบุคคล

_____ อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก

Name : Mr. Nithikun Suntaranurak
Independent Study Title : Prediction of Personal Loan Default using Data Mining
Major Field : Management Information Systems
King Mongkut's University of Technology North Bangkok
Independent Study Advisor : Assistant Professor Dr. Watchareewan Jitsakul
Academic Year : 2025

ABSTRACT

This research compare the performance of data mining techniques in predicting personal loan borrowers who are likely to default. The dataset used in this study consists of 20,000 records with a total of 11 features. All data were subjected to data cleaning, data transformation, and class balancing using the SMOTE technique. The dataset covers personal loan information from January 2022 to July 2023, with repayment behavior monitored for an additional 18 months through February 2025. Borrowers were classified into two groups: non-delinquent and delinquent. This research compares three algorithms: Graph Convolutional Network (GCN), Extreme Gradient Boosting (XGBoost), and a Hybrid model. Model performance was evaluated using accuracy, precision, recall, and F1-score as performance metrics. The results indicate that the Extreme Gradient Boosting (XGBoost) model achieved the best performance in both the training and testing phases. During the training phase, the model achieved an accuracy of 92.44%, with a precision of 0.93, and both recall and F1-score of 0.92. Furthermore, when evaluated using unseen data that were not included in the training process, the model achieved an accuracy of 86.8%, with precision, recall, and F1-score values of 0.87

(Total 89 Pages)

Keywords: Graph Convolutional Network (GCN), Extreme Gradient Boosting (XGBoost), Personal Loan Default

_____ Advisor

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยความร่วมมือ และการสนับสนุนจากหลายฝ่ายข้าพเจ้าขอขอบพระคุณอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุล ซึ่งให้คำชี้แนะ และข้อเสนอแนะอันเป็นประโยชน์ต่องานวิจัยนี้เป็นอย่างยิ่ง ตลอดจนประธานสอบ ผู้ช่วยศาสตราจารย์ ดร.เดช ธรรมศิริ และกรรมการสอบค้นคว้าอิสระ ผู้ช่วยศาสตราจารย์ ดร.มณเฑียร รัตนศิริวงศ์วุฒิ ที่ให้ข้อคิดเห็น และข้อเสนอแนะที่มีคุณค่าอันนำไปสู่การพัฒนางานวิจัยให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณฝ่ายบริหารความเสี่ยงเครดิต ธนาคารออมสิน และ คุณธนพร อัสถิ ที่ให้การสนับสนุนในการอำนวยความสะดวกในการเก็บรวบรวมข้อมูล ซึ่งมีส่วนสำคัญต่อความสำเร็จของงานวิจัยฉบับนี้

ข้าพเจ้ายังรู้สึกซาบซึ้งต่อเพื่อนร่วมงานและเพื่อนร่วมชั้นทุกคนที่ช่วยในการสนับสนุนและคอยให้คำปรึกษารวมทั้งความช่วยเหลือตลอดการศึกษาและการค้นคว้าอิสระ ให้กำลังใจตลอดระยะเวลาการทำวิจัย

สุดท้ายนี้ ข้าพเจ้าขอขอบพระคุณครอบครัวที่เป็นกำลังใจตลอดระยะเวลาการทำวิจัย ข้าพเจ้าหวังเป็นอย่างยิ่งว่างานวิจัยฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจศึกษาในสาขาที่เกี่ยวข้อง หากงานวิจัยฉบับนี้มีข้อบกพร่องประการใด ข้าพเจ้ายินดีรับข้อเสนอแนะเพื่อนำไปพัฒนาต่อไป

นิธิคุณ สุนทรานุรักษ์

สารบัญ

หน้า

บทคัดย่อภาษาไทย ข
บทคัดย่อภาษาอังกฤษ ค
กิตติกรรมประกาศ ง
สารบัญตาราง ช
สารบัญภาพ ซ
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญ 1
1.2 วัตถุประสงค์ของการวิจัย 2
1.3 ขอบเขตของการวิจัย 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ 2
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 3
2.1 การพยากรณ์ลูกหนี้ค้างชำระ (Credit Risk Prediction) 3
2.2 โครงข่ายประสาทเทียมกราฟคอนโวลูชัน Graph Convolutional Network (GCN) 4
2.3 เอ็กตรีมกราเดี้ยนบูทติ้ง Extreme Gradient Boosting (XGBoost) 5
2.4 การผสมผสานระหว่างโครงข่ายประสาทเทียมกราฟคอนโวลูชัน และเอ็กตรีมกรเดี้ยนบูทติ้ง (GCN + XGBoost) 5
2.5 งานวิจัยที่เกี่ยวข้อง 7
บทที่ 3 วิธีดำเนินงานวิจัย 21
3.1 การออกแบบกรอบแนวความคิด (Conceptual Framework Design) 21
3.2 ขั้นตอนการรวบรวมข้อมูลและจัดเตรียมข้อมูล (Data Collection and preparation) 23
3.3 การพัฒนา และวัดประสิทธิภาพของแบบจำลอง (Model Development & Evaluation) 29
3.4 การนำแบบจำลองไปใช้งาน (Model Deployment) 33
บทที่ 4 ผลการดำเนินงานวิจัย 35
4.1 ผลการเก็บรวบรวมข้อมูลและการจัดเตรียมข้อมูล 35
4.2 ผลการพัฒนาและทดสอบประสิทธิภาพของแบบจำลอง 38
4.3 ผลการนำแบบจำลองไปใช้งาน 78
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 82
5.1 สรุปผลการวิจัย 82
5.2 การอภิปรายผลการวิจัย 83

สารบัญ (ต่อ)

หน้า

5.3 ข้อเสนอแนะ 85

บรรณานุกรม 86

ภาคผนวก ก 88

ประวัติผู้จัดทำการค้นคว้าอิสระ 89

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

สารบัญตาราง

ตารางที่		หน้า
2-1	งานวิจัยที่เกี่ยวข้อง	14
3-1	การแปลงข้อมูลวงเงินกู้	23
3-2	การแปลงข้อมูลสินเชื่อคงเหลือ	24
3-3	การแปลงข้อมูลรายได้	24
3-4	การแปลงข้อมูลการศึกษา	24
3-5	การแปลงข้อมูลอายุ	25
3-6	การแปลงข้อมูลสถานะสมรส	25
3-7	การแปลงข้อมูลระยะเวลาขอกู้	25
3-8	การแปลงข้อมูลวัตถุประสงค์การกู้	26
3-9	การแปลงข้อมูลอาชีพ	26
3-10	การแปลงข้อมูลรายจ่าย	26
3-11	การแปลงข้อมูลอายุงาน	27
3-12	การแปลงข้อมูลกรรมสิทธิ์ในสถานประกอบการ	27
3-13	การแปลงข้อมูลประเภทที่อยู่อาศัย	27
3-14	การแปลงข้อมูลประเภทผู้กู้	27
3-15	ค่าความสำคัญของตัวแปร	28
3-16	ค่าพารามิเตอร์สำหรับแบบจำลอง GCN	30
3-17	ค่าพารามิเตอร์สำหรับแบบจำลอง XGBoost	35
4-1	สรุปข้อมูลที่ใช้การพัฒนาและทดสอบแบบจำลอง	36
4-2	ตัวอย่างข้อมูลที่ใช้ในการฝึกฝนแบบจำลอง	36
4-3	ค่าพารามิเตอร์สำหรับแบบจำลอง GCN	43
4-4	ผลลัพธ์ของแบบจำลอง GCN	43
4-5	ค่าพารามิเตอร์สำหรับแบบจำลอง XGBoost	44
4-6	ผลลัพธ์ของแบบจำลอง XGBoost	45
4-7	ผลลัพธ์ของแบบจำลอง Hybrid GCN + XGBoost	77
4-8	ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง	77
4-9	ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง	77

สารบัญรูปภาพ

ภาพที่		หน้า
2-1	โครงสร้างแบบจำลอง GCN	4
2-2	โครงสร้างแบบจำลอง XGBoost	5
2-3	โครงสร้างแบบจำลอง Hybrid	9
3-1	Conceptual Framework Design	25
3-2	Workflow Design	26
3-3	User Interface Design	38
4-1	การแสดงผลหน้า User Interface	79
4-2	การแสดงผลการทำนายประเภทลูกค้าไม่ผิดนัดชำระหนี้	80
4-3	การแสดงผลการทำนายประเภทลูกค้ามีความเสี่ยงผิดนัดชำระหนี้	80

พระจอมเกล้าพระนครเหนือ มหาวิทยาลัยเทคโนโลยี

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

ในปี พ.ศ. 2566 ระดับหนี้ครัวเรือนของประเทศไทยปรับตัวเพิ่มขึ้นสู่ระดับร้อยละ 91.3 ต่อผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) ซึ่งมีสาเหตุหลักมาจากการขยายตัวของสินเชื่อเพื่อการอุปโภคบริโภคและสินเชื่อเพื่อที่อยู่อาศัย ขณะเดียวกัน สัดส่วนหนี้ที่ไม่ก่อให้เกิดรายได้ (Non-Performing Loans: NPLs) ในกลุ่มสินเชื่ออุปโภคบริโภค อสังหาริมทรัพย์ และสินเชื่อยานยนต์ มีแนวโน้มเพิ่มสูงขึ้นอย่างต่อเนื่องนับตั้งแต่ช่วงก่อนการแพร่ระบาดของโรคโควิด-19 [1] ส่งผลให้ประเทศไทยจัดอยู่ในกลุ่มประเทศที่มีระดับหนี้ครัวเรือนต่อ GDP สูงเป็นอันดับต้น ๆ ของโลก

นอกจากนี้ ภาระการชำระหนี้ต่อรายได้ของครัวเรือน (Debt Service Coverage Ratio: DSCR) ของประเทศไทยอยู่ที่ระดับร้อยละ 22.3 ซึ่งสูงกว่าค่าเฉลี่ยของประเทศเศรษฐกิจหลักที่อยู่ที่ประมาณร้อยละ 9.8 อย่างมีนัยสำคัญ ประกอบกับอัตราดอกเบี้ยนโยบายที่ปรับเพิ่มขึ้นสู่ระดับสูงสุดในรอบกว่า 10 ปี ส่งผลให้ความสามารถในการชำระหนี้ของภาคครัวเรือนลดลง และเพิ่มความเปราะบางต่อเสถียรภาพของระบบการเงินและเศรษฐกิจโดยรวมในระยะยาว โดยเฉพาะเมื่อโครงสร้างหนี้ครัวเรือนของไทยส่วนใหญ่กระจุกตัวอยู่ในสินเชื่อบัตรเครดิตและสินเชื่อส่วนบุคคล

ในปัจจุบัน ความต้องการใช้สินเชื่อส่วนบุคคลมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง เนื่องจากประชาชนใช้สินเชื่อดังกล่าวเพื่อรองรับค่าใช้จ่ายในชีวิตประจำวันและความจำเป็นทางการเงินต่าง ๆ อย่างไรก็ตาม การให้สินเชื่อย่อมนำมาซึ่งความเสี่ยงด้านการผิดนัดชำระหนี้ ซึ่งอาจนำไปสู่การเพิ่มขึ้นของ NPLs ในระบบสถาบันการเงิน อีกทั้งภายใต้หลักเกณฑ์การรายงานทางการเงิน TFRS 9 ที่กำหนดโดยธนาคารแห่งประเทศไทย สถาบันการเงินจำเป็นต้องตั้งสำรองผลขาดทุนด้านเครดิตที่คาดว่าจะเกิดขึ้น (Expected Credit Loss: ECL) ส่งผลให้ต้นทุนทางการเงินเพิ่มสูงขึ้น และอาจกระทบต่อสภาพคล่องและความสามารถในการดำเนินธุรกิจ

ด้วยเหตุนี้ สถาบันการเงินจึงได้นำเทคนิคการทำเหมืองข้อมูล (Data Mining) และ Machine Learning มาใช้ในการพยากรณ์ความเสี่ยงด้านเครดิตและโอกาสการผิดนัดชำระหนี้ของลูกค้า งานวิจัยจำนวนมากแสดงให้เห็นถึงประสิทธิภาพของแบบจำลองต่าง ๆ เช่น XGBoost [1–4] แบบจำลองโครงข่ายประสาทเชิงกราฟ ได้แก่ Graph Neural Network (GNN) และ Graph Convolutional Network (GCN) [5–8] รวมถึงแนวทางแบบผสมผสาน (Hybrid Models) [5, 8–10] ซึ่งสามารถเพิ่มความแม่นยำในการพยากรณ์ลูกหนี้ค้างชำระได้อย่างมีประสิทธิภาพ

ดังนั้น งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์การผิดนัดชำระหนี้ของลูกค้าสินเชื่อส่วนบุคคล โดยประยุกต์ใช้เทคนิคเหมืองข้อมูล ในการวิเคราะห์ข้อมูลลักษณะประชากรศาสตร์และพฤติกรรมทางการเงินของลูกค้า เพื่อสนับสนุนการบริหารความเสี่ยงด้านเครดิต ลดโอกาสการเกิดหนี้เสีย และช่วยให้สถาบันการเงินสามารถบริหารการตั้งสำรองภายใต้กรอบ TFRS 9 ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อสร้างแบบจำลองในการพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคล

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับการพยากรณ์ลูกหนี้ค้างชำระ

1.2.3 เพื่อนำเสนอผลการพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคล

1.3 ขอบเขตของการวิจัย

1.3.1 ข้อมูลที่ใช้เป็นข้อมูลพอร์ตสินเชื่อบุคคล ของสถาบันการเงินแห่งหนึ่ง 50,000 ระเบียน ข้อมูลจะประกอบด้วย จุดประสงค์การกู้ วงเงินกู้ ยอดคงเหลือ อายุ อาชีพ รายได้ ค่าใช้จ่าย ประเภทของผู้กู้ ระยะเวลาขอกู้ สถานะสมรส ระดับการศึกษา ประเภทที่อยู่อาศัย อายุงานและ กรรมสิทธิ์ในสถานประกอบการ เป็นต้น

1.3.2 เทคนิคการเรียนรู้เชิงลึกที่ใช้พัฒนาแบบจำลอง และเปรียบเทียบเพื่อหาแบบจำลองที่เหมาะสมสำหรับการพยากรณ์ลูกหนี้ค้างชำระ ประกอบด้วย 3 เทคนิค ได้แก่

1.3.2.1 โครงข่ายประสาทเทียมกราฟคอนโวลูชัน (Graph Convolutional Network: GCN)

1.3.2.2 เอ็กตรีมกราเดี้ยนบูทติ้ง (Extreme Gradient Boosting : XGBoost)

1.3.2.3 Hybrid of Graph Convolutional Network and Extreme Gradient Boosting (Hybrid GCN + XGBoost)

1.3.3 การประเมินประสิทธิภาพแบบจำลองสำหรับการพยากรณ์ลูกหนี้ค้างชำระ

1.3.3.1 ค่าความถูกต้อง (Accuracy)

1.3.3.2 ค่าความแม่นยำ (Precision)

1.3.3.3 ค่าระลึก (Recall)

1.3.3.4 ค่าการถ่วงดุล (F1-Score)

1.3.4 เครื่องมือที่ใช้ในการประมวลผลสำหรับการพยากรณ์ลูกหนี้ค้างชำระ

1.3.4.1 เครื่องมือในการเก็บรวบรวมข้อมูล : MS Excel

1.3.4.2 เครื่องมือในการเขียน Code : Visual Studio Code

1.3.4.3 ภาษาที่ใช้ในพัฒนาแบบจำลอง และนำเสนอ: Python, HTML, JS

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้แบบจำลองที่เหมาะสมสำหรับการพยากรณ์ลูกหนี้ค้างชำระ

1.4.2 เพื่อใช้เป็นข้อมูลสนับสนุนสำหรับหน่วยงานที่เกี่ยวข้องสามารถนำผลการวิจัยไปวางแผนวิเคราะห์ หรือกำหนดมาตรการต่าง ๆ เพื่อช่วยในการให้สินเชื่อสำหรับสินเชื่อบุคคล

1.4.3 เพื่อช่วยป้องกันความเสี่ยงที่เกิดจากการผิดนัดชำระหนี้ให้แก่สถาบันการเงินได้ดีขึ้น

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การดำเนินงานวิจัยการพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมือง ผู้วิจัยได้ศึกษาหลักการ ทฤษฎี และงานวิจัยที่เกี่ยวข้องเพื่อเป็นพื้นฐานสำหรับวิเคราะห์ความคิดเห็น ดังนี้

2.1 การพยากรณ์ลูกหนี้ค้างชำระ (Credit Risk Prediction)

2.2 โครงข่ายประสาทเทียมกราฟคอนโวลูชัน Graph Convolutional Network (GCN)

2.3 เอ็กตรีมกราเดี้ยนบูทติ้ง (Extreme Gradient Boosting : XGBoost)

2.4 การผสมผสานระหว่างโครงข่ายประสาทเทียมกราฟคอนโวลูชันและเอ็กตรีมกราเดี้ยนบูทติ้ง (GCN + XGBoost)

2.5 งานวิจัยที่เกี่ยวข้อง

2.1 การพยากรณ์ลูกหนี้ค้างชำระ (Credit Risk Prediction)

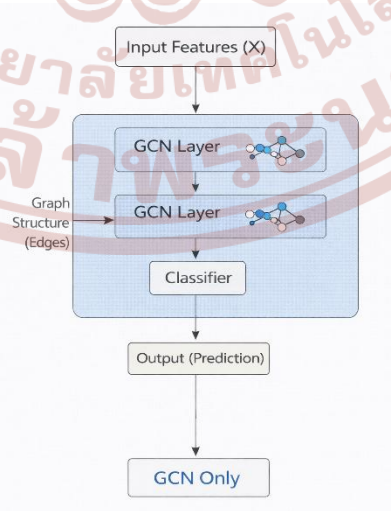
การพยากรณ์ (Forecasting) คือ กระบวนการประมาณค่าหรือคาดการณ์เหตุการณ์ในอนาคต โดยอาศัยข้อมูลในอดีตและปัจจุบันร่วมกับแบบจำลองทางสถิติหรือแบบจำลองเชิงคณิตศาสตร์ จุดประสงค์ของการพยากรณ์ คือ ช่วยสนับสนุนการตัดสินใจภายใต้ความไม่แน่นอน (Uncertainty) โดยเฉพาะในด้านธุรกิจ การเงิน และการบริหารความเสี่ยง [1-3]

ในบริบทของการพยากรณ์ลูกหนี้ค้างชำระ หรือ การพยากรณ์ความเสี่ยงเครดิต (Credit Risk Prediction) คือ การพยากรณ์มีลักษณะเป็นการจำแนกประเภท (Classification) โดยใช้กระบวนการ ข้อมูลและแบบจำลองทางสถิติหรืออัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning) เพื่อทำนายความน่าจะเป็นที่ลูกค้าจะผิดนัดชำระหนี้ (Default) หรือไม่สามารถชำระหนี้ตามกำหนดเวลาได้ โดยมีเป้าหมายเพื่อช่วยให้สถาบันการเงินสามารถบริหารความเสี่ยง ลดการสูญเสียเงินทุน และตัดสินใจในการปล่อยสินเชื่อได้อย่างมีประสิทธิภาพ โดยการจำแนกประเภท จะแบ่งลูกค้าออกเป็นกลุ่ม เช่น กลุ่มปกติ (Non-Default) กลุ่มผิดนัดชำระ (Default)

การพยากรณ์ลูกหนี้ค้างชำระโดยจะใช้ความสัมพันธ์ระหว่างตัวแปรเพื่อมองหาความเชื่อมโยงเชิงสถิติหรือเชิงโครงสร้างระหว่างปัจจัยต่าง ๆ ที่มีผลต่อการผิดนัดชำระหนี้ แล้วนำความสัมพันธ์เหล่านั้นมาใช้สร้างแบบจำลองการทำนาย ตัวอย่างเช่น รายได้ต่อเดือนกับภาระหนี้มักสัมพันธ์กันกับความสามารถในการชำระหนี้ หากลูกค้ามีรายได้ต่ำแต่ภาระหนี้สูงโอกาสผิดนัดก็จะมากขึ้น หรือสถานภาพสมรสกับพฤติกรรมการชำระหนี้ก็อาจมีความสัมพันธ์ที่สะท้อนถึงความเสี่ยงเครดิตได้เช่นกัน ในเชิงเทคนิค วิธีการนี้สามารถทำได้หลายแนวทาง เช่น การวิเคราะห์ความสัมพันธ์เชิงสถิติ หรือการใช้โมเดลที่สามารถเรียนรู้ความสัมพันธ์ที่ไม่เชิงเส้น เช่น Decision Tree, Random Forest, XGBoost ซึ่งจะจับความสัมพันธ์ระหว่างตัวแปรได้โดยอัตโนมัติ หรือ ใช้โครงสร้างกราฟ (Graph-based Models) เช่น GCN ที่มองลูกค้าเป็นโหนดและความสัมพันธ์ระหว่างลูกค้าเป็นเส้นเชื่อม โมเดลจะเรียนรู้ว่าลูกค้าที่เชื่อมโยงกับผู้ที่ผิดนัดมีโอกาสสูงที่จะผิดนัดเช่นกัน ซึ่งเป็นการใช้ “ความสัมพันธ์ระหว่างตัวแปร” ในเชิงโครงสร้างเครือข่าย ไม่ใช่แค่เชิงสถิติ

2.2 โครงข่ายประสาทเทียมกราฟคอนโวลูชัน Graph Convolutional Network (GCN)

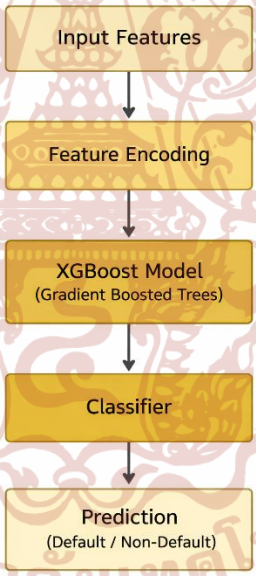
Graph Convolutional Network (GCN) แนวคิดพื้นฐานของ GCN เกิดจากข้อจำกัดของโมเดลแบบดั้งเดิมที่มองข้อมูลเป็นตาราง โดยถือว่าตัวอย่างแต่ละแถวเป็นอิสระต่อกัน แต่ในโลกความจริง ข้อมูลจำนวนมากมี “ความสัมพันธ์” ซ่อนอยู่ เช่น ลูกค้าที่มีลักษณะใกล้เคียงกันอยู่ในกลุ่มเดียวกัน หรือมีพฤติกรรมคล้ายกัน ซึ่งความสัมพันธ์เหล่านี้ไม่สามารถอธิบายได้ด้วยฟีเจอร์ของแต่ละตัวอย่างเพียงอย่างเดียว GCN จึงถูกออกแบบมาเพื่อเรียนรู้ข้อมูลในรูปแบบกราฟ ซึ่งประกอบด้วยโหนด (ตัวอย่าง) และเส้นเชื่อม (ความสัมพันธ์ระหว่างตัวอย่าง) [5-6,8] ในบริบทของ GCN มีปัญหาสำคัญประการหนึ่งคือ ข้อมูลเชิงตารางส่วนใหญ่มักไม่ได้มีกราฟหรือความสัมพันธ์กำหนดมาให้โดยตรง ดังนั้นจึงจำเป็นต้องมีวิธี “สร้างกราฟ” ขึ้นจากข้อมูลก่อน ซึ่งแนวคิดของ k-Nearest Neighbors (KNN) จึงถูกนำมาใช้ในขั้นตอนนี้ โดยพิจารณาความใกล้เคียงของแต่ละฟีเจอร์ที่มีลักษณะคล้ายกันมากที่สุดจำนวน k ตัวเป็นเพื่อนบ้าน ความคล้ายคลึงนี้มักวัดจากระยะทางเชิงตัวเลข เช่น Euclidean Distance หลังจากที่ฟีเจอร์ถูกปรับสเกลให้อยู่ในมาตรฐานเดียวกัน ซึ่งตัวอย่างแต่ละตัวจะถูกมองเป็นโหนด และเส้นเชื่อมจะถูกสร้างขึ้นระหว่างโหนดกับเพื่อนบ้านที่ใกล้ที่สุดของมัน [7] แนวคิดนี้สะท้อนสมมติฐานว่าตัวอย่างที่มีคุณลักษณะใกล้เคียงกันมีแนวโน้มที่จะมีพฤติกรรมหรือสถานะคล้ายกันด้วย เมื่อกราฟลักษณะนี้ถูกป้อนเข้า GCN กระบวนการ Neighborhood Aggregation จะไม่ใช่เพียงการรวมข้อมูลจากเพื่อนบ้านแบบสุ่ม แต่เป็นการรวมข้อมูลจากตัวอย่างที่ “คล้ายกันจริงในเชิงข้อมูล” ส่งผลให้ Embedding ที่ได้สามารถสะท้อนทั้งคุณสมบัติของตัวอย่างเองและบริบทของกลุ่มตัวอย่างที่มีลักษณะใกล้เคียงกัน กล่าวอีกนัยหนึ่ง kNN ทำหน้าที่เป็นสะพานเชื่อมระหว่างข้อมูลเชิงตารางกับโลกของกราฟ โดยแปลงความใกล้เคียงเชิงตัวเลขให้กลายเป็นโครงสร้างความสัมพันธ์ที่ GCN สามารถเรียนรู้และใช้ประโยชน์ได้อย่างเป็นระบบเชิงตัวเลขให้กลายเป็นโครงสร้างความสัมพันธ์ที่ GCN สามารถเรียนรู้และใช้ประโยชน์ได้อย่างเป็นระบบ สถาปัตยกรรมของแบบจำลอง GCN แสดงดังภาพที่ 2-1



ภาพที่ 2-1 สถาปัตยกรรมของแบบจำลอง GCN

2.3 เอ็กตรีมกราเดี้ยนบูทติ้ง Extreme Gradient Boosting (XGBoost)

XGBoost มีแนวคิดพื้นฐานที่แตกต่างอย่างชัดเจน XGBoost เป็นโมเดลในตระกูล Gradient Boosting ที่สร้างแบบจำลองจากการรวมต้นไม้ตัดสินใจจำนวนมากเข้าด้วยกัน โดยหลักการคือ การสร้างต้นไม้ทีละต้น เพื่อแก้ไขข้อผิดพลาดของต้นไม้ก่อนหน้า โมเดลจะเริ่มจากการทำนายแบบหยาบ จากนั้นค่อย ๆ เพิ่มต้นไม้ใหม่ที่โฟกัสไปยังตัวอย่างที่ทำนายผิดหรือทำนายได้ไม่ดีในรอบก่อน การอัปเดตนี้อาศัยแนวคิดของการไล่ระดับของฟังก์ชันความสูญเสีย ทำให้โมเดลเรียนรู้ได้อย่างมีประสิทธิภาพและควบคุมการ Overfitting ได้ดี [1] จุดแข็งของ XGBoost อยู่ที่ความสามารถในการจับความสัมพันธ์แบบไม่เชิงเส้น การตั้งเงื่อนไขแบบ Threshold และการเรียนรู้ปฏิสัมพันธ์ระหว่างฟีเจอร์หลายตัวพร้อมกัน เช่น ฟีเจอร์หนึ่งจะมีความหมายเฉพาะเมื่ออีกฟีเจอร์หนึ่งอยู่ในช่วงค่าหนึ่ง ซึ่งเป็นสิ่งที่ต้นไม้ตัดสินใจทำได้ดีมาก นอกจากนี้ XGBoost ยังมีโครงสร้างที่เหมาะกับข้อมูลเชิงตาราง มีความเสถียร และตีความได้ผ่าน Feature Importance สถาปัตยกรรมของแบบจำลอง XGBoost แสดงดังภาพที่ 2-2



ภาพที่ 2-2 โครงสร้างแบบจำลอง XGBoost

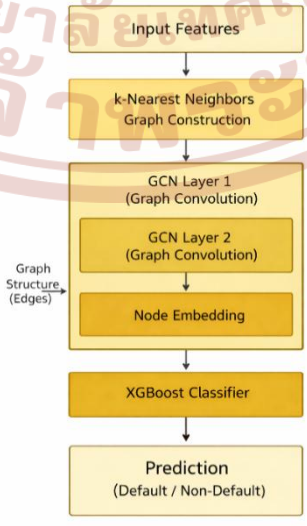
2.4 การผสมผสานระหว่างโครงข่ายประสาทเทียมกราฟคอนโวลูชันและเอ็กตรีมกราเดี้ยนบูทติ้ง (GCN + XGBoost)

เกิดจากการรวมข้อดีของสองอัลกอลิทึมเข้าด้วยกัน โดยให้ Graph Convolutional Network (GCN) ทำหน้าที่เรียนรู้โครงสร้างความสัมพันธ์ระหว่างตัวอย่าง และให้ XGBoost ทำหน้าที่เป็นตัวตัดสินใจสุดท้ายที่แข็งแรงกับข้อมูลตารางและปฏิสัมพันธ์แบบไม่เชิงเส้น (Non-linear Interactions) แนวคิดสำคัญของ hybrid คือ การใช้ GCN เป็น “ตัวสร้างฟีเจอร์เชิงกราฟ” หรือ Representation ใหม่ ๆ ที่สะท้อนบริบทของแต่ละตัวอย่างในเครือข่าย แล้วค่อยป้อน Representation นั้นให้ XGBoost ซึ่งเป็นโมเดลที่มักให้ความแม่นยำสูงกับ tabular data อยู่แล้ว

สรุปให้เป็นภาษาเชิงแนวคิดคือ GCN ทำหน้าที่ “เรียนรู้คุณลักษณะจากความสัมพันธ์” ส่วน XGBoost ทำหน้าที่ “จัดประเภท/พยากรณ์ผลลัพธ์” อย่างมีประสิทธิภาพ

ขั้นตอนแรกของวิธีนี้คือการแปลงตารางให้เป็นกราฟ โดยกำหนดให้แต่ละแถวข้อมูลเป็นโหนด (Node) หนึ่งโหนดหรือก็คือ ลูกหนี้หนึ่งราย จากนั้นต้องนิยามเส้นเชื่อม (Edge) ระหว่างโหนดเพื่อสะท้อนความสัมพันธ์ที่เราต้องการให้โมเดลรับรู้ โดยสร้าง Edge แบบ k-Nearest Neighbors (kNN) ซึ่งอาศัยความใกล้กันของเวกเตอร์ฟีเจอร์หลังสเกล เช่น StandardScaler เพื่อเชื่อมลูกหนี้ที่มีลักษณะคล้ายกันในเชิงตัวเลขเข้าด้วยกัน นอกจากนี้ยังสามารถเพิ่ม Edge แบบ Group-based [15] ได้ เช่น เชื่อมลูกหนี้ที่มี “Status of Residence” เดียวกัน เพื่อให้เกิดชุมชนย่อยที่สะท้อนกลุ่มประชากรในมิติที่เราสนใจ เมื่อทำสองส่วนนี้รวมกัน เราได้กราฟที่โหนดแต่ละตัวไม่ได้มีแค่ฟีเจอร์ของตัวเอง แต่ยัง “มีเพื่อนบ้าน” ที่บอกบริบทว่าโหนดนี้อยู่ใกล้ใคร/อยู่กลุ่มไหน

องค์ประกอบหลักที่ทำให้โมเดลนี้เป็น Hybrid [5,8-9] จริง ๆ คือการนำผลลัพธ์ที่ได้จาก GCN ซึ่งก็คือ Embedding มารวมเข้ากับข้อมูลดั้งเดิม ก่อนจะส่งผลทั้งหมดให้ XGBoost เพื่อทำนาย แนวคิดนี้เกิดจากที่ GCN มีความสามารถโดดเด่นในการเรียนรู้ “บริบทจากความสัมพันธ์” ระหว่างตัวอย่าง เช่น ลูกหนี้ที่มีลักษณะคล้ายกันหรืออยู่ในกลุ่มเดียวกันจะมีอิทธิพลต่อกันผ่านโครงสร้างกราฟ แต่ในขณะเดียวกัน กระบวนการรวมข้อมูลจากเพื่อนบ้านของ GCN อาจทำให้รายละเอียดเชิงตัวเลขบางอย่างของแต่ละตัวอย่าง เช่น ค่ายอดเงินหรือความแตกต่างเล็ก ๆ ในฟีเจอร์สำคัญ ถูกเฉลี่ยหรือถูกลดความเด่นลงไป การเก็บฟีเจอร์ดั้งเดิมไว้คู่กับ Embedding จึงช่วยรักษาข้อมูลทั้งสองมุมมองไว้พร้อมกัน ซึ่ง Embedding จาก GCN ทำหน้าที่สรุปภาพรวมเชิงโครงสร้างและความสัมพันธ์ของตัวอย่าง ส่วนฟีเจอร์ดั้งเดิมยังคงให้รายละเอียดเชิงเนื้อหาของแต่ละตัวอย่างอย่างครบถ้วน จากนั้น XGBoost ซึ่งมีจุดเด่นในการเรียนรู้ความสัมพันธ์แบบไม่เชิงเส้นและการตั้งเงื่อนไขแบบ threshold จะนำข้อมูลทั้งสองส่วนนี้มาผสมผสานกันเพื่อสร้างแบบจำลองที่แม่นยำ กล่าวอย่างง่ายคือ GCN ทำหน้าที่สร้าง “ภาษากลาง” ที่สะท้อนบริบทของข้อมูล ในขณะที่ XGBoost ทำหน้าที่ “ตีความภาษา” นี้ร่วมกับข้อมูลดั้งเดิม เพื่อทำนายผลลัพธ์ได้อย่างมีประสิทธิภาพสูงสุด สถาปัตยกรรมของแบบจำลอง Hybrid แสดงดังภาพที่ 2-3



ภาพที่ 2-3 โครงสร้างแบบจำลอง Hybrid

2.5 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยทำการศึกษาค้นคว้า บทความวิจัยที่มีความเกี่ยวข้องกับการวิเคราะห์ข้อความด้วยเทคนิคต่าง ๆ รวมถึงแนวทางขั้นตอนวิธีการในการเปรียบเทียบ เพื่อใช้เป็นแนวทางสำหรับงานวิจัยที่เกี่ยวข้องกับการพยากรณ์การผิดนัดชำระหนี้ของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล ดังต่อไปนี้

กมลชนก [1] ได้สร้างแบบจำลองการพยากรณ์ลูกหนี้ที่ค้างชำระหนี้ของสินเชื่อเพื่อที่อยู่โดยเปรียบเทียบอัลกอริทึม 5 อัลกอริทึม ได้แก่ 1) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน 2) อัลกอริทึมเพื่อนบ้านใกล้ที่สุด 3) อัลกอริทึมเอ็กตรีมกราเดี้ยนบูทติ้ง 4) อัลกอริทึมป่าสุ่ม และ 5) อัลกอริทึมการเรียนรู้เชิงลึก และเทคนิคการเรียนรู้แบบรวมกลุ่มด้วยวิธีการลงคะแนนเสียงข้างมาก โดยใช้ข้อมูลพอร์ตสินเชื่อเพื่อที่อยู่อาศัย ตั้งแต่เดือนธันวาคม 2561 ถึง เดือนธันวาคม 2562 จำนวน 28,418 ระเบียน ประกอบด้วยคุณลักษณะ จำนวน 10 คุณลักษณะ จากนั้นแบ่งข้อมูลเป็น 2 ส่วน คือ ข้อมูลสำหรับสร้างแบบจำลองจำนวน 19,896 ระเบียน และข้อมูลสำหรับทดสอบแบบจำลองจำนวน 8,522 ระเบียน โดยนำไปทดลองประสิทธิภาพอัลกอริทึมที่เหมาะสมสำหรับการพยากรณ์ด้วยค่าความถูกต้อง ค่าความแม่นยำ ค่าระลึก และค่าการถ่วงดุล ผลการวิจัยพบว่า อัลกอริทึมเอ็กตรีมกราเดี้ยนบูทติ้งให้ประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องร้อยละ 86.84 ค่าความแม่นยำเท่ากับ 0.87 ค่าระลึกเท่ากับ 0.87 และค่าการถ่วงดุลเท่ากับ 0.87 ตามลำดับ

เครือวัลย์ และ ศิริสรรพ [2] ได้ศึกษาการทำนายลูกหนี้ที่มีโอกาสในการผิดนัดชำระกับทางธนาคาร โดยการทดลองกับชุดข้อมูลการทำธุรกรรมสินเชื่อบัตรเครดิตซึ่งประกอบด้วยข้อมูลจำนวนทั้งหมด 307,511 แถว และคอลัมน์ทั้งหมด 122 คอลัมน์ โดยการแบ่งข้อมูลออกเป็น 2 กลุ่มใหญ่ๆ คือ กลุ่มลูกหนี้ปกติ คือกลุ่มลูกหนี้ที่ไม่ได้มีการผิดนัดชำระกับทางธนาคารและกลุ่มลูกหนี้ที่ไม่ปกติ คือกลุ่มลูกหนี้ ที่มีการผิดนัดชำระกับทางธนาคารเครื่องมือหลักที่นักวิจัยใช้ ได้แก่ Logistic Regression, XGBoostClassifier, K-nearest Neighbors, Random Forest , Support Vector Classifier (SVC), Gradient Boosting เป็นต้น โดยผลการศึกษาพบว่า การพัฒนาแบบจำลอง โดยการใช้เทคนิควิธี XGBoost ให้ค่าความไว (Recall) ที่มากที่สุด ซึ่งมีค่าเท่ากับ 0.97 มีค่าความถูกต้อง (Accuracy) เท่ากับ 0.37 และมีค่า F1-Score ที่ใช้ในการวัดความสามารถของแบบจำลอง เท่ากับ 0.51 แต่เทคนิควิธี KNN ให้ค่าความไว (Recall) ที่น้อยที่สุด ซึ่งมีค่าเท่ากับ 0.57 มีค่าความถูกต้อง(Accuracy) เท่ากับ 0.55 และมีค่า F1-Score เท่ากับ 0.46

Chang and et al. [3] ได้ศึกษาวิธีการในการพยากรณ์สถานะการผิดนัดชำระหนี้ของลูกค้าบัตรเครดิต โดยใช้การประยุกต์อัลกอริทึมการเรียนรู้ของเครื่องหลายรูปแบบ ได้แก่ โครงข่ายประสาทเทียม (Neural Networks) การถดถอยแบบลอจิสติก (Logistic Regression) AdaBoost XGBoost และ LightGBM โดยใช้ตัวชี้วัดประสิทธิภาพ เช่น Accuracy, Precision, Recall, F1 Score, ROC และ MCC เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมแต่ละแบบผลการศึกษาพบว่า XGBoost มีประสิทธิภาพเหนือกว่าโมเดลอื่น ๆ โดยให้ค่า Accuracy สูงถึง 99.4% ผลลัพธ์จากการศึกษานี้ชี้ให้เห็นว่า การวิเคราะห์ความเสี่ยงด้านเครดิตที่มีประสิทธิภาพจะช่วยสนับสนุนการตัดสินใจในการปล่อยสินเชื่อได้อย่างรอบคอบ และการนำอัลกอริทึมการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกมาใช้ได้ช่วยเพิ่มความแม่นยำในการพยากรณ์อย่างมีนัยสำคัญ

Gunnarsson et al. [4] การศึกษาความเหมาะสมของอัลกอริทึม Deep Learning สำหรับการให้คะแนนเครดิต โดยในการศึกษานี้ได้สร้างสถาปัตยกรรม Deep Learning จำนวนสองรูปแบบ ได้แก่ โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Perceptron: MLP) และโครงข่ายความเชื่อเชิงลึก (Deep Belief Network: DBN) จากนั้นจึงนำประสิทธิภาพของโมเดลดังกล่าวไปเปรียบเทียบกับ Random Forestc และ XGBoost ซึ่งจากงานวิจัยพบว่า XGBoost ให้ประสิทธิภาพดีที่สุดสำหรับการให้คะแนนเครดิตสูงสุด เมื่อเทียบกับทุกวิธีที่นำมาศึกษา โดยมีค่า AUC อยู่ที่ 98.7%

Sukharev and et al. [5] ต้องการพัฒนาโมเดล Credit Scoring ที่แม่นยำมากขึ้น โดยใช้ข้อมูลการโอนเงินระหว่างลูกค้าในระบบธนาคาร เพื่อแสดงให้เห็นว่าความเชื่อมโยงระหว่างลูกค้าและการโอนเงินสามารถช่วยปรับปรุงคุณภาพของการให้คะแนนเครดิตได้อย่างมีนัยสำคัญ เมื่อเปรียบเทียบกับวิธีที่ใช้เฉพาะข้อมูลของลูกค้า เพียงอย่างเดียว โดยใช้อัลกอริทึมในการเปรียบเทียบคือ Graph Convolutional Networks (GCN), GraphSAGE, Graph Attention Networks (GAT), Edge Enhanced GNN (EGNN), recurrent neural network (RNN) และ Light Gradient-Boosting Machine (LightGBM) ซึ่งผลลัพธ์ คือ โมเดล EWS-GCN ซึ่งเป็นการ Combine ระหว่าง GCN และ RNN สามารถประมวลผลข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ และให้ผลลัพธ์ดีกว่าโมเดล GNN ทั่วไป โดยผลลัพธ์ค่า ROC อยู่ที่ 89.26%

Sun and et al. [6] ได้เสนอแนวทางใหม่ในการพยากรณ์ความเสี่ยงด้านเครดิต โดยใช้ Graph Convolutional Neural Networks (GCNNs) เพื่อประเมินความน่าเชื่อถือของผู้ขอสินเชื่อ โดยใช้เทคนิค ข้อมูลขนาดใหญ่ (Big Data) และ ปัญญาประดิษฐ์ (AI) ซึ่งงานวิจัยเริ่มต้นด้วยการแปลงข้อมูลของผู้ขอสินเชื่อให้เป็นข้อมูลเชิงกราฟ โดยผู้กู้และความสัมพันธ์ จะถูกแปลงเป็น โหนด (Nodes) และขอบ (Edges) ตามลำดับ จากนั้นใช้ Graph Convolutional Networks (GCN), Support Vector Machine (SVM), Decision Tree, backpropagation และ SGCNs ซึ่งเป็นการ hybrid GCNN ซึ่งผลลัพธ์ของการวิจัย คือ โมเดล SGCNs มีประสิทธิภาพสูงสุด โดยค่าความแม่นยำ (Accuracy) เฉลี่ยอยู่ที่ 87.52%, ค่า F1-score เฉลี่ย 93.13% และค่า G-mean เฉลี่ย 85.59%

Shi and et al. [7] ได้พยายามพัฒนาแบบจำลองในการพยากรณ์ความเสี่ยงด้านเครดิตโดยใช้ Graph Representation Learning ร่วมกับ Graph Transformation โดยการวิจัยนี้ใช้ข้อมูลภายในของลูกค้าจาก German Credit, Lending Club และ Home Credit เช่น รายได้, อายุ, ประวัติการชำระหนี้, จำนวนบัญชี, อัตราส่วนหนี้สิน, ระยะเวลาการมีบัญชี และมาแปลงเป็นคุณลักษณะของ node ในกราฟ โดยการวิจัยนี้จะใช้อัลกอริทึม kNN Hybrid GNN เทียบกับ Logistic Regression, Random Forest และXGBoost โดยผลลัพทธ์ของงานวัจัย โมเดล kNN–GNN แบบไฮบริด ให้ผลลัพธ์ดีกว่าโมเดล Logistic Regression และ Random Forest โดยมีค่า Accuracy 88.9% ค่า F1 Score 85.7% และ ค่า AUC 91.3%

Wang and et al. [8] ได้ออกแบบโมเดลแบบ Hybrid ที่รวมอัลกอริทึมหลายประเภทเข้าด้วยกันเพื่อจัดการกับข้อมูลผู้ใช้ที่มีทั้งโครงสร้างแบบกราฟและลักษณะเชิงเวลา โดยจะใช้ข้อมูลจากจริงจาก Alipay ซึ่งงานวิจัยนี้ถูกจัดทำเพื่อนำไปใช้จริงในการคัดกรองความเสี่ยงของผู้กู้ในแอปพลิเคชัน Alipay โดยจะใช้อัลกอริทึม Xgboost, GraphSAGE, LSTM, GAT, GCN-LSTM, ST-GCN และ TemGNN (เป็นการ Hybrid Model ST-GCN-LSTM) จากงานวิจัยพบว่า TemGNN

เป็นแบบจำลองที่มีประสิทธิภาพสูงสุด โดยมีค่า Accuracy 84.3% ค่า F1 Score 85.6% และ ค่า AUC 87.9%

Liu Zhang and Fan. [9] ได้ทำการศึกษา เทคนิคการพยากรณ์ความเสี่ยงด้านเครดิต เพื่อวัดความน่าจะเป็นในการผิดนัดของผู้ขอสินเชื่อด้วยวิธี Machine Learning และ Big Data เพื่อสร้างโมเดลที่มีประสิทธิภาพในการพยากรณ์ความเสี่ยงด้านเครดิต โดยจะใช้อัลกอริทึม XGBoost, GNN (forgeNet) และ Hybrid ระหว่าง 2 อัลกอริทึมเพื่อดูประสิทธิภาพสูงสุด โดยใช้ข้อมูล สินเชื่อจริงจาก Lending Club ในช่วงปี 2550–2559 ได้ผลลัพธ์เฉลี่ยที่ดีที่สุด ในด้านความแม่นยำ (Accuracy), ค่า F1-score และ G-mean ที่ 87.52%, 93.13% และ 85.59% ตามลำดับและ ความสามารถในการพยากรณ์ดีขึ้นจากค่าเฉลี่ยเดิมอยู่ที่ 6.14, 7.59 และ 6.18 เมื่อเปรียบเทียบกับโมเดลอื่น ๆ

Albanesi และ Vamossy .[10] ได้พัฒนาโมเดลเพื่อคาดการณ์การผิดนัดของผู้บริโภค โดยอิงจากการเรียนรู้เชิงลึก (Deep Learning) โดยใช้ข้อมูลที่อยู่ในช่วงวิกฤตบ้านในช่วงปี 2550 - 2552 โดยใช้การเปรียบเทียบ Logistic Regression, Decision Trees, Gradient Boosting (GBM), Deep Neural Network (DNN) และ DNN + GBM โดยผลลัพธ์ของการพัฒนา คือ การ Hybird DNN และ GBM เป็นโมเดลที่มีผลลัพธ์ที่ดีที่สุด โดยค่า Accuracy: มากกว่า 86% ทุกช่วงเวลา ค่า AUC-Score: มากกว่า 92% และ ค่า Gini Coefficient: ประมาณ 86% โดยอัลกอริทึมที่ได้ผลลัพธ์ แย่ที่สุด คือ Logistic Regression

ทิพย์ธิดา [11] ได้จัดทำเหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อของ บริษัท กรุงไทย คาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน)เพื่อช่วยลดปริมาณหนี้สูญ โดยรวบรวมข้อมูลประวัติลูกค้าที่ผ่านการพิจารณาให้สินเชื่อ การเตรียมและคัดเลือกข้อมูลให้สมบูรณ์เหมาะสำหรับการทำเหมืองข้อมูล ได้แก่ การโอนย้ายข้อมูล การลดขนาดของข้อมูล และการทำความสะอาดข้อมูล หลังจากนั้นจึงได้ทำการทดสอบข้อมูล โดยใช้เทคนิคการจำแนกกลุ่มเพื่อหากฎที่ใช้ในการจำแนกลูกค้ากลุ่มดีและลูกค้ากลุ่มไม่ดี ซึ่งได้ทดลองจำแนกกลุ่มข้อมูลด้วยเทคนิคการจำแนกกลุ่ม 3 เทคนิค ซึ่งให้ผลลัพธ์เป็นกฎ ได้แก่ โมเดลการจำแนกกลุ่มแบบ Decision Tree และ โมเดลการจำแนกกลุ่มแบบ Bayes : Naïve Bayes ผลลัพธ์ที่ได้คือกฎที่ใช้ในการจำแนก ลูกค้ากลุ่มดี และลูกค้ากลุ่มไม่ดี ซึ่งผลการทดลองพบว่า การจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Tree ให้ผลลัพธ์ดีที่สุด เนื่องจากมีความถูกต้องมากที่สุดและสามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนมากที่สุด เมื่อเทียบกับอีกสองเทคนิค ซึ่งท้ายสุดจึงนำกฎที่ได้จากผลลัพธ์มาเป็นแนวทางและเงื่อนไข ในการนำมาเป็นตัวอย่างที่สำคัญสำหรับการเขียนโปรแกรมเป็นระบบการตัดสินใจอนุมัติสินเชื่อออนไลน์ได้อย่างมีประโยชน์

ญาณินี และ พุธษดี [12] ได้จัดทำโมเดลทำนายความสามารถในการชำระหนี้เงินกู้ของสมาชิกสำนักงานส่งเสริมสวัสดิการและสวัสดิการภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี โดยการรวบรวมปัจจัยที่มีคาดว่าจะมีผลต่อการความสามารถในการชำระเงินกู้จำนวน 8 ปัจจัย ซึ่งจากการวิเคราะห์ความสัมพันธ์ของปัจจัยโดยใช้การวิเคราะห์การถดถอยพหุคูณ ผลการศึกษาพบว่าปัจจัยที่มีความสัมพันธ์กับความสามารถในการชำระเงินกู้ได้แก่จำนวนบุตร เพศ อายุ รายได้ต่อเดือน จำนวนเงินกู้ ภาระหนี้ การทำประกัน เงินกู้ จำนวนเงินงวด

ข้อมูลได้จากการรวบรวมประวัติการกู้เงินระหว่างปี พ.ศ.2547 - 2557 จำนวนทั้งสิ้น 11,732 รายการ การสร้างโมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้โดยใช้ต้นไม้การตัดสินใจ เครือข่ายความเชื่อเบย์และโครงข่ายประสาทเทียม ผลการศึกษาพบว่า ประสิทธิภาพการทำนายของโมเดลโครงข่ายประสาทเทียมพบค่าความแม่นยำสูงสุดที่ 98.77 โมเดลเครือข่ายความเชื่อเบย์ที่ค่าความแม่นยำ 98.76 และโมเดลต้นไม้การตัดสินใจ ที่ค่าความแม่นยำ 98.74 ตามลำดับ เมื่อเปรียบเทียบประสิทธิภาพการทำนายแล้ว พบว่า โมเดลโครงข่ายประสาทเทียมมีประสิทธิภาพการทำนายสูงที่สุด

ขวัญนภา [13] ได้ศึกษาการนำเทคนิคเหมืองข้อมูลมาใช้เพื่อระบุปัจจัยสำคัญในการพิจารณาสินเชื่อ โดยพิจารณาจากข้อมูลผู้ที่ยื่นสินเชื่อโครงการธนาคารประชาชนจำนวน 14 ปัจจัย จากข้อมูลผู้ยื่นสินเชื่อทั้งหมด 1,000 ตัวอย่าง โดยมีวัตถุประสงค์เพื่อเปรียบเทียบกระบวนการอนุมัติสินเชื่อของธนาคารโดยไม่ใช้วิธีการคัดเลือกคุณลักษณะ และใช้วิธีการคัดเลือกคุณลักษณะ ซึ่งประกอบด้วยวิธี Chi-Square และวิธี Information Gain จากนั้นนำข้อมูลที่ได้ไปสร้างโมเดลด้วยวิธี Support Vector Machine, Multi-Layer Perceptron และ Decision Tree เพื่อใช้สำหรับพิจารณาการได้สินเชื่อของธนาคาร ในการทดสอบได้เลือกใช้วิธี K-Fold Cross-Validation โดยกำหนดให้ K มีค่าเท่ากับ 10 จากการทดลองโดยไม่ใช้วิธีการคัดเลือกคุณลักษณะ พบว่าวิธี Support Vector Machine มีความถูกต้องสูงที่สุดที่97.12% จากนั้นทดสอบด้วยวิธีการคัดเลือกคุณลักษณะพิเศษ พบว่าวิธี Chi-Square คัดเลือกคุณลักษณะพิเศษจำนวนทั้งสิ้น 10 คุณลักษณะ และ วิธีInformation Gain คัดเลือกคุณลักษณะพิเศษจำนวนเพียง 3 คุณลักษณะ จากนั้นได้นำข้อมูลไปสร้างโมเดลด้วยวิธี Support Vector Machine, Multi-Layer Perceptron และ Decision Tree จากการทดลองพบกว่า วิธี Chi-Square ร่วมกับ Multi-Layer Perceptron มีความถูกต้อง 90.40% และวิธี Information Gain ร่วมกับ Multi-Layer Perceptron มีความถูกต้อง 91.70% จากการทดลองจึงสรุปได้ว่าการพิจารณาเพื่ออนุมัติสินเชื่อของธนาคารด้วยวิธี Support Vector Machine มีความถูกต้องสูงที่สุดและไม่จำเป็นต้องใช้กระบวนการคัดเลือกคุณลักษณะพิเศษ

ลัลธริดา [14] ได้พัฒนาแบบจำลองในการพยากรณ์ความล้มเหลวและภาวะกดดันทางการเงินของบริษัทในตลาดหลักทรัพย์แห่งประเทศไทยในช่วงปี พ .ศ.2550 - พ .ศ.2560 ด้วยแบบจำลองที่ใช้ทั้งข้อมูลอัตราส่วนทางการเงิน (Financial ratio Based) และข้อมูลราคาตลาด (Market Based) จากการเก็บรวบรวมข้อมูลบริษัทที่เผชิญกับภาวะล้มเหลวและกดดันทางการเงินตามเกณฑ์ที่กำหนดทั้งสิ้น 80 บริษัทและบริษัทที่ดำเนินการปกติทั้งสิ้น 160 บริษัท โดยทำการคัดเลือกด้วยการจับคู่แบบ 1:2 (REHABCO : NON-REHABCO) ผลการศึกษาสำหรับการพยากรณ์ล่วงหน้า 1 ปีก่อนที่บริษัทจะประสบภาวะล้มเหลวทางการเงินด้วย Binary Logistic Regression พบว่ามี 3 ตัวแปรอิสระที่สามารถพยากรณ์ภาวะความล้มเหลวและกดดันทางการเงินได้อย่างมีนัยสำคัญที่ระดับ 0.05 อันได้แก่ อัตราส่วนที่ใช้วัดความสามารถในการทำกำไร NIMTA), อัตราส่วนทางการเงินที่แสดงถึงสภาพคล่องของกิจ (CASHMTA) และ ขนาดของบริษัท (RSIZE) โดยกำหนด Cutting Point เท่ากับ 0.6 แบบจำลองสามารถพยากรณ์ความล้มเหลวทางการเงินของบริษัทที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยได้แม่นยำถึง 89.36% ไม่มี Type I Error และมี Type II Error เท่ากับ 12.5% จากการทดสอบโดยตัวอย่าง Holdout Sample จำนวน 47 บริษัท

นพมาศ [15] ได้ศึกษาปัจจัยที่ทำให้เกิดปัญหาหนี้ค้าง (NPL) ผ่านโครงการพัฒนาระบบสินเชื่อดิจิทัล (Digital Lending) ในกลุ่มลูกค้าของธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร (ธ.ก.ส.) จังหวัดลพบุรี และเพื่อประยุกต์เหมืองข้อมูลในการอนุมัติสินเชื่อในการแก้ไขปัญหาการจ่ายเงินกู้ โดยเก็บรวบรวมข้อมูลของลูกค้าเงินกู้ที่เข้าร่วมโครงการสินเชื่อฉุกเฉินที่มีอยู่ในระบบสารสนเทศของธนาคารเพื่อการเกษตรและสหกรณ์การเกษตร จังหวัดลพบุรี จำนวน 9,002 ราย โดยทำการวิเคราะห์ข้อมูลโดยการสร้างแบบจำลองการเรียนรู้ผ่านโปรแกรม Weka เพื่อหาความสำคัญของปัจจัยที่ส่งผลให้เกิดปัญหาหนี้ค้าง (NPL) ผลการศึกษา จากการทดลองใช้แบบจำลองพยากรณ์ จำนวน 3 แบบ ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree), เทคนิคเบย์อย่างง่าย (Naïve Bayesain classifier) และเทคนิคซัพพอร์ตเวกเตอร์เมชชีน (Support Vector Machine; SVM) พบว่า เทคนิคต้นไม้ตัดสินใจ (Decision Tree) มีค่าความถูกต้องในการทำนาย (Correctly Classified Instances) มากที่สุด

รังรองรัตน์ [16] ได้ทำการศึกษาค้นหาตัวแปรที่มีผลต่อการวิเคราะห์สินเชื่อเพื่อใช้ในการพัฒนาแพลตฟอร์มการกู้ยืมระหว่างบุคคลกับบุคคลผ่านช่องทางออนไลน์จากข้อมูลทางเลือกแบบทดสอบบุคลิกภาพและแบบประเมินเจตคติต่อเงิน กลุ่มตัวอย่างที่ใช้ในการวิจัย คือบุคคลทั่วไปที่มีข้อมูลเครดิตบูโร จำนวน 400 คน โดยทำการคัดเลือกกลุ่มตัวอย่างแบบโควตา ผสมผสานกับการคัดเลือกกลุ่มตัวอย่างแบบสะดวก โดยแบ่งออกเป็นผู้ที่มีประวัติเครดิตบูโรในระดับความเสี่ยงสูงกว่าหรือเท่ากับระดับความเสี่ยง CC จำนวน 250 คน (ร้อยละ 62.5) และ ผู้ที่มีประวัติเครดิตบูโรต่ำกว่าหรือเท่ากับระดับความเสี่ยง DD จำนวน 150 คน (ร้อนละ 37.5) ซึ่งทำการวิเคราะห์ข้อมูลด้วยวิธี Logistic regression ซึ่งได้ผลลัพธ์ออกมาว่า ตัวแปรข้อมูลทั่วไปได้แก่ ประเภทที่อยู่อาศัย สถานภาพ อาชีพ รายได้ต่อเดือน บุคลิกภาพแบบไม่มั่นคงทางอารมณ์และบุคลิกภาพแบบเปิดรับประสบการณ์ใหม่ โดยประเภทที่อยู่อาศัยประเภททาวน์เฮ้าและคอนโดจะส่งผลให้คะแนนการพิจารณาให้สินเชื่อลดลง 1.427 และ 1.311 เมื่อเทียบกับตัวแปรประเภทที่อยู่อาศัยแบบบ้านที่เป็น 0 อย่างมีนัยสำคัญ สถานภาพแต่งงานแล้วมีผลให้คะแนนการพิจารณาสินเชื่อเพิ่มขึ้น 0.663 เมื่อเทียบกับตัวแปร reference สถานภาพโสดที่เป็น 0 อย่างมีนัยสำคัญ อาชีพธุรกิจส่วนตัวส่งผลให้คะแนนการพิจารณาสินเชื่อลดลง 1.822 เมื่อเทียบกับตัวแปร Reference อาชีพข้าราชการที่เป็น 0 อย่างมีนัยสำคัญ และรายได้ต่อเดือนสูงมีผลดีต่อการพิจารณาให้สินเชื่อตามค่าเบต้าที่เป็นบวกในสมการ

Zhang and et al. [17] ได้ทำการพัฒนาโมเดลการให้คะแนนเครดิตสำหรับผู้ใช้ในระบบการเงินออนไลน์ โดยใช้ข้อมูลจากบริษัทการเงินออนไลน์จำนวน 24,660 ตัวอย่าง โดยมีผู้ใช้ที่มีความเสี่ยงที่จะมีโอกาสผิดนัดชำระอยู่ที่ 26.4% ซึ่งข้อมูลจะประกอบด้วย ข้อมูลพื้นฐานส่วนบุคคล ประวัติการกู้ยืม ข้อมูลบัตรเครดิต ประวัติการค้างชำระและหนี้คงค้าง โดยทำการเปรียบเทียบอัลกอริทึม Logistic Regression, Random Forest, CatBoost และ Weighted Voting Ensemble โดยผลลัพธ์ของการวิจัยวัดประสิทธิภาพจากค่า ROC โดย ผลลัพธ์เป็น 81%, 85%,87% แล 89% ตามลำดับ ซึ่งผลลัพธ์แสดงให้เห็นว่า Weighted voting ensemble ให้ผลลัพธ์ดีที่สุด โดยสามารถจำแนกผู้กู้ที่มีความเสี่ยงได้แม่นยำกว่าการใช้โมเดลเดี่ยว และยังช่วยลดความลำเอียงจากข้อมูลที่ไม่สมดุลได้อย่างมีประสิทธิภาพ

Bae and Kim [18] ได้นำเสนอโมเดลการพยากรณ์คะแนนเครดิตส่วนบุคคลโดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ภายใต้สภาพแวดล้อมการคำนวณแบบอัจฉริยะและแพร่หลาย (Smart Ubiquitous Computing Environment) โดยการวิจัยนี้จะใช้เทคนิค Ubiquitous Data Mining (UDM) ซึ่งรวมอัลกอริทึมหลายๆ ตัวเข้าด้วยกัน ซึ่งการวิจัยนี้จะใช้ข้อมูลเครดิต เช่น ประวัติการชำระหนี้, รายได้, อายุ, ประวัติการกู้ยืม เป็นต้น และข้อมูลสินเชื่อส่วนบุคคล เช่น จำนวนเงินกู้, ระยะเวลาการกู้ ประเภทสินเชื่อ อัตราดอกเบี้ย สถานะการชำระหนี้ จาก Financial Holding Company ในการเปรียบเทียบอัลกอริทึมจะใช้ Bayesian Frequency Matrix, Decision Tree, Neural Network rule extraction, Logistic Regression, Multilayer Perceptron และ วิธี UDM ของทุกอัลกอริทึมเข้าด้วยกัน โดยผลลัพธ์ของการวิจัย คือ การใช้วิธี UDM ของทุกอัลกอริทึมเข้าด้วยกัน ซึ่งมีค่า Accuracy 87.9% ค่า Precision 85.6% และ ค่า Recall 83.8% เป็นอัลกอริทึมที่มีผลลัพธ์ที่ดีที่สุด และผลลัพธ์ที่แย่ที่สุด คือ Logistic Regression ซึ่งมีค่า Accuracy 78.6% ค่า Precision 76.2% และ ค่า Recall 74.9%

Ampountolas and et al. [19] ได้พัฒนาโมเดลการให้คะแนนเครดิตสำหรับผู้กู้รายย่อย (micro-credit) โดยใช้เทคนิค Machine Learning เพื่อแก้ปัญหาการขาดข้อมูลเครดิตในกลุ่มผู้ไม่มีประวัติทางการเงิน โดยข้อมูลนำมาใช้พัฒนาจะเป็นข้อมูลลูกค้ารายย่อยที่หาได้ทั่วไป ประกอบด้วยข้อมูลพื้นฐานที่หาได้ง่าย เช่น:อายุ อาชีพ ที่อยู่ และ ข้อมูลประชากรอื่น ๆ ซึ่งข้อมูลที่ใช้จะไม่ใช้ข้อมูลจากเครดิตบูโรหรือประวัติสินเชื่อ ซึ่งชุดข้อมูลจะเหมาะสำหรับกลุ่มผู้กู้ที่ไม่มีข้อมูลเครดิตในระบบ ซึ่งตรงกับวัตถุประสงค์ในการวิจัย ซึ่งงานวิจัยนี้จะใช้อัลกอริทึม AdaBoost, XGBoost, Random Forest, Multilayer Perceptron และ Weighted voting ensemble ซึ่งผลลัพธ์ของการวิจัย พบว่า แบบจำลองที่ Weighted Voting Ensemble (การ Ensemble ของ 3 แบบจำลอง คือ Random Forest, XGBoost, AdaBoost) เป็นแบบจำลองที่มีประสิทธิภาพสูงสุด โดยมีค่า มีค่า Accuracy 80.0%

Minakawa and et al. [20] มุ่งพัฒนาโมเดลการพยากรณ์การล้มละลายของบริษัท (Firm Default) โดยใช้ Graph Neural Network (GNN) บนเครือข่ายระหว่างบริษัทขนาดใหญ่โดยจะใช้ข้อมูลจาก Tokyo Shoko Research Ltd. (TSR) เป็นฐานข้อมูลขนาดใหญ่ที่ครอบคลุมบริษัทในญี่ปุ่นมากกว่า 1 ล้านแห่ง โดยลักษณะของข้อมูลจะประกอบด้วย ยอดขาย จำนวนพนักงาน สินทรัพย์ อัตรากำไรและค่า Return of Equity เป็นต้น งานวิจัยนี้จะใช้อัลกอริทึม Logistic Regression, Random Forest, Xgboost, GraphSAGE แบบ Uniform Sampling และ GraphSAGE แบบ Gravity Sampling ซึ่งผลลัพธ์ของงานวิจัย คือ GraphSAGE แบบ Gravity Sampling โดยมีค่า Accuracy 81.2% ค่า F1 Score 79.5%

ชลลดา สุรศักดิ์ และ นิเวศ [21] ได้นำเสนอแบบจำลองการให้คะแนนสินเชื่อ และทดสอบประสิทธิภาพของแบบจำลองการให้คะแนนสินเชื่อโดยใช้เทคนิคเหมืองข้อมูล โดยทดสอบกับชุดข้อมูล 2 กลุ่ม คือ คะแนนสินเชื่อประเทศออสเตรเลียและประเทศเยอรมัน สร้างแบบจำลองการให้คะแนนสินเชื่อด้วยอัลกอริทึม 5 วิธี ได้แก่ ต้นไม้การตัดสินใจ การจำแนกกลุ่มนาอีฟเบย์โครงข่ายประสาท การถดถอยแบบโลจิสติกและซับพอร์ตเวคเตอร์แมชชีนทำการทดสอบประสิทธิภาพของแบบจำลองจากค่าความแม่นยำ ผลการทดสอบพบว่าแบบจำลองที่ดีที่สุดคือโครงข่ายประสาทเทียม

ทั้ง 2 กลุ่มข้อมูล โดยแบบจำลองมีประสิทธิภาพสูงสุดได้ประสิทธิภาพ 90.14% กับชุดข้อมูลจากประเทศออสเตรเลีย และ 90.08% กับชุดข้อมูลจากประเทศเยอรมันตามลำดับ

ปารดา [22] ได้ศึกษาปัจจัยที่ไม่ก่อให้เกิดรายได้ NPLsและการประยุกต์เหมืองข้อมูลในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อกรณีศึกษา ธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่ ใช้กระบวนการวิเคราะห์ข้อมูลมาตรฐาน เรียกว่า CrossIndustry Standard Process (CRIS-P) ประกอบด้วย 6ขั้น ตอน ได้แก่ เข้าใจปัญหาของธุรกิจ เข้าใจข้อมูล เตรียมข้อมูลพัฒนาแบบจำลอง การประเมิน และการนำไปใช้จริง ข้อมูลที่ใช้รวบรวมข้อมูลจากลูกค้าที่ได้รับสินเชื่อ Covid-19 ของธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่ ตั้งแต่เดือน กรกฎาคม 2564 ถึง ธันวาคม 2564 จำนวน 7,215ราย การเตรียมและคัดเลือกข้อมูลให้สมบูรณ์เหมาะสำหรับการทำเหมืองข้อมูลด้วยกระบวนการ Extract Transform Load (ETL) จากนั้นทำการโอนย้ายข้อมูลการลดขนาดของข้อมูลและการทำความสะอาดข้อมูล ใช้โปรแกรม WEKA เป็นเครื่องมือ วิเคราะห์หาความสัมพันธ์และใช้อัลกอริทึมการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ3แบบ ได้แก่ Decision Tree, Naïve Bayes และ Random Forest การทดสอบประสิทธิภาพอัลกอริทึมด้วยการหารากของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) เพื่อหาอัลกอริทึมที่ดีที่สุด และวัดผล ผลการศึกษา พบว่า ปัจจัยที่มีผลต่อการเกิดหนี้ที่ไม่ก่อให้เกิดรายได้ได้แก่อาชีพ สถานะภาพ ระดับการศึกษา รายได้ เพศ และ ผลการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ จากอัลกอริทึม Decision Tree (J48) มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 96.15% เมื่อเทียบกับอัลกอริทึม Regression และ Random Forest ถูกต้องร้อยละ 94.52 % และ96.10 % ตามลำดับ

ตารางที่ 2-1 งานวิจัยที่เกี่ยวข้อง

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
การสร้างแบบจำลองการพยากรณ์ลูกหนี้ที่ค้างชำระหนี้ของสินเชื่อเพื่อที่อยู่ [1]	2565	- SVM - kNN - XGBoost - Random Forest - DNN	ผลการวิจัยพบว่า XGBoost ให้ประสิทธิภาพสูงสุด โดยมีค่าความถูกต้องร้อยละ 86.84 ค่าแม่นยำเท่ากับ 0.87 ค่าระลึกเท่ากับ 0.87 และค่าการถ่วงดุลเท่ากับ 0.87
การวิเคราะห์ความเสี่ยงในการผิดนัดชำระของลูกหนี้บัตรเครดิต [2]	2566	- SVM - kNN - Naive Bayes - LR - Artificial neural network - XGBoost - GBM	ผลการศึกษาพบว่า การพัฒนาแบบจำลอง โดยการใช้เทคนิควิธี XGBoost ให้ค่าความไว (Recall) ที่มากที่สุด ซึ่งมีค่าเท่ากับ 0.97 มีค่า Accuracy เท่ากับ 0.37 และมีค่า F1-Score เท่ากับ 0.51
Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers [3]	2025	- Logistic Regression - Neural Networks - AdaBoost - XGBoost - LightGBM	การศึกษาพบว่า XGBoost มีประสิทธิภาพเหนือกว่าโมเดลอื่น ๆ โดยให้ค่า Accuracy สูงถึง 99.4%
Deep Learning for Credit Scoring do or don't. [4]	2021	- XGBoost - Multilayer Perceptron - Deep Belief Network - LR - Random Forest - SVM	งานวิจัยพบว่า XGBoost ให้ประสิทธิภาพดีที่สุดสำหรับการให้คะแนนเครดิตสูงสุด เมื่อเทียบกับทุกวิธีที่นำมาศึกษา โดยมีค่า AUC อยู่ที่ 98.7%

ตารางที่ 2-1 (ต่อ)

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
EWS-GCN: Edge Weight-Shared Graph Convolutional Network for Transactional Banking Data [5]	2020	- GCN - GraphSAGE - GAT - EGNN - RNN - LightGBM	ผลการวิจัยพบว่า คือ โมเดล EWS-GCN ซึ่งเป็นการ Hybrid ระหว่าง GCN และ RNN สามารถประมวลผลข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพที่สุดโดยผลลัพธ์ค่า ROC อยู่ที่ 89.26%
Applying Hybrid Graph Neural Networks to Strengthen Credit Risk Analysis [6]	2024	- GCN - SVM - Decision Tree - backpropagation - SGCNs	ผลลัพธ์ของการวิจัย คือ โมเดล SGCNs มีประสิทธิภาพสุงสุด โดยค่าความแม่นยำ (Accuracy) เฉลี่ยอยู่ที่ 87.52%, ค่า F1-score เฉลี่ย 93.13% และค่า G-mean: เฉลี่ย 85.59%
Improved Credit Risk Prediction Based on an Integrated Graph Representation Learning Approach with Graph Transformation [7]	2024	- kNN - GNN - LR - Random Forest - XGBoost	โมเดล kNN–GNN แบบไฮบริด ให้ผลลัพธ์ดีกว่าโมเดล Logistic Regression และ Random Forest โดยมีค่า Accuracy 88.9% ค่า F1 Score 85.7% และ ค่า AUC 91.3%
Temporal-Aware Graph Neural Network for Credit Risk Prediction [8]	2021	- Xgboost - GraphSAGE - LSTM - ST-GCN - GCN-LSTM - GAT	งานวิจัยพบว่า TemGNN เป็นแบบจำลองที่มีประสิทธิภาพสูงสุด โดยมีค่า Accuracy 84.3% ค่า F1 Score 85.6% และ ค่า AUC 87.9%

ตารางที่ 2-1 (ต่อ)

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
A Two-Stage Hybrid Credit Risk Prediction Model Based On Xgboost And Graph-Based Deep Neural Network [9]	2022	- XGBoost - DNN - GNN	ผลลัพธ์เฉลี่ยที่ดีที่สุดในด้านความแม่นยำ (Accuracy), ค่า F1-score และ G-mean ที่ 87.52%, 93.13% และ 85.59% ตามลำดับและความสามารถในการพยากรณ์ดีขึ้นจากค่าเฉลี่ยเดิมอยู่ที่ 6.14, 7.59 และ 6.18
Predicting Consumer Default: A Deep Learning Approach [10]	2019	- DNN - XGBoost - LR - Decision Trees - GBM	การ Hybird DNN และ GBM เป็นโมเดลที่มีผลลัพธ์ที่ดีที่สุด โดยค่า Accuracy มากกว่า 86% ทุกช่วงเวลา ค่า AUC-Score มากกว่า 92% และค่า Gini Coefficient 86%
การทำเหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อของ บริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน) [11]	2014	- Decision Tree, - Naive Bayes - Decision Rule,	ผลการวิจัยพบว่า Decision Tree ให้ผลลัพธ์ดีที่สุด เนื่องจากมีความถูกต้องมากที่สุด และสามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนมากที่สุดเมื่อเทียบกับอีกสองเทคนิค
โมเดลทำนายความสามารถในการชำระหนี้เงินกู้ของสมาชิกสำนักงานส่งเสริมสวัสดิการและสวัสดิการภาพครูและบุคลากรทางการศึกษาจังหวัดอุดรธานี [12]	2018	- Artificial neural network - Decision Tree, - Naive Bayes	ผลการวิจัยพบว่า โมเดลโครงข่ายประสาทเทียมมีประสิทธิภาพการทำนายสูงที่สุด มีความแม่นยำสูงสุดที่ 98.77

ตารางที่ 2-1 (ต่อ)

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
การใช้เหมืองข้อมูลสำหรับพิจารณาการให้สินเชื่อสำหรับธนาคาร [13]	2021	- SVM - MLP - Decision Tree	ผลการวิจัยพบว่า การพิจารณาเพื่ออนุมัติสินเชื่อของธนาคารด้วยวิธี SVM มีความถูกต้องสูงที่สุดและไม่จำเป็นต้องใช้กระบวนการคัดเลือกคุณลักษณะพิเศษ
การพยากรณ์ความล้มเหลวและภาวะกดดันทางการเงินของบริษัทในตลาดหลักทรัพย์แห่งประเทศไทยโดยแบบจำลอง CAMPBELL [14]	2018	- Descriptive Statistics - Binary Logistic Regression - Multicollinearity	ผลการวิจัยพบว่า มี 3 ตัวแปรอิสระที่สามารถพยากรณ์ภาวะความล้มเหลวและกดดันทางการเงินได้อย่างมีนัยสำคัญที่ระดับ 0.05ได้แม่นยำถึง 89.36% ไม่มี Type I error และมี Type II error เท่ากับ 12.5%
การประยุกต์เหมืองข้อมูลเพื่อพยากรณ์ปัจจัยที่ทำให้เกิดหนี้ค้าง ภายใต้การให้บริการด้านสินเชื่อโครงการสินเชื่อดิจิทัล [15]	2021	- Decision Tree - Naive Bayes	ผลการวิจัยพบว่า เทคนิคต้นไม้ตัดสินใจ (Decision Tree) มีค่าความถูกต้องในการทำนายมากที่สุด
แพลตฟอร์มการกู้ยืมระหว่างบุคคลกับบุคคลผ่านช่องทางออนไลน์จากข้อมูลทางเลือกแบบทดสอบบุคลิกภาพและแบบประเมินเจตคติต่อเงิน [16]	2021	- Correlation analysis - LR	ผลการวิจัยพบว่า ตัวแปรข้อมูลทั่วไปได้แก่ ประเภทที่อยู่อาศัย สถานภาพ อาชีพ รายได้ต่อเดือน บุคลิกภาพแบบไม่มั่นคงทางอารมณ์และบุคลิกภาพแบบเปิดรับประสบการณ์ใหม่ โดยประเภทที่อยู่อาศัยประเภททาวน์เฮาและคอนโดจะส่งผลให้คะแนนการพิจารณาให้สินเชื่อลดลง 1.427 และ 1.311 เมื่อเทียบกับตัวแปร

ตารางที่ 2-1 (ต่อ)

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
			ประเภทที่อยู่อาศัยแบบบ้านที่เป็น 0 อย่างมีนัยสำคัญ สถานภาพแต่งงานแล้วมีผลให้คะแนนการพิจารณาสินเชื่อเพิ่มขึ้น 0.663 เมื่อเทียบกับตัวแปรสถานภาพโสดที่เป็น 0 อย่างมีนัยสำคัญ อาชีพธุรกิจส่วนตัวส่งผลให้คะแนนการพิจารณาสินเชื่อลดลง 1.822 เมื่อเทียบกับตัวแปรอาชีพข้าราชการที่เป็น 0 อย่างมีนัยสำคัญ และรายได้ต่อเดือนสูงมีผลดีต่อการพิจารณาให้สินเชื่อตามค่าเบต้าที่เป็นบวกในสมการ
Research On Personal Credit Scoring Model Based On Multisource Data [17]	2021	- Random Forest - Catboost - LR - Weighted voting ensemble	ผลลัพธ์ของการวิจัยวัดประสิทธิภาพจากค่า ROC โดย ผลลัพธ์เป็น 81%, 85%,87% และ 89% ตามลำดับ ซึ่งผลลัพธ์แสดงให้เห็นว่า Weighted voting ensemble ให้ผลลัพธ์ดีที่สุด
A Personal Credit Rating Prediction Model Using Data Miningin Smart Ubiquitous Environments [18]	2015	- Bayesian Frequency Matrix, - Decision Tree, - Neural Network - Logistic Regression - Multilayer Perceptron	ผลลัพธ์ของการวิจัย คือ การใช้วิธี UDM ของทุกอัลกอริทึมเข้าด้วยกัน ซึ่งมีค่า Accuracy 87.9% ค่า Precision 85.6% และ ค่า Recall 83.8% เป็นอัลกอริทึมที่มีผลลัพธ์ที่ดีที่สุด

ตารางที่ 2-1 (ต่อ)

งานวิจัย	ปี	เทคนิค / แบบจำลอง	ผลการวิจัย
A Machine Learning Approach for Micro-Credit Scoring [19]	2021	- AdaBoost - XGBoost - Random Forest - Multilayer Perceptron - Weighted voting ensemble	ผลลัพธ์ของการวิจัย พบว่า แบบจำลองที่ Weighted voting ensemble ของ Random Forest, XGBoost, AdaBoost) เป็นแบบจำลองที่มีประสิทธิภาพสูงสุด โดยมีค่ามีค่า Accuracy 80.0%
Firm Default Prediction by GNN with Gravity-Model Informed Neighbor Sampling [20]	2021	- Logistic Regression - Random Forest - Xgboost - GraphSAGE-Uniform Sampling - GraphSAGE-Gravity Sampling	ผลลัพธ์ของงานวิจัย คือ GraphSAGE แบบ Gravity Sampling โดยมีค่า Accuracy 81.2% ค่า F1 Score 79.5%
การพัฒนาแบบจำลองการให้คะแนนสินเชื่อโดยใช้เทคนิคเหมืองข้อมูล [21]	2021	- SVM - Decision Tree, - Naive Bayes - LR - Artificial neural network	ผลการวิจัยพบว่า Artificial Neural Network ทั้ง 2 กลุ่มข้อมูล โดยแบบจำลองมีประสิทธิภาพสูงสุดได้ประสิทธิภาพ 90.14% กับชุดข้อมูลจากประเทศออสเตรเลีย และ 90.08% กับชุดข้อมูลจากประเทศเยอรมันตามลำดับ
การประยุกต์เหมืองข้อมูลในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อกรณีศึกษาธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่ [22]	2022	- Decision Tree, - Naive Bayes - Random Forest	ผลการวิจัยพบว่า Decision Tre (มีค่าถูกต้องสูงสุด คิดเป็นร้อยละ 96.15%เมื่อเทียบกับอัลกอริทึม Regression และ Random Forest ถูกต้องร้อยละ 94.52 และ 96.10ตามลำดับ

จากตารางที่ 2-1 งานวิจัยข้างต้นที่ผู้วิจัยได้ศึกษามา เห็นได้ว่ามีทฤษฎีและแนวคิดต่าง ๆ ที่ผู้วิจัยสามารถปรับใช้กับการพยากรณ์การค้างชำระของสินเชื่อบุคคลได้ ซึ่งอัลกอริทึมที่ผู้วิจัยสนใจ คือ อัลกอริทึมเอ็กซ์ตรีมกราเดียนต์บูสติ้ง (eXtreme Gradient Boosting : XGBoost), อัลกอริทึมการเรียนรู้เชิงลึกโครงข่ายประสาทเทียมกราฟคอนโวลูชัน Graph Convolutional Network (GCN) และวิธีการ Hybrid GCN และ XGBoost เพื่อหาอัลกอริทึมที่มีประสิทธิภาพดีที่สุดในการนำมาวิเคราะห์ลูกหนี้ค้าชำระของสินเชื่อบุคคล ซึ่งสามารถผลลัพธ์ไปใช้ประโยชน์ในการกำหนดแนวทาง นโยบายและวิธีรับมือเกี่ยวกับความเสี่ยงที่เกิดจากลูกหนี้ค้างชำระได้อย่างมีประสิทธิภาพมากขึ้น

พระจอมเกล้าพระนครเหนือ มหาวิทยาลัยเทคโนโลยี

บทที่ 3
วิธีดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับการพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคล โดยมีกรอบแนวคิดแสดงดังภาพที่ 1

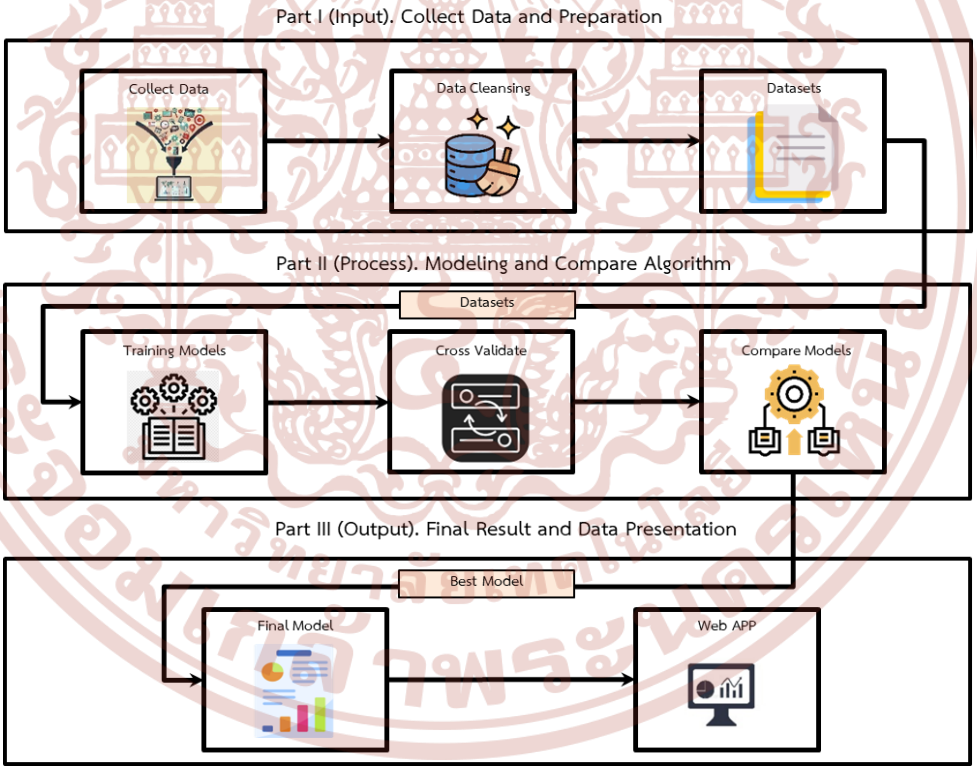
3.1 การออกแบบกรอบแนวความคิด (Conceptual Framework Design)

3.2 ขั้นตอนการรวบรวมข้อมูลและจัดเตรียมข้อมูล (Data Collection and Preparation)

3.3 การพัฒนา และวัดประสิทธิภาพของแบบจำลอง (Model Development & Evaluation)

3.4 การนำแบบจำลองไปใช้งาน (Model Deployment)

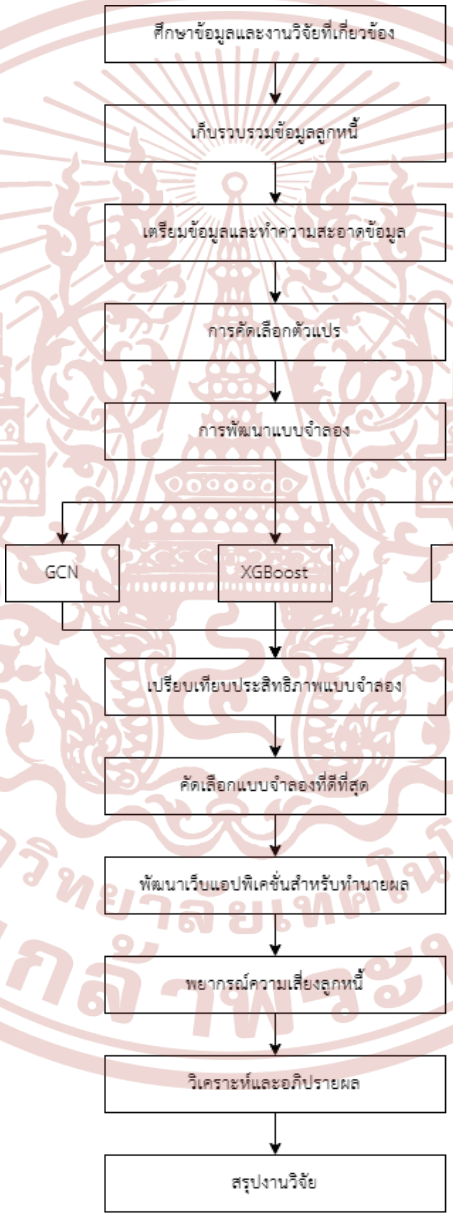
3.1 การออกแบบกรอบแนวความคิด (Conceptual Framework Design)



ภาพที่ 3-1 Conceptual Framework Design

จากภาพที่ 3-1 งานวิจัยนี้นำเสนอกรอบแนวคิดการวิเคราะห์พฤติกรรมการผ่อนชำระของสินเชื่อบุคคลโดยใช้เทคนิคเหมืองข้อมูล กระบวนการวิจัยแบ่งออกเป็น 3 ส่วนหลัก คือ การเก็บรวบรวมข้อมูลและการจัดเตรียมข้อมูล การทำแบบจำลองและทดสอบประสิทธิภาพของแบบจำลอง และ การนำแบบจำลองไปใช้งาน ในการเตรียมข้อมูลจะใช้ข้อมูลพอร์ตสินเชื่อบุคคลของธนาคารแห่งหนึ่ง โดยเป็นข้อมูล ณ ปี 2565 ถึง 2568 ซึ่งเป็นข้อมูลจะประกอบด้วย อายุ ยอดหนี้คงเหลือ รายได้

วงเงินและประวัติการชำระหนี้ เป็นต้น ซึ่งจะใช้อัลกอริทึม GCN, อัลกอริทึม XGBoost และการ Hybrid ทั้ง 2 อัลกอริทึม โดยเปรียบเทียบประสิทธิภาพโดยวิธี Cross Validate โดยวัดจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และ ค่าการถ่วงดุล (F-Measure) เพื่อคัดเลือกแบบจำลองที่มีประสิทธิภาพดีที่สุดไปประยุกต์ใช้กับข้อมูลต่อไป โดยมีขั้นตอนการดำเนินงาน แสดงดังภาพที่ 3-2



ภาพที่ 3-2 Workflow Design

3.2 ขั้นตอนการรวบรวมข้อมูลและจัดเตรียมข้อมูล (Data Collection and preparation)

3.2.1 การรวบรวมข้อมูล (Collect Data)

รวบข้อมูลที่จะนำมาวิเคราะห์ ข้อมูลที่ใช้เป็นข้อมูลพฤติกรรมการผ่อนชำระของสินเชื่อบุคคล โดยข้อมูลจะประกอบด้วย จุดประสงค์การกู้ วงเงินกู้ ยอดคงเหลือ อายุ อาชีพ รายได้ ค่าใช้จ่าย ประเภทของผู้กู้ ระยะเวลาขอกู้ สถานะสมรส ระดับการศึกษา ประเภทที่อยู่อาศัย อายุงานและกรรมสิทธิ์ในสถานประกอบการ จำนวน 50,000 รายการ โดยกำหนดสถานะ Performance 18 เดือน เพื่อแบ่งแยกบัญชีที่เป็นลูกค้าปกติ (Good) และลูกค้าที่ผิดนัดชำระ (Bad)

3.2.2 การทำความสะอาดข้อมูล (Data Cleansing)

หลังจากเตรียมข้อมูลสำหรับการวิเคราะห์พฤติกรรมการผ่อนชำระเพื่อพยากรณ์การค้างชำระหนี้ จะต้องมีการตรวจสอบความถูกต้องของข้อมูลและทำความสะอาดข้อมูล โดยจะตรวจสอบความครบถ้วน ของข้อมูล การเติมข้อมูลสูญหาย (Missing Value) การแปลงข้อมูลให้อยู่ในรูปแบบช่วงข้อมูลและการ Adjust เพื่อทำให้ข้อมูลสมบูรณ์มากที่สุดและอยู่ในรูปแบบที่เหมาะสมพร้อมสำหรับการใช้งาน

ข้อมูลได้รับการทำความสะอาดโดยการลบระเบียนที่มีข้อมูลสูญหายหรือไม่สมบูรณ์ออก หลังจากกระบวนการทำความสะอาดเสร็จสิ้น พบว่ามีข้อมูลที่สมบูรณ์คงเหลือจำนวนทั้งสิ้น 13,596 ระเบียน ประกอบด้วยลูกค้าที่ไม่ผิดนัดชำระหนี้จำนวน 10,000 ราย และลูกค้าที่ผิดนัดชำระหนี้จำนวน 3,596 ราย จากตัวแปรทั้งหมด

3.2.3 การแปลงข้อมูล (Data Transformation)

หลังจากทำความสะอาดข้อมูลเรียบร้อยแล้ว ข้อมูลจะถูกแปลงให้อยู่ในรูปแบบเชิงตัวเลขเพื่อรองรับการพัฒนาโมเดล โดยตัวแปรเชิงหมวดหมู่ (Categorical Variables) เพื่อให้ใช้ได้อย่างเหมาะสมและมีประสิทธิภาพ โดยการแปลงข้อมูล จะแปลงดังตารางที่ 3-1 ถึง ตารางที่ 3-11

ตารางที่ 3-1 การแปลงข้อมูลวงเงินกู้

วงเงินกู้	**Value (บาท)**
1	1,000-10,000
2	มากกว่า 10,000-15,000
3	มากกว่า 15,000-20,000
4	มากกว่า 20,000-25,000
5	มากกว่า 25,000-50,000
6	มากกว่า 50,000-100,000
7	มากกว่า 100,000

ตารางที่ 3-2 การแปลงข้อมูลยอดคงเหลือ

ยอดคงเหลือ	**Value (บาท)**
1	ไม่มียอดคงเหลือ
2	น้อยกว่าหรือเท่ากับ 5,000
3	มากกว่า 5,000-10,000
4	มากกว่า 10,000-30,000
5	มากกว่า 30,000-50,000
6	มากกว่า 50,000-100,000
7	มากกว่า 100,000

ตารางที่ 3-3 การแปลงข้อมูลรายได้

รายได้	**Value (บาท)**
1	น้อยกว่าหรือเท่ากับ 10,000
2	มากกว่า 10,000-20,000
3	มากกว่า 20,000-30,000
4	มากกว่า 30,000-50,000
5	มากกว่า 50,000-100,000
6	มากกว่า 100,000

ตารางที่ 3-4 การแปลงข้อมูลระดับการศึกษา

ระดับการศึกษา	**Value**
0	ไม่มีการศึกษา
1	ประถมต้น
2	ประถมปลาย
3	มัธยมต้น
4	มัธยมปลาย/ปวช.
5	ปวส./อนุปริญญา
6	ปริญญาตรี
7	ปริญญาโท
8	ปริญญาเอก

ตารางที่ 3-5 การแปลงข้อมูลอายุ

อายุ	**Value (ปี)**
1	น้อยกว่า 25 ปี
2	26-35 ปี
3	36-40 ปี
4	41-45 ปี
5	46-50 ปี
6	51-55 ปี
7	มากกว่า 55 ปี

ตารางที่ 3-6 การแปลงข้อมูลสถานะสมรส

สถานะสมรส	**Value**
1	สมรสจดทะเบียน
2	โสด
3	หย่า
4	สมรสไม่มีทะเบียน
5	หม้าย

ตารางที่ 3-7 การแปลงข้อมูลระยะเวลาขอกู้

ระยะเวลาขอกู้	**Value**
1	น้อยกว่า 2 ปี
2	2-3 ปี
3	มากกว่า 3-4 ปี
4	มากกว่า 4-5 ปี
5	มากกว่า 5-8 ปี
6	8 ปีขึ้นไป

ตารางที่ 3-8 การแปลงข้อมูลจุดประสงค์การกู้

จุดประสงค์การกู้	Value
1	เพื่อใช้ส่วนตัว
2	เงินทุนหมุนเวียนธุรกิจระหว่างประเทศ
3	refinance
4	ซื้อที่ดิน/สิ่งปลูกสร้างในการทำธุรกิจ
5	ซื้ออุปกรณ์ในการทำธุรกิจ
6	ลงทุนในอื่นๆในการทำธุรกิจ
7	ลงทุนในหลักทรัพย์
8	ลงทุนในอื่นๆ
9	เพื่ออื่นๆ
10	ทุนหมุนเวียน

ตารางที่ 3-9 การแปลงข้อมูลอาชีพ

อาชีพ	Value
0	ว่างงาน
5	ผู้ประกอบธุรกิจ
6	ประกอบอาชีพอิสระ
7	ผู้รับจ้าง
8	เกษตรกร

ตารางที่ 3-10 การแปลงข้อมูลรายจ่าย

รายได้	Value (บาท)
1	น้อยกว่าหรือเท่ากับ 10,000
2	มากกว่า 10,000-20,000
3	มากกว่า 20,000-30,000
4	มากกว่า 30,000-50,000
5	มากกว่า 50,000-70,000
6	มากกว่า 70,000

ตารางที่ 3-11 การแปลงข้อมูลอายุงาน

อายุงาน	**Value**
1	1ปี
2	2-3ปี
3	<3-5ปี
4	<5-8 ปี
5	<8-10ปี
6	<10-15ปี
7	<15-20ปี
8	<20ปี

ตารางที่ 3-12 การแปลงข้อมูลกรรมสิทธิ์ในสถานประกอบการ

กรรมสิทธิ์ในสถานประกอบการ	**Value**
1	ไม่มีกรรมสิทธิ์
2	เป็นเจ้าของ
3	เช่า
4	เป็นลูกจ้างชั่วคราว

ตารางที่ 3-13 การแปลงข้อมูลประเภทที่อยู่อาศัย

ประเภทที่อยู่อาศัย	**Value**
1	ไม่มีกรรมสิทธิ์
2	เป็นเจ้าของ
3	เช่า
4	เป็นลูกจ้างชั่วคราว

ตารางที่ 3-14 การแปลงข้อมูลประเภทผู้กู้

ประเภทผู้กู้	**Value**
1	ผู้กู้คนเดียว
2	มีผู้กู้ร่วม
3	มีผู้กู้ร่วมมากว่า 1 คนขึ้นไป

3.2.4 การปรับสมดุลข้อมูล (Data Balancing)

หลังการทำความสะอาดและแปลงรูปแบบข้อมูลแล้ว ชุดข้อมูลยังคงมีปัญหาความไม่สมดุลของคลาส (Class Imbalance) อย่างชัดเจน โดยกลุ่มลูกค้าที่ไม่ผิดนัดชำระหนี้มีสัดส่วนมากกว่ากลุ่มลูกค้าที่ผิดนัดชำระหนี้ เพื่อแก้ไขปัญหาดังกล่าว จึงได้นำเทคนิค Synthetic Minority Over-sampling Technique (SMOTE) มาใช้ในการปรับสมดุลของข้อมูล SMOTE เป็นวิธีการสร้างข้อมูลสังเคราะห์ (Synthetic Samples) โดยอาศัยการประมาณค่าระหว่างข้อมูลตัวอย่างเดิมของกลุ่มคลาสส่วนน้อย ทำให้ได้การกระจายตัวของข้อมูลลูกค้าที่ผิดนัดชำระหนี้ที่มีความสมจริงมากขึ้น และช่วยเพิ่มประสิทธิภาพในการเรียนรู้ของโมเดล

ในการศึกษานี้ จำนวนระเบียนของลูกค้าที่ผิดนัดชำระหนี้ซึ่งมีอยู่ 3,596 ระเบียน ได้รับการเพิ่มจำนวนให้เท่ากับกลุ่มลูกค้าที่ไม่ผิดนัดชำระหนี้ซึ่งมี 10,000 ระเบียน ส่งผลให้มีการสร้างข้อมูลสังเคราะห์ของกลุ่มลูกค้าที่ผิดนัดชำระหนี้เพิ่มเติมจำนวน 6,404 ระเบียน

3.2.5 การคัดเลือกตัวแปรที่เหมาะสมกับการพัฒนาแบบจำลอง

หลังจากได้ข้อมูลที่สมบูรณ์แล้ว ผู้วิจัยได้ทำการคัดเลือกตัวแปรที่จากทั้งหมด 14 ตัวแปร คือ จุดประสงค์การกู้ วงเงินกู้ ยอดคงเหลือ อายุ อาชีพ รายได้ ค่าใช้จ่าย ประเภทของผู้กู้ ระยะเวลาขอกู้ สถานะสมรส ระดับการศึกษา ประเภทที่อยู่อาศัย อายุงานและ กรรมสิทธิ์ในสถานประกอบการ โดยคัดเลือกเฉพาะตัวแปรที่เหมาะสมในการพัฒนาแบบจำลอง

ตารางที่ 3-15 ค่าความสำคัญของตัวแปร

ตัวแปร	F-Score
วงเงินกู้	888.94
อายุ	748.67
ระยะเวลาการกู้	567.39
สถานะสมรส	405.17
ระดับการศึกษา	273.08
กรรมสิทธิ์ในสถานประกอบการ	169.79
ยอดคงเหลือ	136.60
รายได้	113.69
อาชีพ	109.74
อายุงาน	52.58
จุดประสงค์การกู้	50.38
ประเภทที่อยู่อาศัย	33.57
รายจ่าย	16.77
ประเภทของผู้กู้	11.52

โดยผู้วิจัยจะคัดเลือกตัวแปรสำหรับการสร้างแบบจำลองโดยกำหนดเกณฑ์ (Threshold) ของค่า F-Score มากกว่า 50 เพื่อเลือกตัวแปรที่มีความสำคัญต่อแบบจำลอง มาใช้ในการพัฒนาแบบจำลอง ซึ่งตัวแปรสุดท้ายที่ใช้พัฒนาแบบจำลองมีจำนวน 11 ตัวแปร ได้แก่ วัตถุประสงค์ของการกู้เงิน วงเงินกู้ ยอดคงเหลือ อายุ อาชีพ รายได้ ระยะเวลาการกู้ สถานภาพสมรส ระดับการศึกษา อายุงานและกรรมสิทธิ์ในสถานประกอบการ การกำหนดเกณฑ์ดังกล่าวช่วยลดความซับซ้อนของโมเดล ลดปัญหา overfitting และเพิ่มความสามารถในการอธิบายผลลัพธ์ โดยตัวแปรที่ถูกเลือกครอบคลุมทั้งด้านการเงิน ประชากร อาชีพ และพฤติกรรมของผู้กู้ ซึ่งเป็นปัจจัยสำคัญในการประเมินความเสี่ยงด้านเครดิต

3.3 การพัฒนา และวัดประสิทธิภาพของแบบจำลอง (Model Development & Evaluation)

ทำการเปรียบเทียบอัลกอริทึม 3 อัลกอริทึม ได้แก่ อัลกอริทึมโครงข่ายประสาทเทียมกราฟคอนโวลูชัน (Graph Convolutional Network : GCN) อัลกอริทึมเอ็กตรีมกราเดี้ยนบูทติ้ง (Extreme Gradient Boosting : XGBoost) และการ Hybrid ทำการปรับค่าพารามิเตอร์ตามที่กำหนดไว้ เพื่อคัดเลือกอัลกอริทึมที่ดีที่สุดไปสร้างแบบจำลองในการพยากรณ์ลูกหนี้ค้างชำระ และทดสอบประสิทธิภาพของแบบจำลอง มีรายละเอียดของแต่ละขั้นตอนดังต่อไปนี้

3.3.1 การพัฒนาโมเดลด้วยอัลกอลิทึม GCN

หลักการของ GCN เริ่มจาก “การสร้างกราฟ” โดยใช้แนวคิดหลักคือ ความใกล้เคียงเชิงตัวเลขผ่าน k-Nearest Neighbors มาสร้างความสัมพันธ์ระหว่างกลุ่มตัวอย่าง [7] เมื่อได้กราฟโค้ดจะมองข้อมูลแต่ละแถวเป็นโหนด และเส้นเชื่อมคือบริบทของข้อมูลนั้น โดยการทำนายขึ้นอยู่กับ Embedding ที่สะท้อนโครงสร้างความสัมพันธ์ในกราฟเป็นหลัก แสดงดังในสมการที่ 3-1

$$\hat{y}_i = \text{softmax}\left(\mathbf{W}_{out}\,\sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}\mathbf{W})_i\right) \tag{3-1}$$

โดยที่ $\mathbf{X}$ คือ เมทริกซ์ฟีเจอร์เริ่มต้นของทุกโหนด (เช่น รายได้, อายุ, ยอดเงิน)

$\mathbf{A}$ คือ adjacency matrix ของกราฟ แสดงความเชื่อมโยงระหว่างโหนด

$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ คือ adjacency matrix ที่เพิ่ม self-loop

$\tilde{\mathbf{D}}$ คือ degree matrix ของ $\tilde{\mathbf{A}}$ ใช้สำหรับ normalize อิทธิพลของเพื่อนบ้าน

$\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$ คือ ตัวดำเนินการที่ทำให้การรวมข้อมูลจากเพื่อนบ้านมีความสมดุล

$\mathbf{W}$ คือ weight matrix ที่เรียนรู้ได้ของชั้น GCN

$\sigma(\cdot)$ คือ ฟังก์ชันไม่เชิงเส้น เช่น ReLU

$\mathbf{W}_{out}$ คือ weight ของชั้นจำแนก (classifier layer)

$\text{softmax}(\cdot)$ คือ แปลงผลลัพธ์เป็นความน่าจะเป็นของแต่ละคลาส

ตารางที่ 3-16 ค่าพารามิเตอร์สำหรับแบบจำลอง GCN

Parameter	Value
ค่า K (Number of Nearest Neighbors)	3, 5, 7, 9, 11, 13
Dropout Rate	0.1, 0.2, 0.5
Epochs	60, 100, 150
Hidden Dimension	32, 64, 128

3.3.2 การพัฒนาโมเดลด้วยอัลกอลิทึม XGBoost

สร้างโมเดลจากต้นไม้ตัดสินใจหลายต้นเรียงต่อกัน ต้นไม้แต่ละต้นไม่ได้เรียนรู้ข้อมูลใหม่ทั้งหมด แต่ถูกฝึกขึ้นมาเพื่อแก้ไขข้อผิดพลาดที่โมเดลก่อนหน้าทำไว้ ทำให้การเรียนรู้ค่อย ๆ ดีขึ้นเป็นลำดับ โดยระหว่างการฝึก โมเดลจะประเมินคุณภาพของการทำนายด้วยค่า Loss หรือ Metric ต่าง ๆ เพื่อควบคุมไม่ให้โมเดลซับซ้อนเกินไป เมื่อฝึกเสร็จ โมเดล XGBoost จะให้ผลลัพธ์เป็นค่าความน่าจะเป็นของคลาสเป้าหมาย ซึ่งเปิดโอกาสให้ปรับจุดตัด (Threshold) ได้ตามบริบทการใช้งาน แนวคิดทั้งหมดนี้ทำให้ XGBoost เหมาะกับข้อมูลเชิงตารางที่มีความสัมพันธ์ไม่เชิงเส้น และมีปฏิสัมพันธ์ระหว่างฟีเจอร์จำนวนมาก [4] แสดงดังในสมการที่ 3-2

$$\hat{y}_i = \sigma\left(\sum_{m=1}^{M} f_m(\mathbf{x}_i)\right), \qquad f_m \in \mathcal{F}_{tree} \tag{3-2}$$

โดยที่ $\mathbf{x}_i$ คือ เวกเตอร์ฟีเจอร์ของตัวอย่างที่ i

$f_m(\cdot)$ คือ ต้นไม้ตัดสินใจลำดับที่ m

M คือ จำนวนต้นไม้ทั้งหมดในโมเดล

$\mathcal{F}_{tree}$ คือ เซตของฟังก์ชันที่เป็น Decision Tree

$\sum_{m=1}^{M} f_m(\mathbf{x}_i)$ คือ การรวมผลลัพธ์จากต้นไม้ทุกต้น (boosting)

$\sigma(\cdot)$ คือ sigmoid function ใช้แปลงค่า logit เป็นความน่าจะเป็น

ตารางที่ 3-17 ค่าพารามิเตอร์สำหรับแบบจำลอง XGBoost

Parameter	Value
Number of Boosting Trees	200, 400, 600, 800
Learning Rate	0.03, 0.05, 0.1
Maximum Tree Depth	4, 6, 8, 10
Minimum Child Weight	1, 2, 4
Subsample Ratio of Training Instances	0.1, 0.2, 0.5, 1
Column Subsample Ratio per Tree	0.01, 0.02, 0.05, 1
Alpha	0, 0.1
Lambda	0, 1
Gamma	0, 0.1

3.3.3 การพัฒนาโมเดลด้วยอัลกอลิทึม Hybrid

หลักการของ hybrid คือ “แยกหน้าที่กันทำ” โดยใช้ GCN เพื่อสกัดบริบทเชิงโครงสร้าง และใช้ XGBoost เพื่อทำการตัดสินใจเชิงพยากรณ์ โดยขั้นตอนการ hybrid คือ การนำ Embedding จาก GCN มาต่อรวมกับฟีเจอร์ดั้งเดิม ก่อนส่งเข้า XGBoost นี่สะท้อนแนวคิดเชิงว่า Embedding จากกราฟให้ภาพรวมของบริบท ขณะที่ ฟีเจอร์ดิบยังให้รายละเอียดเชิงตัวเลขที่สำคัญ การรวมทั้งสองส่วนทำให้โมเดลสุดท้ายไม่สูญเสียข้อมูลด้านใดด้านหนึ่งจากมุมมองเชิงการพัฒนาโมเดล ซึ่งการใช้ GCN เป็น Feature Extractor เชิงโครงสร้าง และใช้ XGBoost เป็น Decision Model เชิงพยากรณ์ ซึ่งเป็นการผสานข้อดีของ Deep Learning บนกราฟ และ Tree-based Model [9] บนข้อมูลตารางเข้าด้วยกันอย่างเป็นระบบ โดยค่าพารามิเตอร์ที่จะ Set ในโมเดล Hybird คือค่าที่ดีที่สุดที่จะได้มาจาก GCN และ XGBoost ซึ่งแนวคิดหลักของการ Hybrid คือ GCN สร้าง Representation เชิงโครงสร้าง ส่วน XGBoost ใช้ Representation เพื่อเรียนรู้เงื่อนไขการทำนายที่ซับซ้อน แสดงดังในสมการที่ 3-3

$$\hat{y}_i = \sigma\left(\sum_{m=1}^{M} f_m\Big([\mathbf{z}_i \,\|\, \mathbf{x}_i]\Big)\right), \qquad \mathbf{z}_i = \mathrm{GCN}(\mathbf{x}_i, \mathcal{G}) \tag{3-3}$$

โดยที่ $\mathbf{z}_i$ คือ embedding ของโหนดที่ i ที่ได้จาก GCN สะท้อนทั้งฟีเจอร์ของตัวเองและบริบทจากเพื่อนบ้าน

$\mathcal{G}$ คือ โครงสร้างกราฟ (โหนด + เส้นเชื่อม) ที่ใช้ใน GCN

$\mathbf{x}_i$ คือ ฟีเจอร์ดั้งเดิมของตัวอย่างที่ i (มักผ่านการ standardize)

$[\mathbf{z}_i \,\|\, \mathbf{x}_i]$ คือ การ Concat Embedding ฟีเจอร์ดั้งเดิมเพื่อเก็บข้อมูลทั้งเชิงโครงสร้างและเชิงตัวเลข

$f_m(\cdot), M, \sigma(\cdot)$ คือ มีความหมายเดียวกับ XGBoost only

3.3.4 การทดสอบประสิทธิภาพ

ในการประเมินประสิทธิภาพของโมเดลจะใช้ตัวชี้วัดมาตรฐานสำหรับงานจำแนกประเภท ได้แก่ Accuracy, Precision, Recall และ F1-score ซึ่ง ตัวชี้วัดเหล่านี้ถูกเลือกใช้เพื่อให้สามารถประเมินประสิทธิภาพของการจำแนกได้อย่างครอบคลุม ซึ่งมีความหมายดังนี้

Accuracy ใช้วัดความถูกต้องโดยรวมของแบบจำลอง โดยคำนวณจากสัดส่วนของจำนวนตัวอย่างที่ถูกทำนายได้อย่างถูกต้องทั้งหมด เมื่อเทียบกับจำนวนตัวอย่างทั้งหมด ดังแสดงในสมการที่ 3-4

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3-4}$$

โดยที่ TP (True Positive) หมายถึง จำนวนตัวอย่างที่เป็นบวกและถูกทำนายเป็นบวกอย่างถูกต้อง,
TN (True Negative) หมายถึง จำนวนตัวอย่างที่เป็นลบและถูกทำนายเป็นลบอย่างถูกต้อง,
FP (False Positive) หมายถึง จำนวนตัวอย่างที่เป็นลบแต่ถูกทำนายผิดเป็นบวก, และ
FN (False Negative) หมายถึง จำนวนตัวอย่างที่เป็นบวกแต่ถูกทำนายผิดเป็นลบ

Precision ใช้ในการประเมินความน่าเชื่อถือของการทำนายคลาสบวก โดยวัดสัดส่วนของตัวอย่างที่ถูกทำนายเป็นบวกอย่างถูกต้อง เมื่อเทียบกับจำนวนตัวอย่างทั้งหมดที่ถูกทำนายว่าเป็นบวก ดังแสดงในสมการที่ 3-5

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3-5}$$

Recall หรือที่เรียกว่า Sensitivity ใช้สะท้อนความสามารถของแบบจำลองในการตรวจจับตัวอย่างที่เป็นบวกจริง โดยคำนวณจากสัดส่วนของตัวอย่างที่เป็นบวกและถูกทำนายอย่างถูกต้อง เมื่อเทียบกับจำนวนตัวอย่างที่เป็นบวกทั้งหมด ดังแสดงในสมการที่ 3-6

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3-6}$$

ค่า F1-score เป็นค่าเฉลี่ยเชิงฮาร์มอนิกของ Precision และ Recall ซึ่งช่วยให้การประเมินมีความสมดุล โดยคำนึงถึงทั้งข้อผิดพลาดแบบ False Positive และ False Negative พร้อมกัน สมการนิยามของ F1-score ดังแสดงในสมการที่ 3-7

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3-7}$$

3.4 การนำแบบจำลองไปใช้งาน (Model Deployment)

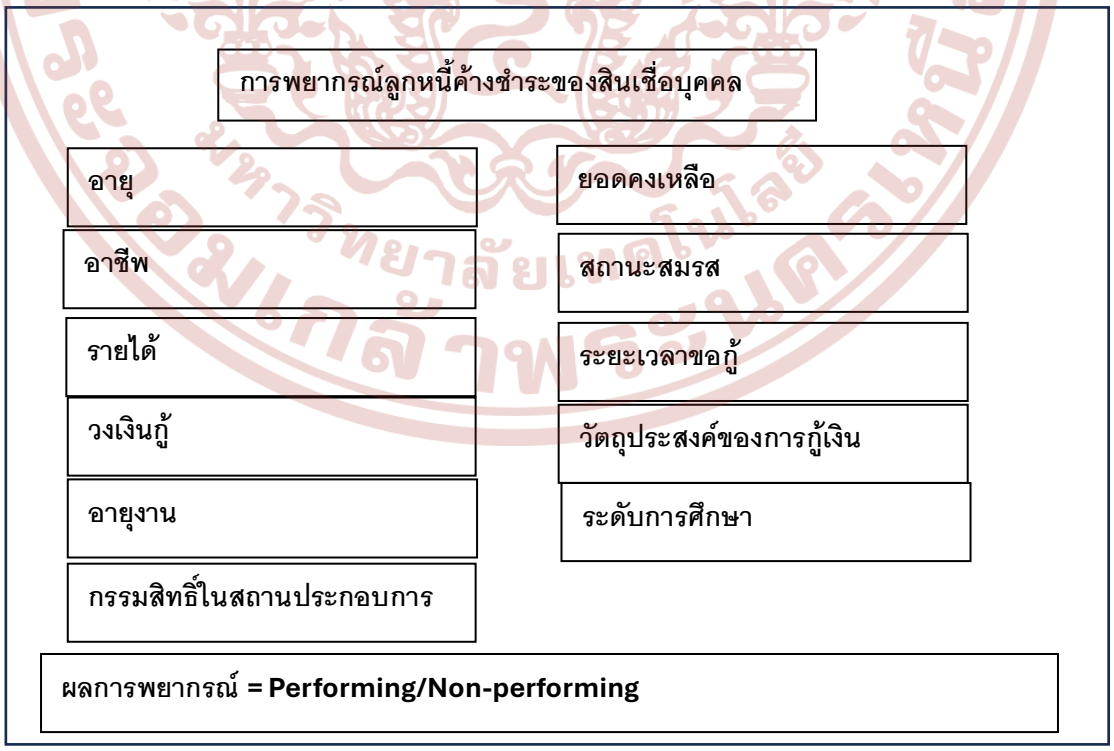
การนำแบบจำลองไปประยุกต์ใช้งานจริง (Deployment) เป็นขั้นตอนสุดท้ายของงานวิจัย เพื่อให้ผู้ปฏิบัติงานและหน่วยงานที่เกี่ยวข้องสามารถนำผลลัพธ์ไปใช้ประโยชน์ในการประเมินความเสี่ยงได้อย่างเป็นรูปธรรมและเกิดประสิทธิภาพสูงสุด ผู้วิจัยได้ดำเนินการออกแบบและพัฒนาระบบการพยากรณ์ตามขั้นตอนดังต่อไปนี้

3.4.1 การบันทึกสถานะแบบจำลอง (Model Serialization)

ในขั้นตอนแรก ผู้วิจัยทำการตรวจสอบความสมบูรณ์ของแบบจำลองขั้นสุดท้ายอีกครั้ง โดยยืนยันลำดับตัวแปร (Feature Order) และกระบวนการเตรียมข้อมูล (Preprocessing) ให้สอดคล้องกับขั้นตอนการฝึกแบบจำลอง จากนั้นได้ทำการบันทึกแบบจำลองในรูปแบบไฟล์ .pkl ด้วยเทคนิค Model Serialization เพื่อให้สามารถเรียกใช้งานภายหลังได้โดยไม่ต้องฝึกโมเดลใหม่ ขั้นตอนนี้มีความสำคัญในการแยกกระบวนการพัฒนาโมเดลออกจากกระบวนการให้บริการจริง

3.4.2 การออกแบบหน้าเว็บ (User Interface)

ผู้วิจัยได้ออกแบบและพัฒนาเว็บแอปพลิเคชันสำหรับใช้ในการทำนายผลการจัดประเภทลูกหนี้ โดยผู้ใช้งานสามารถเลือกกำหนดคุณลักษณะของลูกหนี้ตามตัวแปรที่กำหนดไว้ในแบบจำลองการพยากรณ์ ซึ่งค่าของตัวแปรแต่ละตัวจะถูกแปลงให้อยู่ในช่วงค่าที่สอดคล้องกับรูปแบบข้อมูลที่ใช้ในการฝึกแบบจำลอง เมื่อผู้ใช้งานระบุข้อมูลของลูกหนี้ครบถ้วนตามตัวแปรที่กำหนด ระบบจะทำการประมวลผลผ่านแบบจำลองที่พัฒนาขึ้น และแสดงผลการทำนายว่าลูกหนี้รายดังกล่าวมีแนวโน้มจัดอยู่ในกลุ่ม Performing (ลูกหนี้ปกติ) หรือ Non-performing (ลูกหนี้ผิดนัดชำระ) ทั้งนี้ รูปแบบของหน้าเว็บแอปพลิเคชันแสดงดังภาพประกอบ



ภาพที่ 3-3 User Interface Design

3.4.3 การติดตั้งระบบประมวลผลและการแสดงผลลัพธ์

ผู้วิจัยได้สร้างโปรเจคบน Firebase Console และเปิดใช้งานบริการหลัก ได้แก่ Firebase Hosting และ Cloud Functions ซึ่ง Firebase Hosting ทำหน้าที่เผยแพร่หน้าเว็บ (Frontend) ให้สามารถเข้าถึงผ่านอินเทอร์เน็ต และ Cloud Functions ทำหน้าที่เป็น Backend API สำหรับประมวลผลข้อมูลและเรียกใช้แบบจำลอง โดยลำดับการทำงานภายในประกอบด้วย

1. รับข้อมูลผ่าน HTTP Request จากหน้าเว็บ
2. ดำเนินการแปลงข้อมูลให้อยู่ในรูปแบบที่แบบจำลองรองรับ
3. โหลด Final Model ที่ถูกบันทึกไว้
4. ทำการการพยากรณ์ (Prediction)

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

บทที่ 4
ผลการดำเนินงานวิจัย

ผลการดำเนินงานวิจัยการพัฒนาแบบจำลองการพยากรณ์การผิดนัดชำระหนี้ของลูกค้าสินเชื่อส่วนบุคคล โดยประยุกต์ใช้เทคนิคเหมืองข้อมูล มีรายละเอียดดังต่อไปนี้

4.1 ผลการเก็บรวบรวมข้อมูลและจัดเตรียมข้อมูล

4.2 ผลการพัฒนา และทดสอบประสิทธิภาพของแบบจำลอง

4.3 ผลการนำแบบจำลองไปใช้งาน

4.1 ผลการเก็บรวบรวมข้อมูลและการจัดเตรียมข้อมูล

จากการเก็บรวบรวมข้อมูลและความสะอาดข้อมูลการผ่อนชำระของกลุ่มตัวอย่างแล้ว พบว่าข้อมูลที่เหมาะสมจะนำมาพัฒนาแบบจำลอง มีจำนวนทั้งสิ้น 13,596 รายการและตัวแปรที่หมาะสมจำนวน 11 ตัวแปร ได้แก่ วัตถุประสงค์ของการกู้เงิน วงเงินกู้ ยอดคงค้าง อายุ อาชีพ รายได้ ระยะเวลาการกู้ สถานภาพสมรส ระดับการศึกษา อายุงานและกรรมสิทธิ์ในสถานประกอบการ ซึ่งข้อมูลครอบคลุมระยะเวลาตั้งแต่ มกราคม 2565 – กุมภาพันธ์ 2568 ซึ่งเป็นข้อมูลเปิดบัญชีตั้งแต่ มกราคม 2565 - กันยายน 2566 และติดตามความสามารถในการผ่อนชำระ 18 เดือน จนถึง กุมภาพันธ์ 2568 จากนั้นนำมาผ่านกระบวนการแปลงข้อมูลและการจัดสมดุลข้อมูล (Data Transformation & Data Balancing) ตามระเบียบวิธีวิจัยที่กำหนดไว้ในบทที่ 3 ทำให้ได้ข้อมูลสำหรับฝึกฝนแบบจำลองเพิ่มขึ้นจำนวน 6,404 ระเบียน ซึ่งข้อมูลสุดท้ายสำหรับฝึกฝนแบบจำลองมีจำนวน 20,000 ระเบียน โดยแบ่งเป็นข้อมูลลูกหนี้ปกติจำนวน 10,000 ระเบียนและลูกหนี้ผิดนัดชำระหนี้จำนวน 10,000 ระเบียน เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ของแบบจำลองและป้องกันปัญหาความเอนเอียง (Bias) ผู้วิจัยได้ดำเนินการเตรียมข้อมูลเพิ่มเติมสำหรับทดสอบประสิทธิภาพของแบบจำลองที่ผ่านฝึกฝน โดนข้อมูลจะแบ่งออกเป็น 2 ส่วนหลัก ได้แก่

1. ชุดข้อมูลสำหรับการพัฒนาแบบจำลอง (Training Modeling Dataset) จำนวน 20,000 ระเบียน ซึ่งเป็นข้อมูลที่ผ่านการทำความสะอาด คัดเลือกและจัดความสมดุลของข้อมูลด้วยวิธี SMOTE (Synthetic Minority Over-sampling Technique) เพื่อนำเข้าสู่กระบวนการฝึกสอนอัลกอริทึม โดยแบ่งลูกหนี้ปกติและเสีย เป็นคลาสละ 10,000 ระเบียน

2. ชุดข้อมูลสำหรับทดสอบแบบจำลอง (Unseen Dataset) จำนวน 500 ระเบียน ซึ่งแบ่งลูกหนี้ปกติและเสีย เป็นคลาสละ 250 ระเบียน ซึ่งข้อมูลส่วนนี้จะเป็นข้อมูลที่ไม่ได้มีส่วนในกระบวนการฝึกสอนแบบจำลอง เพื่อใช้เป็นตัวแทนของข้อมูลจริงในการประเมินประสิทธิภาพขั้นสุดท้าย

ตารางที่ 4-1 สรุปข้อมูลที่ใช้การพัฒนาและทดสอบแบบจำลอง

ประเภทของข้อมูล	ลูกหนี้ปกติ	ลูกหนี้ผิดนัดชำระ	จำนวนลูกหนี้ทั้งหมด
ข้อมูลฝึกฝนแบบจำลอง	10,000 ระเบียน	10,000 ระเบียน	20,000 ระเบียน
ข้อมูล Unseen	250 ระเบียน	250 ระเบียน	500 ระเบียน

ตารางที่ 4-2 ตัวอย่างข้อมูลที่ใช้ในการฝึกฝนแบบจำลอง

NO	วงเงินกู้	เงินคงเหลือ	วัตถุประสงค์การกู้	ระยะเวลาขอกู้	อายุ	สถานะสมรส	อาชีพ	ระดับการศึกษา	รายได้	ระยะเวลาการทำงาน	กรรมสิทธิ์ในสถานประกอบการ
1	2	1	1	1	6	4	7	2	3	2	4
2	2	1	1	1	4	2	7	2	3	7	2
3	2	1	1	1	5	4	7	2	2	6	2
4	2	2	1	4	3	1	6	4	5	4	1
5	2	1	1	1	4	4	6	2	2	7	4
6	7	7	10	6	4	2	5	5	6	7	1
7	2	1	1	1	5	2	7	4	3	4	4
8	2	2	1	1	4	1	7	4	2	5	4
9	2	1	1	1	2	2	7	5	2	3	4
10	2	1	1	1	7	2	5	2	4	7	2
11	2	1	1	1	7	3	7	3	3	7	1
12	2	3	1	1	3	1	7	4	1	6	4
13	3	4	1	4	3	1	6	4	4	1	2

NO	วงเงินกู้	เงินคงเหลือ	วัตถุประสงค์การกู้	ระยะเวลาขอกู้	อายุ	สถานะสมรส	อาชีพ	ระดับการศึกษา	รายได้	ระยะเวลาการทำงาน	กรรมสิทธิ์ในสถานประกอบการ
14	2	1	1	2	6	1	5	4	3	5	2
15	3	2	1	3	5	1	7	2	1	2	4
16	2	1	1	1	5	2	7	2	2	7	4
17	2	1	1	2	3	4	6	1	1	6	1
18	2	3	1	1	2	2	6	6	4	6	2
19	2	1	1	1	2	2	7	2	1	5	4
20	2	1	1	1	3	4	6	3	2	2	2
21	2	1	1	1	7	4	6	2	3	8	2
22	2	3	1	2	4	2	6	2	3	5	2
23	2	1	1	1	2	2	5	4	6	3	3
24	2	1	1	1	2	2	6	4	3	3	4

พระจอมเกล้าพระนครเหนือ มหาวิทยาลัยเทคโนโลยี

4.2 ผลการพัฒนาและทดสอบประสิทธิภาพของแบบจำลอง

ในการพัฒนาแบบจำลองของทั้ง 3 รูปแบบ ได้แก่ โครงข่ายประสาทเทียมกราฟคอนโวลูชัน (Graph Convolutional Network: GCN) เอ็กตรีมกราเดี้ยนบูทติ้ง (Extreme Gradient Boosting : XGBoost) Hybrid of Graph Convolutional Network and Extreme Gradient Boosting ในการปรับแต่งแบบจำลอง (Hyperparameter Tuning) ผู้วิจัยได้ดำเนินการกำหนดปรับแต่งค่าพารามิเตอร์ต่างๆในแต่ละรูปแบบเพื่อค้นหาชุดพารามิเตอร์ที่ให้ประสิทธิภาพสูงสุด (Best Parameters) สำหรับแต่ละอัลกอริทึม โดยใช้เทคนิคการตรวจสอบความถูกต้องข้ามกลุ่มแบบ K-Fold (K-Fold Cross-validation) บนชุดข้อมูลฝึกฝนแบบจำลองจำนวน 20,000 รายการ ทั้งนี้ การเปรียบเทียบประสิทธิภาพของแบบจำลองแต่ละรูปแบบ จะถูกพิจารณาจากตัวชี้วัดผลทางสถิติ 4 ประการ ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่าเอฟวัน (F1-Score: F) โดยมีรายละเอียดดังนี้

4.2.1 ผลการพัฒนาแบบจำลอง โครงข่ายประสาทเทียมกราฟคอนโวลูชัน (Graph Convolutional Network: GCN)

ตารางที่ 4.3 ค่าพารามิเตอร์สำหรับแบบจำลอง GCN

Parameter	Value
ค่า K (Number of Nearest Neighbors)	3, 5, 7, 9, 11, 13
Dropout Rate	0.1, 0.2, 0.5
Epochs	60, 100, 150
Hidden Dimension	32, 64, 128

จากตารางที่ 4.3 ผู้วิจัยได้ทำการปรับค่าพารามิเตอร์จำนวน 162 รูปแบบโดยกำหนดค่า K เท่ากับ [3, 5, 7, 9, 11, 13] ค่า dropout เท่ากับ [0.1, 0.2, 0.5] ค่า Epochs เท่ากับ [60, 100, 150] และค่า Hidden dim เท่ากับ [32, 64, 128]

ตารางที่ 4-4 ผลลัพธ์ของแบบจำลอง GCN

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
3	0.1	128	150	78.105	0.800	0.749	0.781
3	0.2	128	150	77.960	0.801	0.745	0.779
3	0.5	128	150	77.570	0.799	0.738	0.775

ตารางที่ 4-4 (ต่อ)

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
3	0.1	128	100	77.260	0.790	0.742	0.772
3	0.1	64	150	77.240	0.793	0.737	0.772
3	0.2	128	100	77.140	0.790	0.739	0.771
3	0.2	64	150	77.015	0.794	0.729	0.770
3	0.5	128	100	76.965	0.791	0.733	0.769
3	0.5	64	150	76.915	0.796	0.723	0.769
3	0.1	32	150	76.790	0.790	0.729	0.768
5	0.2	128	150	76.680	0.779	0.746	0.767
5	0.1	128	150	76.650	0.777	0.749	0.766
3	0.2	64	100	76.650	0.789	0.728	0.766
3	0.2	32	150	76.580	0.791	0.722	0.765
3	0.1	128	60	76.430	0.784	0.730	0.764
3	0.1	64	100	76.395	0.787	0.724	0.764
3	0.5	64	100	76.335	0.790	0.717	0.763
5	0.5	128	150	76.275	0.778	0.735	0.763
5	0.1	128	100	76.265	0.780	0.733	0.762
3	0.2	128	60	76.255	0.785	0.724	0.762
3	0.1	32	100	76.220	0.786	0.720	0.762
3	0.2	32	100	76.205	0.787	0.718	0.762
3	0.5	128	60	76.175	0.785	0.722	0.761
3	0.5	32	150	76.005	0.789	0.710	0.759
3	0.5	32	100	75.990	0.789	0.710	0.759
5	0.2	128	100	75.905	0.775	0.731	0.759
3	0.5	64	60	75.840	0.783	0.715	0.758
5	0.1	64	150	75.785	0.773	0.730	0.758
5	0.2	64	150	75.780	0.774	0.728	0.758
3	0.1	64	60	75.680	0.775	0.724	0.757
5	0.5	128	100	75.660	0.771	0.729	0.756
5	0.5	64	150	75.625	0.772	0.728	0.756
7	0.2	128	150	75.600	0.767	0.735	0.756
5	0.1	64	100	75.595	0.775	0.722	0.756
3	0.2	64	60	75.575	0.775	0.721	0.755

ตารางที่ 4-4 (ต่อ)

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
7	0.1	128	150	75.435	0.765	0.734	0.754
7	0.5	128	150	75.415	0.767	0.731	0.754
5	0.1	32	150	75.355	0.771	0.721	0.753
9	0.1	128	150	75.235	0.760	0.738	0.752
3	0.5	32	60	75.250	0.771	0.719	0.752
5	0.2	64	100	75.180	0.769	0.720	0.752
5	0.1	128	60	75.180	0.770	0.718	0.752
9	0.2	128	150	75.110	0.760	0.734	0.751
7	0.1	128	100	75.060	0.758	0.735	0.751
5	0.5	64	100	75.060	0.770	0.715	0.750
5	0.2	128	60	75.050	0.769	0.716	0.750
5	0.2	32	150	75.035	0.769	0.715	0.750
5	0.5	32	150	75.000	0.772	0.709	0.750
7	0.1	64	150	74.960	0.763	0.725	0.749
7	0.2	128	100	74.960	0.764	0.722	0.749
5	0.5	128	60	74.945	0.768	0.715	0.749
3	0.2	32	60	74.935	0.765	0.720	0.749
9	0.5	128	150	74.820	0.756	0.734	0.748
7	0.2	64	150	74.815	0.762	0.722	0.748
7	0.5	128	100	74.815	0.762	0.722	0.748
3	0.1	32	60	74.715	0.760	0.724	0.747
5	0.2	64	60	74.670	0.762	0.718	0.746
11	0.1	128	150	74.645	0.754	0.733	0.746
5	0.2	32	100	74.675	0.767	0.709	0.746
5	0.1	32	100	74.625	0.765	0.711	0.746
7	0.5	64	150	74.590	0.762	0.715	0.746
5	0.5	32	100	74.560	0.765	0.709	0.745
11	0.2	128	150	74.470	0.752	0.730	0.745
5	0.5	64	60	74.470	0.761	0.714	0.744
5	0.1	64	60	74.460	0.760	0.715	0.744
9	0.2	64	150	74.445	0.754	0.726	0.744
7	0.5	64	100	74.415	0.758	0.717	0.744

ตารางที่ 4-4 (ต่อ)

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
7	0.2	32	150	74.365	0.756	0.721	0.743
7	0.1	32	150	74.355	0.759	0.714	0.743
9	0.1	64	150	74.340	0.755	0.720	0.743
11	0.5	128	150	74.310	0.751	0.728	0.743
9	0.1	128	100	74.305	0.755	0.720	0.743
9	0.5	128	100	74.270	0.751	0.726	0.743
9	0.2	128	100	74.255	0.753	0.722	0.742
13	0.2	128	150	74.235	0.750	0.727	0.742
7	0.2	64	100	74.225	0.756	0.715	0.742
7	0.1	128	60	74.220	0.755	0.718	0.742
11	0.2	64	150	74.180	0.752	0.722	0.742
7	0.2	128	60	74.150	0.756	0.714	0.741
7	0.1	64	100	74.145	0.755	0.715	0.741
5	0.2	32	60	74.120	0.756	0.712	0.741
13	0.1	128	150	74.090	0.745	0.732	0.741
7	0.1	32	100	74.105	0.757	0.710	0.741
11	0.2	128	100	74.070	0.749	0.724	0.741
11	0.1	64	150	74.030	0.745	0.731	0.740
11	0.1	128	100	73.985	0.743	0.733	0.740
7	0.5	128	60	74.000	0.754	0.713	0.740
7	0.5	32	150	74.000	0.758	0.706	0.740
5	0.5	32	60	73.975	0.756	0.708	0.739
7	0.5	32	100	73.890	0.755	0.708	0.739
9	0.5	64	150	73.855	0.755	0.706	0.738
11	0.5	128	100	73.775	0.747	0.718	0.738
5	0.1	32	60	73.770	0.750	0.714	0.738
9	0.1	32	150	73.755	0.748	0.716	0.737
11	0.5	64	150	73.740	0.748	0.717	0.737
7	0.2	64	60	73.715	0.752	0.708	0.737
9	0.2	64	100	73.680	0.750	0.712	0.737
13	0.5	128	150	73.670	0.747	0.716	0.737
9	0.2	32	150	73.650	0.749	0.711	0.736

ตารางที่ 4-4 (ต่อ)

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
9	0.1	128	60	73.585	0.748	0.711	0.736
13	0.2	128	100	73.570	0.744	0.718	0.736
13	0.2	64	150	73.565	0.744	0.720	0.736
9	0.2	128	60	73.565	0.749	0.709	0.735
7	0.2	32	100	73.555	0.749	0.708	0.735
9	0.1	64	100	73.530	0.747	0.711	0.735
13	0.1	128	100	73.520	0.745	0.715	0.735
9	0.5	64	100	73.515	0.746	0.714	0.735
11	0.1	64	100	73.510	0.743	0.718	0.735
7	0.5	64	60	73.525	0.751	0.705	0.735
13	0.1	64	150	73.490	0.742	0.721	0.735
11	0.2	128	60	73.475	0.746	0.711	0.735
9	0.5	32	150	73.435	0.749	0.705	0.734
7	0.1	64	60	73.420	0.746	0.710	0.734
9	0.5	128	60	73.415	0.749	0.704	0.734
13	0.5	128	100	73.400	0.744	0.713	0.734
7	0.2	32	60	73.355	0.747	0.707	0.733
13	0.1	64	100	73.335	0.742	0.716	0.733
13	0.1	32	150	73.295	0.745	0.708	0.733
11	0.2	64	100	73.285	0.743	0.712	0.733
11	0.5	64	100	73.290	0.745	0.709	0.733
7	0.1	32	60	73.260	0.747	0.704	0.732
11	0.2	32	150	73.240	0.744	0.708	0.732
9	0.2	32	100	73.230	0.745	0.707	0.732
13	0.5	64	150	73.220	0.743	0.709	0.732
9	0.5	32	100	73.215	0.746	0.703	0.732
11	0.1	32	150	73.190	0.742	0.711	0.732
11	0.1	128	60	73.170	0.742	0.711	0.732
9	0.1	32	100	73.125	0.744	0.706	0.731
11	0.5	128	60	73.105	0.744	0.704	0.731
11	0.2	32	100	73.085	0.743	0.706	0.731
11	0.5	32	150	73.045	0.741	0.708	0.730

ตารางที่ 4-4 (ต่อ)

Parameter				Performance			
K	Dropout	Hidden	Epochs	Acc(%)	Pre	Rec	F1
13	0.2	64	100	73.045	0.742	0.706	0.730
9	0.1	64	60	72.980	0.742	0.705	0.730
9	0.5	64	60	72.910	0.739	0.710	0.729
13	0.5	64	100	72.900	0.740	0.706	0.729
13	0.2	32	150	72.900	0.740	0.707	0.729
13	0.5	32	150	72.870	0.737	0.710	0.729
7	0.5	32	60	72.875	0.738	0.710	0.729
11	0.5	32	100	72.860	0.740	0.705	0.728
9	0.2	64	60	72.850	0.739	0.707	0.728
11	0.5	64	60	72.815	0.740	0.704	0.728
13	0.1	128	60	72.785	0.739	0.706	0.728
11	0.2	64	60	72.710	0.738	0.706	0.727
13	0.2	128	60	72.705	0.739	0.703	0.727
13	0.5	32	100	72.695	0.736	0.707	0.727
11	0.1	32	100	72.685	0.739	0.701	0.727
11	0.1	64	60	72.685	0.740	0.701	0.727
13	0.5	128	60	72.660	0.737	0.705	0.726
13	0.1	32	100	72.615	0.739	0.700	0.726
9	0.1	32	60	72.610	0.738	0.703	0.726
13	0.2	32	100	72.505	0.735	0.704	0.725
11	0.2	32	60	72.485	0.735	0.705	0.725
13	0.5	64	60	72.410	0.737	0.698	0.724
9	0.5	32	60	72.295	0.731	0.706	0.723
13	0.2	64	60	72.305	0.735	0.698	0.723
13	0.1	64	60	72.155	0.731	0.702	0.721
11	0.5	32	60	72.075	0.731	0.699	0.721
13	0.2	32	60	71.955	0.731	0.695	0.719
9	0.2	32	60	71.875	0.730	0.696	0.719
11	0.1	32	60	71.770	0.727	0.698	0.718
13	0.1	32	60	71.620	0.726	0.696	0.716
13	0.5	32	60	71.190	0.720	0.694	0.712

จากตารางที่ 4.4 ที่แสดงผลลัพธ์ทั้ง 162 รูปแบบ แสดงให้เห็นว่าค่าพารามิเตอร์ K เท่ากับ 3 ค่า Dropout 0.1 ค่า Hidden dim 128 ค่า Epochs 150 ให้ประสิทธิภาพดีที่สุดด้วยค่าความถูกต้องร้อยละ 78.10

4.2.1 ผลการพัฒนาแบบจำลอง เอ็กตรีมกราเดี้ยนบูทติ้ง (Extreme Gradient Boosting : XGBoost)

ตารางที่ 4.5 ค่าพารามิเตอร์สำหรับแบบจำลอง XGBoost

Parameter	Value
Number of Boosting Trees	200, 400, 600, 800
Learning Rate	0.03, 0.05, 0.1
Maximum Tree Depth	4, 6, 8, 10
Minimum Child Weight	1, 2, 4
Subsample Ratio of Training Instances	0.1, 0.2, 0.5, 1
Column Subsample Ratio per Tree	0.01, 0.02, 0.05, 1
Alpha	0, 0.1
Lambda	0, 1
Gamma	0, 0.1

จากตารางที่ 2 ผู้วิจัยได้ทำการปรับค่าพารามิเตอร์จำนวน 18,432 รูปแบบโดยกำหนดค่า n_estimators (trees) เท่ากับ [200, 400, 600, 800] ค่า learning_rate [0.03, 0.05, 0.1] ค่า max_depth [4, 6, 8, 10] ค่า min_child_weight [1, 2, 4] ค่า subsample [0.1, 0.2, 0.5, 1] ค่า colsample_bytree [0.01, 0.02, 0.05, 1] ค่า reg_alpha [0, 0.1] ค่า reg_lambda [0, 1] และ ค่า gamma [0, 0.1]

ตารางที่ 4-6 ผลลัพธ์ของแบบจำลอง XGBoost

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
1	0	0.1	10	1	800	0	0	1	92.440	0.924	0.1	10
1	0	0.1	10	1	600	0	0	1	92.430	0.924	0.1	10
1	0	0.1	8	1	800	0	0	1	91.705	0.917	0.1	8
1	0	0.1	8	1	600	0	0	1	91.100	0.911	0.1	8
0.5	0.1	0.05	6	1	800	0	0	0.5	89.330	0.893	0.05	6
0.5	0	0.05	6	1	800	0.1	0	0.5	89.290	0.893	0.05	6
0.5	0	0.05	6	1	800	0	0	0.5	89.285	0.893	0.05	6
0.5	0.1	0.05	6	1	800	0.1	0	0.5	89.245	0.892	0.05	6
0.5	0.1	0.05	6	1	800	0.1	1	0.5	88.645	0.886	0.05	6
0.5	0	0.05	6	1	800	0.1	1	0.5	88.590	0.886	0.05	6
0.5	0.1	0.05	6	1	800	0	1	0.5	88.535	0.885	0.05	6
0.5	0	0.05	6	2	800	0	0	0.5	88.530	0.885	0.05	6
0.5	0.1	0.05	6	2	800	0	0	0.5	88.530	0.885	0.05	6
0.5	0	0.05	6	1	800	0	1	0.5	88.520	0.885	0.05	6
0.5	0	0.05	6	2	800	0.1	0	0.5	92.440	0.885	0.05	6

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	2	800	0.1	0	0.5	88.400	0.885	0.884	0.884
0.5	0	0.05	6	1	600	0	0	0.5	88.180	0.882	0.882	0.882
0.5	0.1	0.05	6	2	800	0	1	0.5	88.170	0.882	0.882	0.882
0.5	0	0.05	6	1	600	0.1	0	0.5	88.120	0.882	0.881	0.881
0.5	0	0.05	6	2	800	0	1	0.5	88.070	0.881	0.881	0.881
0.5	0.1	0.05	6	1	600	0.1	0	0.5	88.055	0.881	0.881	0.881
0.5	0.1	0.05	6	2	800	0.1	1	0.5	88.050	0.881	0.881	0.880
0.5	0	0.05	6	2	800	0.1	1	0.5	88.010	0.881	0.880	0.880
0.5	0.1	0.05	6	1	600	0	0	0.5	88.005	0.881	0.880	0.880
0.5	0.1	0.05	6	1	800	0	0	0.2	87.715	0.878	0.877	0.877
0.5	0	0.05	6	1	800	0	0	0.2	87.590	0.876	0.876	0.876
0.5	0	0.05	6	1	800	0.1	0	0.2	87.535	0.876	0.875	0.875
0.5	0.1	0.05	6	1	800	0.1	0	0.2	87.410	0.875	0.874	0.874
0.5	0.1	0.05	6	2	600	0	0	0.5	87.315	0.873	0.873	0.873
0.5	0	0.05	6	1	600	0	1	0.5	87.300	0.873	0.873	0.873
0.5	0	0.05	6	2	600	0	0	0.5	87.290	0.873	0.873	0.873

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	6	1	600	0.1	1	0.5	87.265	0.873	0.873	0.873
0.5	0.1	0.03	6	1	800	0	0	0.5	87.255	0.873	0.873	0.873
0.5	0.1	0.05	6	2	600	0.1	0	0.5	87.250	0.873	0.873	0.872
0.5	0.1	0.05	6	1	600	0.1	1	0.5	87.240	0.873	0.872	0.872
0.5	0.1	0.03	6	1	800	0.1	0	0.5	87.215	0.872	0.872	0.872
0.5	0	0.05	6	2	600	0.1	0	0.5	87.205	0.872	0.872	0.872
0.5	0	0.03	6	1	800	0	0	0.5	87.155	0.872	0.872	0.872
0.5	0.1	0.05	6	1	600	0	1	0.5	87.145	0.872	0.871	0.871
0.5	0	0.05	6	5	800	0	0	0.5	86.940	0.870	0.869	0.869
0.5	0	0.03	6	1	800	0.1	0	0.5	86.930	0.870	0.869	0.869
0.5	0.1	0.05	6	5	800	0	0	0.5	86.900	0.869	0.869	0.869
0.5	0.1	0.05	6	1	800	0	1	0.2	86.895	0.869	0.869	0.869
0.5	0	0.05	6	2	600	0	1	0.5	86.885	0.869	0.869	0.869
0.5	0	0.05	6	5	800	0.1	0	0.5	86.855	0.869	0.869	0.869
0.5	0.1	0.05	6	5	800	0.1	0	0.5	86.835	0.869	0.868	0.868
0.5	0	0.05	6	1	800	0	1	0.2	86.820	0.869	0.868	0.868

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	5	800	0	1	0.5	86.730	0.868	0.867	0.867
0.5	0	0.05	6	1	800	0.1	1	0.2	86.720	0.868	0.867	0.867
0.5	0.1	0.05	6	1	800	0.1	1	0.2	86.715	0.867	0.867	0.867
0.5	0.1	0.05	6	2	600	0	1	0.5	86.710	0.867	0.867	0.867
0.5	0	0.05	6	2	600	0.1	1	0.5	86.695	0.867	0.867	0.867
0.5	0.1	0.05	6	5	800	0.1	1	0.5	86.690	0.867	0.867	0.867
0.5	0	0.05	6	5	800	0.1	1	0.5	86.660	0.867	0.867	0.867
0.5	0	0.05	6	5	800	0	1	0.5	86.615	0.866	0.866	0.866
0.5	0.1	0.05	6	2	600	0.1	1	0.5	86.560	0.866	0.866	0.866
0.5	0	0.05	6	2	800	0	0	0.2	86.525	0.866	0.865	0.865
0.5	0.1	0.03	6	1	800	0	1	0.5	86.480	0.865	0.865	0.865
0.5	0	0.03	6	2	800	0	0	0.5	86.450	0.865	0.865	0.864
0.5	0	0.03	6	2	800	0.1	0	0.5	86.390	0.864	0.864	0.864
0.5	0.1	0.05	6	1	600	0	0	0.2	86.380	0.864	0.864	0.864
0.5	0.1	0.03	6	2	800	0	0	0.5	86.375	0.864	0.864	0.864
0.5	0	0.05	5	1	800	0.1	0	0.5	86.325	0.864	0.863	0.863

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	2	800	0	0	0.2	86.315	0.863	0.863	0.863
0.5	0	0.03	6	1	800	0	1	0.5	86.310	0.863	0.863	0.863
0.5	0	0.05	6	1	600	0	0	0.2	86.285	0.863	0.863	0.863
0.5	0.1	0.05	5	1	800	0	0	0.5	86.285	0.863	0.863	0.863
0.5	0.1	0.05	6	2	800	0.1	0	0.2	86.260	0.863	0.863	0.863
0.5	0.1	0.05	6	1	600	0.1	0	0.2	86.255	0.863	0.863	0.863
0.5	0.1	0.05	5	1	800	0.1	0	0.5	86.220	0.862	0.862	0.862
0.5	0.1	0.03	6	2	800	0.1	0	0.5	86.215	0.862	0.862	0.862
0.5	0	0.05	6	1	600	0.1	0	0.2	86.205	0.862	0.862	0.862
0.5	0	0.05	6	2	800	0.1	0	0.2	86.205	0.862	0.862	0.862
0.5	0	0.03	6	1	800	0.1	1	0.5	86.195	0.862	0.862	0.862
0.5	0	0.05	5	1	800	0	0	0.5	86.160	0.862	0.862	0.862
0.5	0.1	0.03	6	1	800	0.1	1	0.5	86.105	0.861	0.861	0.861
0.5	0	0.05	6	1	400	0	0	0.5	86.080	0.861	0.861	0.861
0.5	0.1	0.05	6	1	400	0	0	0.5	86.015	0.860	0.860	0.860
0.5	0.1	0.05	6	1	400	0.1	0	0.5	85.980	0.860	0.860	0.860

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	6	2	800	0.1	1	0.2	85.955	0.860	0.860	0.860
0.5	0	0.05	6	1	400	0.1	0	0.5	85.935	0.860	0.859	0.859
0.5	0	0.05	6	2	800	0	1	0.2	85.840	0.859	0.858	0.858
0.5	0	0.05	5	2	800	0	0	0.5	85.790	0.858	0.858	0.858
0.5	0.1	0.05	6	2	800	0	1	0.2	85.785	0.858	0.858	0.858
0.5	0	0.03	6	1	800	0	0	0.2	85.775	0.858	0.858	0.858
0.5	0.1	0.03	6	1	600	0	0	0.5	85.775	0.858	0.858	0.858
0.5	0.1	0.03	6	1	600	0.1	0	0.5	85.775	0.858	0.858	0.858
0.5	0	0.03	6	1	600	0	0	0.5	85.750	0.858	0.858	0.857
0.5	0.1	0.03	6	1	800	0.1	0	0.2	85.745	0.858	0.857	0.857
0.5	0	0.03	6	2	800	0	1	0.5	85.725	0.857	0.857	0.857
0.5	0	0.03	6	2	800	0.1	1	0.5	85.725	0.857	0.857	0.857
0.5	0	0.05	6	1	600	0	1	0.2	85.725	0.857	0.857	0.857
0.5	0	0.05	6	1	800	0.1	0	0.1	85.725	0.858	0.857	0.857
0.5	0.1	0.05	5	2	800	0.1	0	0.5	85.725	0.857	0.857	0.857
0.5	0.1	0.05	6	2	800	0.1	1	0.2	85.725	0.857	0.857	0.857

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	6	5	600	0	0	0.5	85.705	0.857	0.857	0.857
0.5	0.1	0.05	6	1	600	0	1	0.2	85.700	0.857	0.857	0.857
0.5	0.1	0.05	6	5	600	0	0	0.5	85.700	0.857	0.857	0.857
0.5	0	0.05	6	5	600	0.1	0	0.5	85.695	0.857	0.857	0.857
0.5	0.1	0.03	6	2	800	0.1	1	0.5	85.685	0.857	0.857	0.857
0.5	0.1	0.03	6	1	800	0	0	0.2	85.675	0.857	0.857	0.857
0.5	0	0.03	6	1	800	0.1	0	0.2	85.655	0.857	0.857	0.857
0.5	0	0.05	5	2	800	0.1	0	0.5	85.625	0.856	0.856	0.856
0.5	0.1	0.03	6	2	800	0	1	0.5	85.620	0.856	0.856	0.856
0.5	0.1	0.05	6	5	600	0.1	0	0.5	85.605	0.856	0.856	0.856
0.5	0.1	0.05	6	1	800	0	0	0.1	85.575	0.856	0.856	0.856
0.5	0.1	0.05	5	1	800	0	1	0.5	85.565	0.856	0.856	0.856
0.5	0	0.03	6	1	600	0.1	0	0.5	85.560	0.856	0.856	0.856
0.5	0	0.05	5	1	800	0.1	1	0.5	85.555	0.856	0.856	0.856
0.5	0.1	0.05	5	2	800	0	0	0.5	85.525	0.855	0.855	0.855
0.5	0	0.05	5	1	800	0	1	0.5	85.495	0.855	0.855	0.855

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	2	600	0	0	0.2	85.455	0.855	0.855	0.855
0.5	0	0.05	6	1	800	0	0	0.1	85.425	0.855	0.854	0.854
0.5	0.1	0.05	5	1	800	0.1	1	0.5	85.425	0.854	0.854	0.854
0.5	0.1	0.05	6	2	400	0	0	0.5	85.425	0.854	0.854	0.854
0.5	0.1	0.05	6	1	600	0.1	1	0.2	85.405	0.854	0.854	0.854
0.5	0.1	0.05	6	5	600	0	1	0.5	85.380	0.854	0.854	0.854
0.5	0	0.05	6	2	400	0.1	0	0.5	85.360	0.854	0.854	0.854
0.5	0	0.05	6	1	600	0.1	1	0.2	85.335	0.854	0.853	0.853
0.5	0.1	0.05	6	5	600	0.1	1	0.5	85.335	0.853	0.853	0.853
0.5	0	0.05	6	5	600	0	1	0.5	85.325	0.853	0.853	0.853
0.5	0	0.05	6	5	600	0.1	1	0.5	85.305	0.853	0.853	0.853
0.5	0	0.05	6	2	600	0.1	0	0.2	85.300	0.853	0.853	0.853
0.5	0.1	0.05	6	2	600	0.1	0	0.2	85.290	0.853	0.853	0.853
0.5	0	0.05	6	1	400	0	1	0.5	85.280	0.853	0.853	0.853
0.5	0	0.05	6	1	400	0.1	1	0.5	85.270	0.853	0.853	0.853
0.5	0	0.05	6	2	600	0	0	0.2	85.260	0.853	0.853	0.853

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	5	2	800	0	1	0.5	85.235	0.852	0.852	0.852
0.5	0.1	0.05	6	1	800	0.1	0	0.1	85.235	0.853	0.852	0.852
0.5	0.1	0.05	6	2	400	0.1	0	0.5	85.235	0.852	0.852	0.852
0.5	0.1	0.05	6	1	400	0	1	0.5	85.195	0.852	0.852	0.852
0.5	0	0.05	6	2	400	0	0	0.5	85.160	0.852	0.852	0.852
0.5	0.1	0.05	6	1	400	0.1	1	0.5	85.145	0.852	0.851	0.851
0.5	0	0.05	5	2	800	0.1	1	0.5	85.115	0.851	0.851	0.851
0.5	0.1	0.05	5	2	800	0.1	1	0.5	85.075	0.851	0.851	0.851
0.5	0	0.05	5	2	800	0	1	0.5	85.030	0.850	0.850	0.850
0.5	0.1	0.03	6	2	600	0	0	0.5	84.980	0.850	0.850	0.850
0.5	0	0.03	6	2	600	0.1	0	0.5	84.965	0.850	0.850	0.850
0.5	0	0.03	6	2	600	0	0	0.5	84.960	0.850	0.850	0.850
0.5	0	0.05	5	1	600	0.1	0	0.5	84.915	0.849	0.849	0.849
0.5	0.1	0.03	6	1	600	0	1	0.5	84.915	0.849	0.849	0.849
0.5	0.1	0.05	5	1	600	0	0	0.5	84.910	0.849	0.849	0.849
0.5	0.1	0.05	5	1	600	0.1	0	0.5	84.905	0.849	0.849	0.849

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	6	2	400	0.1	1	0.5	84.900	0.849	0.849	0.849
0.5	0	0.03	6	5	800	0.1	0	0.5	84.895	0.849	0.849	0.849
0.5	0	0.03	6	1	600	0	1	0.5	84.885	0.849	0.849	0.849
0.5	0.1	0.03	6	5	800	0.1	0	0.5	84.880	0.849	0.849	0.849
0.5	0.1	0.03	6	2	600	0.1	0	0.5	84.850	0.849	0.849	0.848
0.5	0	0.05	6	2	600	0.1	1	0.2	84.845	0.849	0.848	0.848
0.5	0	0.03	6	1	600	0.1	1	0.5	84.820	0.848	0.848	0.848
0.5	0.1	0.03	6	1	600	0.1	1	0.5	84.820	0.848	0.848	0.848
0.5	0.1	0.05	6	1	800	0	1	0.1	84.805	0.848	0.848	0.848
0.5	0	0.03	6	5	800	0	0	0.5	84.785	0.848	0.848	0.848
0.5	0	0.05	6	2	400	0	1	0.5	84.765	0.848	0.848	0.848
0.5	0.1	0.05	5	1	800	0.1	0	0.2	84.755	0.848	0.848	0.848
0.5	0	0.05	5	1	800	0.1	0	0.2	84.750	0.848	0.848	0.847
0.5	0	0.05	6	2	600	0	1	0.2	84.745	0.848	0.847	0.847
0.5	0	0.05	5	1	800	0	0	0.2	84.740	0.848	0.847	0.847
0.5	0.1	0.05	6	2	400	0	1	0.5	84.715	0.847	0.847	0.847

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	2	400	0.1	1	0.5	84.715	0.847	0.847	0.847
0.5	0	0.05	5	1	600	0	0	0.5	84.700	0.847	0.847	0.847
0.5	0.1	0.05	5	1	800	0	0	0.2	84.690	0.847	0.847	0.847
0.5	0.1	0.03	6	5	800	0	0	0.5	84.685	0.847	0.847	0.847
0.5	0.1	0.03	6	1	800	0.1	1	0.2	84.655	0.847	0.847	0.847
0.5	0	0.03	6	1	800	0.1	1	0.2	84.645	0.847	0.846	0.846
0.5	0	0.03	6	2	800	0	0	0.2	84.615	0.846	0.846	0.846
0.5	0	0.05	6	1	800	0	1	0.1	84.615	0.846	0.846	0.846
0.5	0.1	0.05	6	1	600	0	0	0.1	84.615	0.846	0.846	0.846
0.5	0	0.03	6	1	800	0	1	0.2	84.600	0.846	0.846	0.846
0.5	0.1	0.03	6	5	800	0	1	0.5	84.585	0.846	0.846	0.846
0.5	0	0.03	6	5	800	0	1	0.5	84.575	0.846	0.846	0.846
0.5	0	0.05	5	5	800	0	0	0.5	84.570	0.846	0.846	0.846
0.5	0	0.03	6	2	800	0.1	0	0.2	84.560	0.846	0.846	0.846
0.5	0	0.05	5	5	800	0.1	0	0.5	84.535	0.845	0.845	0.845
0.5	0.1	0.03	6	1	800	0	1	0.2	84.535	0.845	0.845	0.845

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	2	600	0	1	0.2	84.530	0.845	0.845	0.845
0.5	0	0.05	6	1	400	0	0	0.2	84.525	0.845	0.845	0.845
0.5	0	0.05	6	1	800	0.1	1	0.1	84.520	0.845	0.845	0.845
0.5	0.1	0.03	6	2	800	0.1	0	0.2	84.500	0.845	0.845	0.845
0.5	0.1	0.05	5	2	600	0.1	0	0.5	84.485	0.845	0.845	0.845
0.5	0.1	0.05	5	5	800	0	0	0.5	84.485	0.845	0.845	0.845
0.5	0.1	0.03	6	2	800	0	0	0.2	84.480	0.845	0.845	0.845
0.5	0.1	0.05	6	2	600	0.1	1	0.2	84.470	0.845	0.845	0.845
0.5	0.1	0.03	6	5	800	0.1	1	0.5	84.460	0.845	0.845	0.845
0.5	0	0.03	6	5	800	0.1	1	0.5	84.440	0.844	0.844	0.844
0.5	0	0.03	6	2	600	0	1	0.5	84.430	0.844	0.844	0.844
0.5	0	0.05	5	2	600	0	0	0.5	84.430	0.844	0.844	0.844
0.5	0.1	0.05	6	1	400	0	0	0.2	84.430	0.844	0.844	0.844
0.5	0.1	0.05	5	5	800	0.1	0	0.5	84.425	0.844	0.844	0.844
0.5	0.1	0.03	6	2	600	0	1	0.5	84.400	0.844	0.844	0.844
0.5	0.1	0.05	6	1	800	0.1	1	0.1	84.395	0.844	0.844	0.844

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.03	6	1	600	0	0	0.2	84.380	0.844	0.844	0.844
0.5	0.1	0.03	6	1	600	0	0	0.2	84.365	0.844	0.844	0.844
0.5	0.1	0.05	6	1	400	0.1	0	0.2	84.360	0.844	0.844	0.844
0.5	0.1	0.03	6	1	600	0.1	0	0.2	84.355	0.844	0.844	0.844
0.5	0.1	0.05	5	2	600	0	0	0.5	84.355	0.844	0.844	0.844
0.5	0.1	0.05	5	1	800	0.1	1	0.2	84.340	0.844	0.843	0.843
0.5	0	0.03	6	1	600	0.1	0	0.2	84.320	0.843	0.843	0.843
0.5	0	0.05	6	1	600	0.1	0	0.1	84.315	0.843	0.843	0.843
0.5	0	0.05	6	1	400	0.1	0	0.2	84.305	0.843	0.843	0.843
0.5	0.1	0.05	6	1	600	0.1	0	0.1	84.300	0.843	0.843	0.843
0.5	0.1	0.05	5	2	800	0	0	0.2	84.295	0.843	0.843	0.843
0.5	0.1	0.03	6	2	600	0.1	1	0.5	84.290	0.843	0.843	0.843
0.5	0.1	0.05	5	1	600	0	1	0.5	84.265	0.843	0.843	0.843
0.5	0	0.05	5	2	600	0.1	0	0.5	84.255	0.843	0.843	0.843
0.5	0	0.05	6	1	600	0	0	0.1	84.240	0.843	0.842	0.842
0.5	0	0.05	5	1	800	0.1	1	0.2	84.235	0.842	0.842	0.842

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	5	2	800	0.1	0	0.2	84.230	0.842	0.842	0.842
0.5	0	0.03	6	2	600	0.1	1	0.5	84.215	0.842	0.842	0.842
0.5	0	0.05	5	1	800	0	1	0.2	84.190	0.842	0.842	0.842
0.5	0	0.05	5	5	800	0	1	0.5	84.190	0.842	0.842	0.842
0.5	0	0.05	5	1	600	0.1	1	0.5	84.180	0.842	0.842	0.842
0.5	0	0.05	5	1	600	0	1	0.5	84.145	0.842	0.841	0.841
0.5	0	0.05	6	5	800	0	0	0.2	84.135	0.841	0.841	0.841
0.5	0.1	0.05	5	5	800	0	1	0.5	84.120	0.841	0.841	0.841
0.5	0	0.05	5	5	800	0.1	1	0.5	84.085	0.841	0.841	0.841
0.5	0	0.05	6	2	800	0.1	0	0.1	84.075	0.841	0.841	0.841
0.5	0.1	0.05	5	1	600	0.1	1	0.5	84.075	0.841	0.841	0.841
0.5	0.1	0.05	5	5	800	0.1	1	0.5	84.065	0.841	0.841	0.841
0.5	0	0.05	6	5	800	0.1	0	0.2	84.060	0.841	0.841	0.841
0.5	0.1	0.05	6	5	800	0	0	0.2	84.055	0.841	0.841	0.841
0.5	0	0.05	6	2	800	0	0	0.1	84.050	0.841	0.841	0.840
0.5	0.1	0.05	6	2	800	0	0	0.1	84.050	0.841	0.841	0.840

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	5	800	0.1	0	0.2	84.050	0.841	0.841	0.840
0.5	0	0.05	5	2	800	0	0	0.2	84.040	0.840	0.840	0.840
0.5	0	0.05	5	2	800	0.1	0	0.2	84.035	0.840	0.840	0.840
0.5	0	0.05	6	5	800	0	1	0.2	84.030	0.840	0.840	0.840
0.5	0	0.05	5	2	600	0.1	1	0.5	84.015	0.840	0.840	0.840
0.5	0.1	0.05	6	5	400	0	0	0.5	83.980	0.840	0.840	0.840
0.5	0.1	0.05	6	5	400	0.1	0	0.5	83.980	0.840	0.840	0.840
0.5	0.1	0.03	5	1	800	0	0	0.5	83.945	0.840	0.839	0.839
0.5	0.1	0.03	6	2	800	0.1	1	0.2	83.935	0.839	0.839	0.839
0.5	0.1	0.05	5	1	800	0	1	0.2	83.935	0.839	0.839	0.839
0.5	0.1	0.03	6	1	800	0	0	0.1	83.930	0.839	0.839	0.839
0.5	0	0.03	6	2	800	0	1	0.2	83.925	0.839	0.839	0.839
0.5	0.1	0.05	6	2	800	0.1	0	0.1	83.920	0.839	0.839	0.839
0.5	0.1	0.05	6	5	800	0	1	0.2	83.900	0.839	0.839	0.839
0.5	0	0.03	6	1	800	0	0	0.1	83.890	0.839	0.839	0.839
0.5	0	0.05	6	5	400	0.1	0	0.5	83.885	0.839	0.839	0.839

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	6	5	800	0.1	1	0.2	83.865	0.839	0.839	0.839
0.5	0	0.03	5	1	800	0.1	0	0.5	83.850	0.839	0.839	0.838
0.5	0	0.03	5	1	800	0	0	0.5	83.845	0.838	0.838	0.838
0.5	0	0.03	6	1	800	0.1	0	0.1	83.845	0.839	0.838	0.838
0.5	0	0.03	6	2	800	0.1	1	0.2	83.845	0.839	0.838	0.838
0.5	0	0.05	6	5	400	0	0	0.5	83.845	0.839	0.838	0.838
0.5	0.1	0.05	5	2	600	0	1	0.5	83.845	0.839	0.838	0.838
0.5	0.1	0.03	6	2	800	0	1	0.2	83.810	0.838	0.838	0.838
0.5	0.1	0.05	5	2	600	0.1	1	0.5	83.790	0.838	0.838	0.838
0.5	0.1	0.03	5	1	800	0.1	0	0.5	83.780	0.838	0.838	0.838
0.5	0	0.05	6	5	800	0.1	1	0.2	83.775	0.838	0.838	0.838
0.5	0.1	0.03	6	1	800	0.1	0	0.1	83.775	0.838	0.838	0.838
0.5	0.1	0.05	6	5	400	0	1	0.5	83.775	0.838	0.838	0.838
0.5	0	0.03	6	5	600	0.1	0	0.5	83.770	0.838	0.838	0.838
0.5	0	0.05	5	2	600	0	1	0.5	83.735	0.837	0.837	0.837
0.5	0.1	0.05	5	2	800	0	1	0.2	83.725	0.837	0.837	0.837

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	6	1	400	0	0	0.5	83.710	0.837	0.837	0.837
0.5	0	0.03	6	1	400	0	0	0.5	83.680	0.837	0.837	0.837
0.5	0.1	0.05	6	2	400	0	0	0.2	83.680	0.837	0.837	0.837
0.5	0.1	0.03	6	5	600	0.1	0	0.5	83.675	0.837	0.837	0.837
0.5	0	0.05	6	1	400	0	1	0.2	83.670	0.837	0.837	0.837
0.5	0.1	0.05	5	1	600	0.1	0	0.2	83.670	0.837	0.837	0.837
0.5	0	0.05	6	5	400	0.1	1	0.5	83.650	0.837	0.837	0.836
0.5	0.1	0.05	6	2	800	0	1	0.1	83.630	0.836	0.836	0.836
0.5	0.1	0.05	6	5	400	0.1	1	0.5	83.625	0.836	0.836	0.836
0.5	0	0.05	5	2	800	0	1	0.2	83.620	0.836	0.836	0.836
0.5	0.1	0.03	6	1	400	0.1	0	0.5	83.620	0.836	0.836	0.836
0.5	0.1	0.05	5	1	600	0	0	0.2	83.580	0.836	0.836	0.836
0.5	0	0.05	5	1	600	0.1	0	0.2	83.575	0.836	0.836	0.836
0.5	0	0.05	6	2	400	0	0	0.2	83.555	0.836	0.836	0.836
0.5	0	0.03	6	1	400	0.1	0	0.5	83.550	0.836	0.836	0.835
0.5	0.1	0.03	6	5	600	0	0	0.5	83.545	0.836	0.835	0.835

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	6	1	600	0	1	0.1	83.540	0.835	0.835	0.835
0.5	0.1	0.05	5	2	800	0.1	1	0.2	83.525	0.835	0.835	0.835
0.5	0.1	0.05	6	1	400	0.1	1	0.2	83.525	0.835	0.835	0.835
0.5	0	0.05	6	2	400	0.1	0	0.2	83.520	0.835	0.835	0.835
0.5	0	0.05	6	5	400	0	1	0.5	83.515	0.835	0.835	0.835
0.5	0	0.03	5	2	800	0	0	0.5	83.510	0.835	0.835	0.835
0.5	0	0.05	5	1	600	0	0	0.2	83.510	0.835	0.835	0.835
0.5	0.1	0.05	6	1	400	0	1	0.2	83.510	0.835	0.835	0.835
0.5	0	0.03	6	5	600	0	0	0.5	83.480	0.835	0.835	0.835
0.5	0.1	0.05	6	1	600	0	1	0.1	83.480	0.835	0.835	0.835
0.5	0	0.05	6	2	800	0	1	0.1	83.475	0.835	0.835	0.835
0.5	0.1	0.05	5	1	800	0	0	0.1	83.470	0.835	0.835	0.835
0.5	0	0.03	5	2	800	0.1	0	0.5	83.465	0.835	0.835	0.835
0.5	0	0.05	5	2	800	0.1	1	0.2	83.465	0.835	0.835	0.835
0.5	0.1	0.03	5	2	800	0	0	0.5	83.465	0.835	0.835	0.835
0.5	0	0.05	6	1	400	0.1	1	0.2	83.440	0.834	0.834	0.834

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.03	6	1	600	0	1	0.2	83.415	0.834	0.834	0.834
0.5	0	0.03	6	5	600	0	1	0.5	83.400	0.834	0.834	0.834
0.5	0.1	0.03	5	2	800	0.1	0	0.5	83.400	0.834	0.834	0.834
0.5	0.1	0.03	6	1	600	0.1	1	0.2	83.400	0.834	0.834	0.834
0.5	0	0.05	5	1	800	0.1	0	0.1	83.385	0.834	0.834	0.834
0.5	0	0.05	6	2	800	0.1	1	0.1	83.380	0.834	0.834	0.834
0.5	0	0.03	6	1	600	0.1	1	0.2	83.375	0.834	0.834	0.834
0.5	0	0.03	6	2	600	0	0	0.2	83.370	0.834	0.834	0.834
0.5	0.1	0.05	6	2	400	0.1	0	0.2	83.365	0.834	0.834	0.834
0.5	0	0.05	5	5	600	0.1	0	0.5	83.355	0.834	0.834	0.834
0.5	0	0.05	5	1	400	0.1	0	0.5	83.350	0.834	0.834	0.833
0.5	0.1	0.03	6	5	600	0	1	0.5	83.345	0.834	0.833	0.833
0.5	0.1	0.05	5	5	600	0	0	0.5	83.335	0.833	0.833	0.833
0.5	0.1	0.03	6	2	600	0	0	0.2	83.330	0.833	0.833	0.833
0.5	0.1	0.03	6	5	600	0.1	1	0.5	83.330	0.833	0.833	0.833
0.5	0	0.05	5	5	600	0	0	0.5	83.315	0.833	0.833	0.833

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	6	2	600	0.1	0	0.2	83.310	0.833	0.833	0.833
0.5	0.1	0.05	6	2	800	0.1	1	0.1	83.310	0.833	0.833	0.833
0.5	0	0.05	5	1	800	0	0	0.1	83.300	0.833	0.833	0.833
0.5	0.1	0.03	6	1	600	0	1	0.2	83.300	0.833	0.833	0.833
0.5	0	0.03	6	5	600	0.1	1	0.5	83.295	0.833	0.833	0.833
0.5	0.1	0.03	5	1	800	0	1	0.5	83.285	0.833	0.833	0.833
0.5	0	0.03	6	2	600	0.1	0	0.2	83.280	0.833	0.833	0.833
0.5	0	0.03	6	1	800	0	1	0.1	83.270	0.833	0.833	0.833
0.5	0	0.03	5	1	800	0	1	0.5	83.240	0.832	0.832	0.832
0.5	0	0.05	6	1	600	0.1	1	0.1	83.235	0.832	0.832	0.832
0.5	0.1	0.05	6	5	600	0	0	0.2	83.230	0.832	0.832	0.832
0.5	0.1	0.05	5	1	400	0	0	0.5	83.225	0.832	0.832	0.832
0.5	0	0.03	6	2	400	0	0	0.5	83.215	0.832	0.832	0.832
0.5	0.1	0.05	5	1	800	0.1	0	0.1	83.205	0.832	0.832	0.832
0.5	0	0.05	5	5	600	0	1	0.5	83.195	0.832	0.832	0.832
0.5	0.1	0.03	6	1	800	0	1	0.1	83.190	0.832	0.832	0.832

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	5	5	600	0.1	0	0.5	83.190	0.832	0.832	0.832
0.5	0	0.05	5	1	400	0	0	0.5	83.165	0.832	0.832	0.832
0.5	0	0.05	6	5	600	0	0	0.2	83.160	0.832	0.832	0.832
0.5	0	0.03	5	1	800	0.1	1	0.5	83.155	0.832	0.832	0.832
0.5	0.1	0.05	6	1	600	0.1	1	0.1	83.145	0.832	0.831	0.831
0.5	0.1	0.05	5	1	400	0.1	0	0.5	83.135	0.831	0.831	0.831
0.5	0.1	0.03	6	2	400	0.1	0	0.5	83.125	0.831	0.831	0.831
0.5	0	0.03	6	2	400	0.1	0	0.5	83.115	0.831	0.831	0.831
0.5	0.1	0.03	5	1	800	0.1	1	0.5	83.110	0.831	0.831	0.831
0.5	0.1	0.03	5	2	800	0.1	1	0.5	83.080	0.831	0.831	0.831
0.5	0.1	0.05	6	2	400	0.1	1	0.2	83.080	0.831	0.831	0.831
0.5	0.1	0.03	6	2	400	0	0	0.5	83.070	0.831	0.831	0.831
0.5	0.1	0.05	5	2	600	0.1	0	0.2	83.065	0.831	0.831	0.831
0.5	0	0.03	5	2	800	0	1	0.5	83.055	0.831	0.831	0.831
0.5	0	0.05	6	2	400	0	1	0.2	83.045	0.831	0.830	0.830
0.5	0	0.05	6	5	600	0	1	0.2	83.010	0.830	0.830	0.830

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.03	5	2	800	0.1	1	0.5	83.005	0.830	0.830	0.830
0.5	0	0.03	6	1	800	0.1	1	0.1	83.000	0.830	0.830	0.830
0.5	0.1	0.03	6	1	400	0	1	0.5	83.000	0.830	0.830	0.830
0.5	0.1	0.05	5	5	600	0.1	1	0.5	83.000	0.830	0.830	0.830
0.5	0	0.05	5	5	600	0.1	1	0.5	82.995	0.830	0.830	0.830
0.5	0	0.05	5	1	600	0	1	0.2	82.990	0.830	0.830	0.830
0.5	0	0.05	5	2	600	0	0	0.2	82.990	0.830	0.830	0.830
0.5	0.1	0.05	5	5	600	0	1	0.5	82.990	0.830	0.830	0.830
0.5	0	0.05	6	5	600	0.1	0	0.2	82.985	0.830	0.830	0.830
0.5	0.1	0.03	5	2	800	0	1	0.5	82.980	0.830	0.830	0.830
0.5	0	0.03	6	1	400	0	1	0.5	82.970	0.830	0.830	0.830
0.5	0	0.05	6	2	600	0.1	0	0.1	82.970	0.830	0.830	0.830
0.5	0.1	0.05	6	2	600	0.1	0	0.1	82.965	0.830	0.830	0.830
0.5	0	0.05	6	2	400	0.1	1	0.2	82.960	0.830	0.830	0.830
0.5	0.1	0.03	6	1	400	0.1	1	0.5	82.960	0.830	0.830	0.830
0.5	0	0.03	6	1	400	0.1	1	0.5	82.950	0.830	0.830	0.829

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.05	5	1	600	0.1	1	0.2	82.925	0.829	0.829	0.829
0.5	0	0.05	5	2	400	0.1	0	0.5	82.920	0.829	0.829	0.829
0.5	0.1	0.03	6	1	800	0.1	1	0.1	82.920	0.829	0.829	0.829
0.5	0	0.03	5	1	800	0.1	0	0.2	82.915	0.829	0.829	0.829
0.5	0.1	0.05	5	1	600	0	1	0.2	82.915	0.829	0.829	0.829
0.5	0	0.03	6	2	600	0.1	1	0.2	82.910	0.829	0.829	0.829
0.5	0	0.05	6	1	400	0	0	0.1	82.895	0.829	0.829	0.829
0.5	0	0.03	5	1	800	0	0	0.2	82.890	0.829	0.829	0.829
0.5	0	0.05	5	1	800	0	1	0.1	82.890	0.829	0.829	0.829
0.5	0.1	0.05	6	2	400	0	1	0.2	82.890	0.829	0.829	0.829
0.5	0.1	0.05	5	1	600	0.1	1	0.2	82.880	0.829	0.829	0.829
0.5	0.1	0.05	6	5	600	0.1	0	0.2	82.880	0.829	0.829	0.829
0.5	0.1	0.03	6	1	600	0	0	0.1	82.870	0.829	0.829	0.829
0.5	0	0.05	6	1	200	0	0	0.5	82.855	0.829	0.829	0.829
0.5	0	0.05	6	2	600	0	0	0.1	82.840	0.829	0.828	0.828
0.5	0.1	0.05	5	2	600	0	0	0.2	82.835	0.828	0.828	0.828

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	6	2	800	0	0	0.1	82.830	0.828	0.828	0.828
0.5	0	0.05	5	2	400	0	0	0.5	82.825	0.828	0.828	0.828
0.5	0.1	0.03	6	2	600	0	1	0.2	82.825	0.828	0.828	0.828
0.5	0	0.05	6	5	600	0.1	1	0.2	82.820	0.828	0.828	0.828
0.5	0.1	0.05	5	2	400	0.1	0	0.5	82.815	0.828	0.828	0.828
0.5	0.1	0.05	6	5	600	0	1	0.2	82.815	0.828	0.828	0.828
0.5	0.1	0.05	6	1	200	0.1	0	0.5	82.810	0.828	0.828	0.828
0.5	0.1	0.05	5	1	800	0.1	1	0.1	82.795	0.828	0.828	0.828
0.5	0	0.05	5	2	600	0.1	0	0.2	82.790	0.828	0.828	0.828
0.5	0.1	0.05	5	2	400	0	0	0.5	82.790	0.828	0.828	0.828
0.5	0	0.03	6	2	400	0	1	0.5	82.780	0.828	0.828	0.828
0.5	0.1	0.05	6	1	400	0	0	0.1	82.780	0.828	0.828	0.828
0.5	0.1	0.03	6	2	400	0	1	0.5	82.770	0.828	0.828	0.828
0.5	0	0.05	5	2	800	0.1	0	0.1	82.750	0.828	0.828	0.827
0.5	0.1	0.05	6	2	600	0	0	0.1	82.745	0.828	0.827	0.827
0.5	0.1	0.03	6	2	800	0.1	0	0.1	82.735	0.827	0.827	0.827

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	6	2	600	0.1	1	0.2	82.725	0.827	0.827	0.827
0.5	0	0.03	6	2	600	0	1	0.2	82.700	0.827	0.827	0.827
0.5	0	0.05	5	1	800	0.1	1	0.1	82.700	0.827	0.827	0.827
0.5	0.1	0.05	6	5	600	0.1	1	0.2	82.700	0.827	0.827	0.827
0.5	0	0.03	6	2	400	0.1	1	0.5	82.695	0.827	0.827	0.827
0.5	0.1	0.05	5	1	800	0	1	0.1	82.690	0.827	0.827	0.827
0.5	0	0.03	5	5	800	0	0	0.5	82.685	0.827	0.827	0.827
0.5	0.1	0.03	5	1	800	0.1	0	0.2	82.685	0.827	0.827	0.827
0.5	0	0.05	6	1	400	0.1	0	0.1	82.680	0.827	0.827	0.827
0.5	0	0.03	5	5	800	0.1	0	0.5	82.670	0.827	0.827	0.827
0.5	0.1	0.03	6	1	400	0.1	0	0.2	82.670	0.827	0.827	0.827
0.5	0.1	0.03	6	1	600	0.1	0	0.1	82.660	0.827	0.827	0.827
0.5	0.1	0.05	4	1	800	0	0	0.5	82.660	0.827	0.827	0.827
0.5	0.1	0.05	6	1	200	0	0	0.5	82.660	0.827	0.827	0.827
0.5	0.1	0.05	5	2	800	0.1	0	0.1	82.655	0.827	0.827	0.827
0.5	0.1	0.03	5	1	800	0	0	0.2	82.650	0.827	0.827	0.826

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.03	6	2	800	0	0	0.1	82.640	0.826	0.826	0.826
0.5	0	0.05	4	1	800	0	0	0.5	82.635	0.826	0.826	0.826
0.5	0.1	0.05	6	1	400	0.1	0	0.1	82.635	0.826	0.826	0.826
0.5	0	0.03	6	1	600	0.1	0	0.1	82.625	0.826	0.826	0.826
0.5	0	0.05	4	1	800	0.1	0	0.5	82.625	0.826	0.826	0.826
0.5	0	0.05	6	1	200	0.1	0	0.5	82.615	0.826	0.826	0.826
0.5	0	0.03	6	1	400	0	0	0.2	82.610	0.826	0.826	0.826
0.5	0.1	0.05	5	1	400	0	1	0.5	82.610	0.826	0.826	0.826
0.5	0	0.03	6	1	600	0	0	0.1	82.605	0.826	0.826	0.826
0.5	0	0.03	6	1	400	0.1	0	0.2	82.600	0.826	0.826	0.826
0.5	0	0.03	6	2	800	0.1	0	0.1	82.595	0.826	0.826	0.826
0.5	0.1	0.05	4	1	800	0.1	0	0.5	82.595	0.826	0.826	0.826
0.5	0	0.03	5	1	600	0.1	0	0.5	82.590	0.826	0.826	0.826
0.5	0	0.03	6	5	800	0.1	0	0.2	82.575	0.826	0.826	0.826
0.5	0.1	0.03	6	1	400	0	0	0.2	82.560	0.826	0.826	0.826
0.5	0.1	0.03	6	2	400	0.1	1	0.5	82.560	0.826	0.826	0.826

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	5	5	800	0.1	0	0.5	82.555	0.826	0.826	0.826
0.5	0.1	0.03	5	5	800	0	0	0.5	82.545	0.826	0.825	0.825
0.5	0	0.05	5	2	600	0.1	1	0.2	82.540	0.825	0.825	0.825
0.5	0	0.03	6	5	800	0	0	0.2	82.530	0.825	0.825	0.825
0.5	0	0.03	5	1	600	0	0	0.5	82.525	0.825	0.825	0.825
0.5	0	0.05	5	1	400	0	1	0.5	82.525	0.825	0.825	0.825
0.5	0.1	0.03	6	5	800	0.1	0	0.2	82.515	0.825	0.825	0.825
0.5	0	0.05	5	1	400	0.1	1	0.5	82.510	0.825	0.825	0.825
0.5	0.1	0.05	5	5	800	0.1	0	0.2	82.510	0.825	0.825	0.825
0.5	0	0.05	5	5	800	0.1	0	0.2	82.500	0.825	0.825	0.825
0.5	0.1	0.05	5	2	600	0.1	1	0.2	82.500	0.825	0.825	0.825
0.5	0.1	0.05	6	2	600	0	1	0.1	82.485	0.825	0.825	0.825
0.5	0	0.05	5	5	800	0	0	0.2	82.455	0.825	0.825	0.825
0.5	0	0.05	5	2	800	0	0	0.1	82.450	0.825	0.825	0.824
0.5	0	0.05	6	2	600	0	1	0.1	82.450	0.825	0.825	0.824
0.5	0.1	0.05	6	2	600	0.1	1	0.1	82.450	0.825	0.825	0.824

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.05	5	2	800	0	0	0.1	82.440	0.824	0.824	0.824
0.5	0	0.05	4	2	800	0.1	0	0.5	82.435	0.825	0.824	0.824
0.5	0.1	0.03	5	1	600	0	0	0.5	82.435	0.824	0.824	0.824
0.5	0.1	0.05	4	2	800	0.1	0	0.5	82.435	0.825	0.824	0.824
0.5	0.1	0.03	5	1	600	0.1	0	0.5	82.425	0.824	0.824	0.824
0.5	0	0.03	6	5	800	0	1	0.2	82.420	0.824	0.824	0.824
0.5	0	0.03	5	1	800	0	1	0.2	82.405	0.824	0.824	0.824
0.5	0.1	0.03	6	5	800	0	0	0.2	82.405	0.824	0.824	0.824
0.5	0	0.05	4	2	800	0	0	0.5	82.400	0.824	0.824	0.824
0.5	0	0.05	6	2	600	0.1	1	0.1	82.385	0.824	0.824	0.824
0.5	0.1	0.05	4	2	800	0	0	0.5	82.385	0.824	0.824	0.824
0.5	0.1	0.05	5	1	600	0	0	0.1	82.370	0.824	0.824	0.824
0.5	0	0.03	6	2	800	0	1	0.1	82.365	0.824	0.824	0.824
0.5	0	0.05	5	1	600	0	0	0.1	82.365	0.824	0.824	0.824
0.5	0.1	0.03	5	1	800	0	1	0.2	82.350	0.824	0.824	0.823
0.5	0.1	0.05	5	1	400	0.1	1	0.5	82.350	0.824	0.824	0.823

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	5	5	800	0	1	0.5	82.335	0.824	0.823	0.823
0.5	0.1	0.05	5	1	600	0.1	0	0.1	82.330	0.823	0.823	0.823
0.5	0.1	0.05	6	2	200	0	0	0.5	82.330	0.823	0.823	0.823
0.5	0	0.05	5	2	400	0.1	1	0.5	82.325	0.823	0.823	0.823
0.5	0.1	0.03	5	5	800	0.1	1	0.5	82.320	0.823	0.823	0.823
0.5	0.1	0.03	6	5	800	0	1	0.2	82.320	0.823	0.823	0.823
0.5	0	0.03	5	2	600	0	0	0.5	82.305	0.823	0.823	0.823
0.5	0	0.03	5	2	800	0.1	0	0.2	82.300	0.823	0.823	0.823
0.5	0.1	0.03	6	2	800	0	1	0.1	82.300	0.823	0.823	0.823
0.5	0.1	0.05	5	2	600	0	1	0.2	82.295	0.823	0.823	0.823
0.5	0	0.05	5	2	800	0.1	1	0.1	82.285	0.823	0.823	0.823
0.5	0.1	0.03	5	2	800	0	0	0.2	82.285	0.823	0.823	0.823
0.5	0.1	0.03	5	2	800	0.1	0	0.2	82.280	0.823	0.823	0.823
0.5	0.1	0.03	5	1	800	0.1	1	0.2	82.270	0.823	0.823	0.823
0.5	0.1	0.05	5	5	800	0	0	0.2	82.270	0.823	0.823	0.823
0.5	0	0.03	5	1	800	0.1	1	0.2	82.265	0.823	0.823	0.823

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0	0.03	5	2	800	0	0	0.2	82.265	0.823	0.823	0.823
0.5	0	0.03	5	5	800	0	1	0.5	82.260	0.823	0.823	0.823
0.5	0	0.03	6	5	800	0.1	1	0.2	82.250	0.823	0.823	0.822
0.5	0	0.05	4	1	800	0	1	0.5	82.250	0.823	0.823	0.822
0.5	0.1	0.03	6	5	800	0.1	1	0.2	82.250	0.823	0.823	0.822
0.5	0.1	0.05	6	2	200	0.1	0	0.5	82.250	0.823	0.823	0.822
0.5	0	0.05	6	2	200	0.1	0	0.5	82.245	0.823	0.822	0.822
0.5	0	0.05	6	2	200	0	0	0.5	82.240	0.822	0.822	0.822
0.5	0.1	0.05	5	5	800	0	1	0.2	82.240	0.823	0.822	0.822
0.5	0	0.05	4	1	800	0.1	1	0.5	82.225	0.822	0.822	0.822
0.5	0.1	0.05	6	1	200	0	1	0.5	82.225	0.822	0.822	0.822
0.5	0	0.05	6	1	200	0.1	1	0.5	82.220	0.822	0.822	0.822
0.5	0.1	0.05	5	2	400	0.1	1	0.5	82.220	0.822	0.822	0.822
0.5	0	0.05	6	1	400	0	1	0.1	82.215	0.822	0.822	0.822
0.5	0.1	0.03	6	2	800	0.1	1	0.1	82.210	0.822	0.822	0.822
0.5	0	0.05	5	1	600	0.1	0	0.1	82.205	0.822	0.822	0.822

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	6	1	600	0	1	0.1	82.200	0.822	0.822	0.822
0.5	0.1	0.05	5	2	400	0	1	0.5	82.200	0.822	0.822	0.822
0.5	0.1	0.05	4	1	800	0	1	0.5	82.195	0.822	0.822	0.822
0.5	0.1	0.05	5	2	800	0	1	0.1	82.195	0.822	0.822	0.822
0.5	0	0.05	5	2	400	0	1	0.5	82.185	0.822	0.822	0.822
0.5	0.1	0.03	5	2	600	0	0	0.5	82.185	0.822	0.822	0.822
0.5	0.1	0.03	5	2	600	0.1	0	0.5	82.175	0.822	0.822	0.822
0.5	0	0.03	5	5	800	0.1	1	0.5	82.165	0.822	0.822	0.822
0.5	0	0.05	5	2	600	0	1	0.2	82.155	0.822	0.822	0.822
0.5	0	0.05	5	5	800	0	1	0.2	82.155	0.822	0.822	0.822
0.5	0.1	0.05	4	1	800	0.1	1	0.5	82.155	0.822	0.822	0.822
0.5	0	0.05	5	5	800	0.1	1	0.2	82.150	0.822	0.822	0.821
0.5	0.1	0.05	4	2	800	0.1	1	0.5	82.135	0.822	0.821	0.821
0.5	0.1	0.05	4	1	800	0	0	0.2	82.120	0.821	0.821	0.821
0.5	0	0.05	4	2	800	0	1	0.5	82.115	0.821	0.821	0.821
0.5	0.1	0.05	5	2	800	0.1	1	0.1	82.115	0.821	0.821	0.821

ตารางที่ 4-6 (ต่อ)

Parameter									Performance			
colsamp	gamma	learning	maxdep	minchil	N_esti	alpha	lambda	subsam	Acc(%)	Pre	Rec	F1
0.5	0.1	0.03	5	1	600	0	1	0.5	82.110	0.821	0.821	0.821
0.5	0.1	0.05	5	1	400	0.1	0	0.2	82.105	0.821	0.821	0.821
0.5	0	0.03	6	2	800	0.1	1	0.1	82.100	0.821	0.821	0.821
0.5	0	0.03	5	2	600	0.1	0	0.5	82.095	0.821	0.821	0.821
0.5	0	0.05	4	1	800	0	0	0.2	82.095	0.821	0.821	0.821

จากตารางที่ 4.6 ที่แสดงผลลัพธ์สูงสุด 500 รูปแบบจาก 18,432 รูปแบบแสดงให้เห็นว่าค่าพารามิเตอร์ n_estimators (trees) เท่ากับ 800 ค่า learning_rate 0.1 ค่า max_depth 10 ค่า min_child_weight 1 ค่า subsample 1 ค่า colsample_bytree 1 ค่า reg_alpha 0 ค่า reg_lambda 0 และ ค่า gamma 0 ประสิทธิภาพดีที่สุดด้วยค่าความถูกต้องร้อยละ 92.44

4.2.3 ผลการพัฒนาแบบจำลอง Hybrid GCN + XGBoost

โดยใช้ค่าพารามิเตอร์ที่ดีที่สุดของแต่ละอัลกอลิทึม โดย GCN กำหนดให้ค่า K เท่ากับ 3 ค่า Dropout 0.1 ค่า Hidden dim 128 ค่า Epochs 150 และ XGBoost กำหนดค่าพารามิเตอร์ ค่า n_estimators (trees) เท่ากับ 800 ค่า learning_rate 0.1 ค่า max_depth 10 ค่า min_child_weight 1 ค่า subsample 1 ค่า colsample_bytree 1 ค่า reg_alpha 0 ค่าlambda 0 และ ค่า gamma 0

ตารางที่ 4-7 ผลลัพธ์ของแบบจำลอง Hybrid GCN + XGBoost

Performance			
Acc(%)	Pre	Rec	F1
85.950	0.874	0.861	0.850

จากตารางที่ 4-7 แสดงให้เห็นว่า GCN + XGBoost ที่นำค่าที่ดีที่สุดของแต่ละอัลกอลิทึม แต่ละตัวมา Hy-brid กัน ค่าความถูกต้องร้อยละ 85.95

ตารางที่ 4-8 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง

Algorithm	Performance			
	Acc(%)	Pre	Re	F1
GCN	78.105	0.800	0.749	0.781
XGBoost	**92.440**	**0.928**	**0.924**	**0.924**
Hybrid	85.950	0.874	0.861	0.845

จากตารางที่ 4-8 เมื่อได้ผลลัพธ์ของ 3 แบบจำลอง ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลอง พบว่าแบบจำลอง **XGBoost** ให้ประสิทธิภาพในการทำนายผลดีที่สุดด้วยค่าความถูกต้องร้อยละ 92.44 ค่าความแม่นยำ 0.9279 ค่าระลึก 0.9244 และค่าการถ่วงดุล 0.9442 ตามลำดับ

4.2.4 ผลการทดสอบประสิทธิภาพของแต่ละแบบจำลองด้วย Unseen Dataset

ตารางที่ 4-9 ผลการทดสอบประสิทธิภาพของแบบจำลองของทั้ง 3 แบบจำลอง

Algorithm	Performance			
	Acc(%)	Pre	Re	F1
GCN	49.20%	0.484	0.240	0.458
XGBoost	86.80%	0.872	0.868	0.868
Hybrid	81.20%	0.755	0.924	0.831

จากตารางที่ 4-9 ผลการทดสอบตัวด้วย Unseen Data จำนวน 500 ระเบียน พบว่าแบบจำลอง XGBoost ให้ประสิทธิภาพในการทำนายผลดีที่สุดด้วยค่าความถูกต้องร้อยละ 86.80 ค่าความแม่นยำ ค่าระลึก ค่าการถ่วงดุล อยู่ที่ประมาณ 0.87

ผลการทดสอบดังกล่าวยืนยันในเชิงประจักษ์ว่า แบบจำลอง XGBoost ที่ผ่านการปรับแต่งพารามิเตอร์ มีความสามารถในการจำแนกกลุ่มลูกค้าผิดนัดชำระหนี้และไม่ผิดนัดชำระหนี้ได้อย่างแม่นยำ อีกทั้งมีทดสอบด้วยข้อมูล Unseen ก็ยังคงมีประสิทธิภาพในการทำนายผลได้ดีที่สุด แสดงให้เห็นว่า แบบจำลองไม่ได้ overfit สำหรับข้อมูลชุดใดชุดหนึ่ง ทำให้สามารถนำไปใช้พยากรณ์ในสถานการณ์จริงได้อย่างมีประสิทธิภาพ ดังนั้น ผู้วิจัยจึงพิจารณาคัดเลือก XGBoost ให้เป็นแบบจำลองขั้นสุดท้าย (Final Model) เพื่อนำไปประยุกต์ใช้งานจริง (Model Deployment) ในรูปแบบของระบบคัดเลือกลูกหนี้ที่จะเข้ามาขอกู้ ซึ่งจะช่วยให้ผู้ปฏิบัติงานสามารถนำกำหนดแนวทางปฏิบัติในการทำงานได้

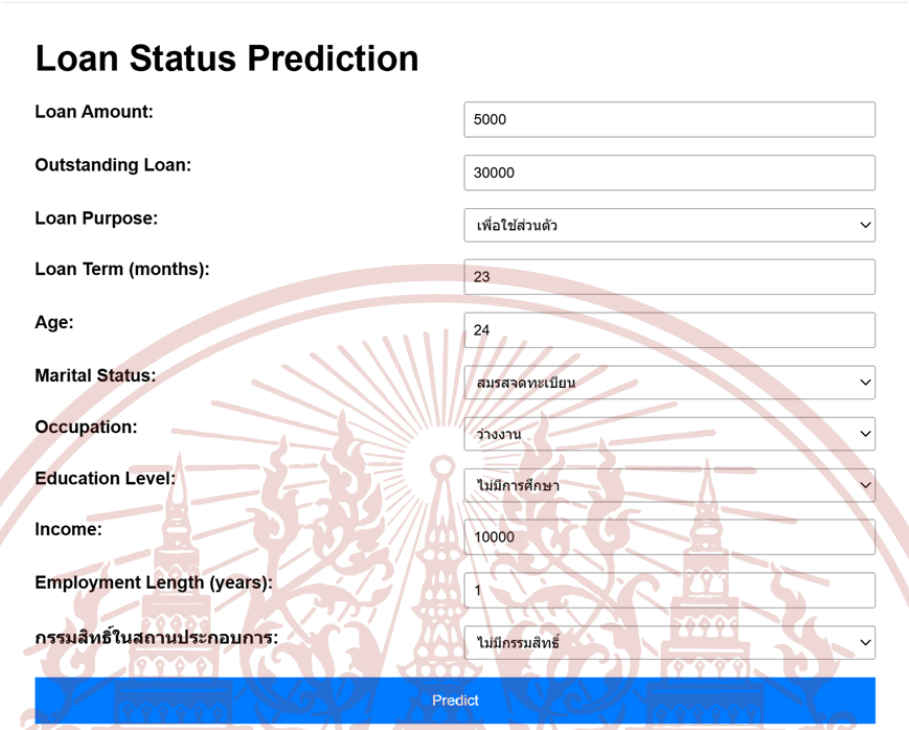
4.3 ผลการนำแบบจำลองไปใช้งาน

ภายหลังจากดำเนินการพัฒนาแบบจำลองและทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมต่าง ๆ สำหรับการพยากรณ์ความเสี่ยงการผิดนัดชำระหนี้สินเชื่อบุคคล ผู้วิจัยได้พิจารณาคัดเลือกแบบจำลองที่มีประสิทธิภาพสูงสุดจากการทดลอง คือ แบบจำลองเอ็กซ์ตรีมกราเดียนท์บูสติง (Extreme Gradient Boosting: XGBoost) เพื่อใช้เป็นแบบจำลองหลักของงานวิจัย

เมื่อได้แบบจำลองที่เหมาะสมแล้ว ผู้วิจัยได้นำแบบจำลอง XGBoost มาทำการบันทึกในรูปแบบไฟล์ .pkl เพื่อรองรับการนำไปใช้งานในสภาพแวดล้อมจริง จากนั้นได้พัฒนาต่อยอดโดยบูรณาการแบบจำลองเข้ากับระบบเว็บแอปพลิเคชัน (Web Application) บน Firebase Console เพื่อจัดทำเป็นระบบประเมินความเสี่ยงสินเชื่อแบบออนไลน์ ซึ่งสามารถให้ผลลัพธ์ได้แบบทันที

4.3.1 User Interface

ระบบถูกออกแบบให้สามารถใช้งานสะดวก โดยผู้ใช้งานสามารถกรอกข้อมูลที่เกี่ยวข้องกับผู้ขอสินเชื่อผ่านแบบฟอร์มบนหน้าเว็บไซต์ เช่น วัตถุประสงค์ของการกู้เงิน วงเงินกู้ ยอดคงค้าง อายุ อาชีพ รายได้ ระยะเวลาการกู้ สถานภาพสมรส ระดับการศึกษา อายุงานและกรรมสิทธิ์ในสถานประกอบการ ซึ่งเป็นข้อมูลที่เป็นตัวแปรสำคัญของแบบจำลอง เมื่อกดยืนยัน ระบบจะดำเนินการส่ง



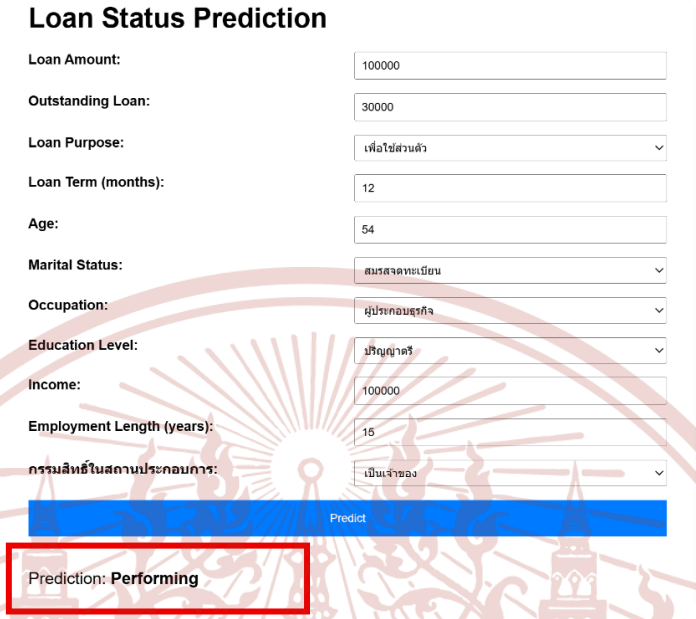
ภาพที่ 4-1 การแสดงผลหน้า User Interface

4.3.2 ระบบประมวลผลข้อมูล (Data Processing System and Result)

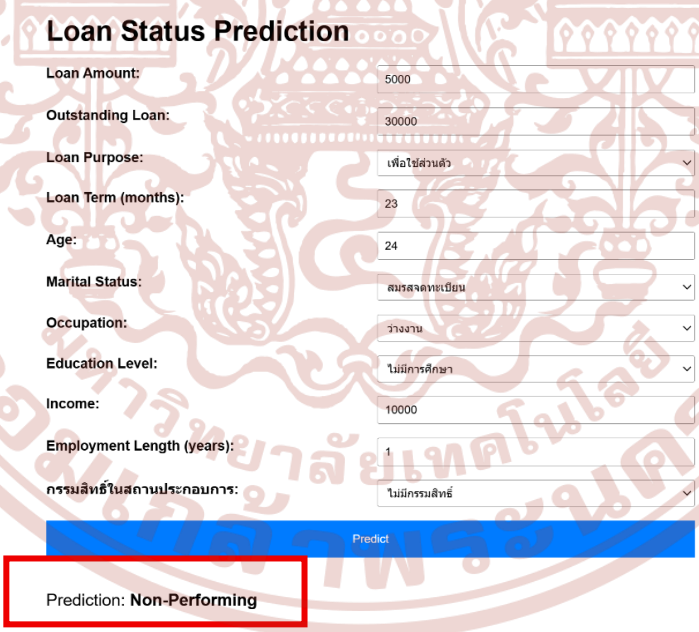
โดยข้อมูลทั้ง 11 รายการจัดถูกเชื่อมต่อกับ XGBoost Final Model ที่เป็นไฟล์ pkl. โดยเมื่อการกรอกข้อมูลครบถ้วน Backend จะเรียกใช้แบบจำลอง XGBoost ที่ถูกติดตั้งไว้เพื่อทำการพยากรณ์ โดยผลลัพธ์ที่ได้จะแสดงสถานะการจำแนกเป็น 2 ระดับ ได้แก่

“Performing” หรือ “ไม่ผิดนัดชำระหนี้” แสดงดังภาพที่ 4-2

“Non- Performing” หรือ “มีความเสี่ยงผิดนัดชำระหนี้” แสดงดังภาพที่ 4-3



ภาพที่ 4-2 การแสดงผลการทำนายประเภทลูกค้าไม่ผิดนัดชำระหนี้



ภาพที่ 4-3 การแสดงผลการทำนายประเภทลูกค้ามีความเสี่ยงผิดนัดชำระหนี้

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องในการพยากรณ์การผิดนัดชำระหนี้สินเชื่อบุคคล โดยใช้ชุดข้อมูลจำนวน 20,000 ระเบียน ซึ่งครอบคลุมข้อมูลระหว่างเดือนมกราคม 2565 ถึงกรกฎาคม 2566 และมีการติดตามพฤติกรรมการชำระหนี้เพิ่มเติมอีก 18 เดือน จนถึงเดือนกุมภาพันธ์ 2568 ข้อมูลประกอบด้วยตัวแปรสำคัญจำนวน 11 ตัวแปร ที่สะท้อนลักษณะทางการเงินและคุณลักษณะของผู้กู้ คือ วัตถุประสงค์ของการกู้เงิน วงเงินกู้ ยอดคงค้าง อายุ อาชีพ รายได้ ระยะเวลาการกู้ สถานภาพสมรส ระดับการศึกษา อายุงาน และกรรมสิทธิ์ในสถานประกอบการ

กระบวนการดำเนินงานเริ่มจากการทำความสะอาดข้อมูล การแปลงตัวแปรให้อยู่ในรูปเชิงตัวเลข และการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE เพื่อแก้ไขปัญหาความไม่สมดุลของคลาส จากนั้นได้ทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึม 3 วิธี ได้แก่ Graph Convolutional Network (GCN), Extreme Gradient Boosting (XGBoost) และแบบจำลองผสมผสาน (GCN + XGBoost) โดยใช้ Stratified K-Fold Cross-Validation (k = 5) เป็นวิธีประเมินผล และวัดประสิทธิภาพด้วย Accuracy, Precision, Recall และ F1-score

ผลของการวิจัยแสดงให้เห็นว่าแบบจำลอง XGBoost มีประสิทธิภาพสูงสุด โดยมีค่า Accuracy ในขั้นตอน Cross-Validation เท่ากับ 92.44% และเมื่อทดสอบกับชุดข้อมูลที่ไม่เคยใช้ในการฝึก จำนวน 500 ตัวอย่าง ยังคงรักษาประสิทธิภาพได้ที่ระดับ 86.80% พร้อมค่า Precision, Recall และ F1-score เฉลี่ยประมาณ 0.87 สะท้อนถึงความสามารถในการทั่วไปของแบบจำลอง และความเหมาะสมในการนำไปใช้งานจริง

ภายหลังการคัดเลือกแบบจำลองขั้นสุดท้าย ผู้วิจัยได้นำ XGBoost ไปบันทึกในรูปแบบไฟล์ .pkl และพัฒนาเป็นระบบเว็บแอปพลิเคชันบนแพลตฟอร์ม Firebase โดยใช้ Cloud Functions ทำหน้าที่เป็น API สำหรับประมวลผลแบบเรียลไทม์ ระบบที่พัฒนาขึ้นสามารถรับข้อมูลจากผู้ใช้งาน ประมวลผลผ่านแบบจำลอง และแสดงผลลัพธ์ในรูปแบบสถานะการจำแนกพร้อมค่าความน่าจะเป็น เพื่อสนับสนุนการตัดสินใจในด้านการให้สินเชื่อ

โดยสรุป งานวิจัยนี้แสดงให้เห็นถึงศักยภาพของการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในการพยากรณ์ความเสี่ยงด้านเครดิตได้อย่างมีประสิทธิภาพ และสามารถต่อยอดจากการวิเคราะห์เชิงวิชาการไปสู่การพัฒนาระบบต้นแบบที่ใช้งานได้จริง ซึ่งมีคุณค่าต่อการบริหารความเสี่ยงของสถาบันการเงินในเชิงปฏิบัติ

บทที่ 5
สรุปผล อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์การผิดนัดชำระหนี้ของลูกค้าสินเชื่อส่วนบุคคล โดยประยุกต์ใช้เทคนิคเหมืองข้อมูล การดำเนินงานวิจัยครอบคลุมกระบวนการเก็บรวบรวมข้อมูล การจัดเตรียมข้อมูล การพัฒนา และประเมินประสิทธิภาพแบบจำลอง และการนำแบบจำลองไปใช้งานจริง ในบทนี้จะนำเสนอข้อมูล 3 ส่วน ดังนี้

5.1 สรุปผลการวิจัย

5.2 การอภิปรายผลการวิจัย

5.3 ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอแนวทางการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อพยากรณ์การผิดนัดชำระหนี้สินเชื่อบุคคล โดยใช้ชุดข้อมูลจำนวน 20,000 ระเบียน ซึ่งครอบคลุมข้อมูลระหว่างเดือนมกราคม 2022 ถึงกรกฎาคม 2023 และติดตามพฤติกรรมการชำระหนี้เพิ่มเติมอีก 18 เดือน จนถึงเดือนกุมภาพันธ์ 2025 ข้อมูลแต่ละระเบียนประกอบด้วยตัวแปรสำคัญจำนวน 11 ตัวแปร ที่เกี่ยวข้องกับลักษณะทางการเงินและคุณลักษณะส่วนบุคคลของผู้กู้

แบบจำลองที่นำมาเปรียบเทียบ ได้แก่ Graph Convolutional Network (GCN), Extreme Gradient Boosting (XGBoost) และแบบจำลองผสมผสาน (GCN + XGBoost) โดยมีการดำเนินการเตรียมข้อมูลอย่างเป็นระบบ ทั้งการทำความสะอาดข้อมูล การแปลงตัวแปรให้อยู่ในรูปเชิงตัวเลข และการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE เพื่อแก้ไขปัญหาความไม่สมดุลของคลาส

ผลการประเมินประสิทธิภาพด้วยเทคนิค Stratified K-Fold Cross-Validation (k = 5) แสดงให้เห็นว่าแบบจำลอง XGBoost ให้ผลลัพธ์ดีที่สุด โดยมีค่า Accuracy เท่ากับ 92.44% ค่า Precision เท่ากับ 0.93 ค่า Recall เท่ากับ 0.92 และค่า F1-score เท่ากับ 0.92

เมื่อทดสอบเพิ่มเติมกับชุดข้อมูลที่ไม่เคยใช้ในการฝึกจำนวน 500 ตัวอย่าง พบว่า XGBoost ยังคงมีประสิทธิภาพสูงสุด โดยให้ค่า Accuracy เท่ากับ 86.80% และมีค่า Precision, Recall และ F1-score เฉลี่ยอยู่ที่ประมาณ 0.87 สะท้อนถึงความสามารถของแบบจำลองในการประยุกต์ใช้กับข้อมูลใหม่ได้อย่างมีประสิทธิภาพ

5.2 การอภิปรายผลการวิจัย

จากงานวิจัยนี้แสดงให้เห็นว่า การนำเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้เพื่อพยากรณ์การผิดนัดชำระหนี้สินเชื่อบุคคล สะท้อนให้เห็นถึงศักยภาพของโมเดลที่แตกต่างกันอย่างชัดเจน โดยเฉพาะอย่างยิ่งโมเดล XGBoost ที่สามารถให้ผลลัพธ์เหนือกว่าโมเดล Graph Convolutional Network (GCN) และโมเดลผสมผสาน GCN+XGBoost ทั้งในขั้นตอนการประเมินด้วย Stratified K-Fold Cross-Validation และการทดสอบกับข้อมูลใหม่ที่ไม่เคยใช้ในการฝึก (Unseen Data)

เมื่อเปรียบเทียบกับ GCN และโมเดล Hybrid พบว่าประสิทธิภาพของทั้งสองโมเดลด้อยกว่า XGBoost โดย GCN แม้จะมีข้อได้เปรียบในการเรียนรู้โครงสร้างเชิงกราฟ แต่ข้อมูลสินเชื่อบุคคลที่ใช้ในงานนี้ไม่ได้มีความสัมพันธ์เชิงโครงสร้าง (Graph Structure) ที่ชัดเจนเพียงพอ ทำให้แบบจำลองดึงประโยชน์จากการเรียนรู้เชิงกราฟออกมาได้ไม่เต็มที่ นอกจากนี้ เมื่อพิจารณาในเชิงลักษณะของข้อมูล พบว่าข้อมูลสินเชื่อบุคคลที่ใช้ในงานวิจัยนี้มีลักษณะเป็นข้อมูลแบบตาราง (Tabular Data) ซึ่งประกอบด้วยตัวแปรเชิงคุณลักษณะของผู้กู้เป็นหลัก เช่น รายได้ อายุ วงเงินกู้ และระยะเวลาการกู้ เป็นต้น โดยตัวแปรเหล่านี้เป็นข้อมูลเชิงอิสระที่สะท้อนสถานะทางการเงินและพฤติกรรมของแต่ละบุคคลเป็นรายตัว มากกว่าจะเป็นข้อมูลที่มีความสัมพันธ์เชิงโครงสร้าง แม้ว่าในงานวิจัยนี้จะมีการสร้างโครงสร้างกราฟโดยใช้วิธี k-Nearest Neighbors (kNN) เพื่อเชื่อมโยงตัวอย่างที่มีความคล้ายคลึงกัน แต่ความสัมพันธ์ดังกล่าวเป็นความสัมพันธ์เชิงคำนวณ (Similarity-based Relationship) ที่ถูกสร้างขึ้นจากระยะห่างของตัวแปรเชิงตัวเลข ไม่ใช่ความสัมพันธ์เชิงโครงสร้างที่เกิดขึ้นจริง ด้วยเหตุนี้ โครงสร้างกราฟที่ได้ไม่สามารถสะท้อนบริบทเชิงความสัมพันธ์ที่มีนัยสำคัญต่อพฤติกรรมการผิดนัดชำระหนี้ได้อย่างแท้จริง ส่งผลให้ GCN ซึ่งพึ่งพาการเรียนรู้จากความสัมพันธ์ระหว่างโหนด (nodes) และบริบทของเพื่อนบ้าน (neighborhood aggregation) ไม่สามารถแสดงศักยภาพได้อย่างเต็มที่ ในขณะที่ XGBoost มีความเหมาะสมกับข้อมูลแบบตาราง เนื่องจากสามารถเรียนรู้ความสัมพันธ์เชิงไม่เชิงเส้น (nonlinear relationships) และเงื่อนไขเชิงเกณฑ์ (threshold-based rules) ระหว่างตัวแปรได้อย่างมีประสิทธิภาพ โดยไม่จำเป็นต้องอาศัยโครงสร้างกราฟเพิ่มเติม ซึ่งสอดคล้องกับงานวิจัยของ Gunnarsson et al.[4] ที่อธิบายว่าโมเดลแบบกลุ่มที่ใช้ต้นไม้ตัดสินใจ (Ensemble Tree-based Models) เช่น XGBoost มักจะแสดงประสิทธิภาพที่เหนือกว่าเมื่อเทียบกับสถาปัตยกรรมเชิงลึก (deep learning) และแบบผสม (hybrid) ในงานด้านการให้คะแนนเครดิต ดังนั้น ประสิทธิภาพที่เหนือกว่าของ XGBoost แบบเดี่ยวที่สังเกตได้ในงานวิจัยนี้จึงสอดคล้องกับหลักฐานเชิงประจักษ์ที่มีอยู่แล้วในวรรณกรรมด้านการสร้างแบบจำลองความเสี่ยงด้านเครดิต และงานวิจัยของกมลชนก [1] ซึ่งรายงานว่า XGBoost แบบเดี่ยวในงานพยากรณ์ความเสี่ยงด้านเครดิตมีผลลัพธ์สูงกว่า voting ensemble ของหลายๆอัลกอริทึม ซึ่งงานวิจัยนี้ตอกย้ำข้อเท็จจริงที่ว่า XGBoost ยังคงเป็นหนึ่งในอัลกอริทึมที่มีประสิทธิภาพสูงสุดสำหรับการพยากรณ์การผิดนัดชำระหนี้ในบริบทของข้อมูลที่มีตัวแปรทางการเงินและคุณลักษณะส่วนบุคคลของผู้กู้ ผลลัพธ์นี้สอดคล้องกับงานของ Liu, Zhang & Fan [9] ที่แสดงให้เห็นว่าแม้โมเดล Hybrid XGBoost+GNN จะมีศักยภาพ แต่ XGBoost แบบเดี่ยวก็ยังคงให้ผลลัพธ์ที่แข็งแกร่งและเสถียรในหลายกรณี ซึ่งงานวิจัยนี้ยืนยันว่า XGBoost เป็นโมเดลที่เหมาะสมที่สุดสำหรับการพยากรณ์การผิดนัดชำระหนี้สินเชื่อบุคคลในบริบทของข้อมูลเชิงตัวแปร และการพยายามสร้างโมเดล Hybrid ที่ไม่สอดคล้องกับลักษณะข้อมูล

และ ลักษณะความสัมพันธ์ของตัวแปรที่ไม่ชัดเจน อาจไม่ได้เพิ่มคุณภาพของการทำนาย แต่กลับลดประสิทธิภาพลง ซึ่งเป็นข้อค้นพบที่มีนัยสำคัญต่อการเลือกใช้โมเดลในงานวิจัยและการประยุกต์ใช้จริงในสถาบันการเงิน

ดังนั้น ผลลัพธ์ที่ XGBoost มีประสิทธิภาพเหนือกว่า GCN และโมเดล Hybrid ในงานวิจัยนี้ จึงสามารถอธิบายได้จากความสอดคล้องระหว่างลักษณะของข้อมูล (tabular data) และกลไกการเรียนรู้ของโมเดล ซึ่งไม่ต้องพึ่งพาความสัมพันธ์เชิงโครงสร้างที่ซับซ้อน ทั้งนี้สะท้อนให้เห็นว่าการเลือกใช้โมเดลควรพิจารณาให้เหมาะสมกับธรรมชาติของข้อมูลเป็นสำคัญ และในกรณีที่ข้อมูลไม่ได้มีโครงสร้างเชิงกราฟที่ชัดเจน การใช้โมเดลเชิงกราฟอาจไม่ก่อให้เกิดประโยชน์เพิ่มเติม และอาจเพิ่มความซับซ้อนโดยไม่จำเป็น

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

5.3 ข้อเสนอแนะ

5.3.1. ข้อเสนอแนะเชิงการประยุกต์ใช้ (Practical Implications)

5.3.1.1 คุณภาพของข้อมูล ข้อมูลต้องมาจากแหล่งที่น่าเชื่อถือ ยิ่งข้อมูลมีคุณภาพก็จะสามารถ เพื่อประสิทธิภาพในการเรียนรู้ของแบบจำลองได้อย่างครอบคลุม

5.3.1.2 ด้วยข้อจำกัดบางประการของข้อมูล ทำให้บางแบบจำลองไม่สามารถถูกนำมาใช้ได้อย่างเต็มประสิทธิภาพตามศักยภาพที่ควรจะเป็น

5.3.1.3 สถาบันการเงินสามารถนำแบบจำลองไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจ (Decision Support Tool) ในกระบวนการพิจารณาอนุมัติสินเชื่อ

5.3.2. ข้อเสนอแนะสำหรับการวิจัยในอนาคต (Future Research Directions)

5.3.2.1 การเพิ่มจำนวนและความหลากหลายของตัวแปร งานวิจัยในอนาคตอาจพิจารณาเพิ่มข้อมูลตัวแปรเพิ่มเติม ข้อมูลเครดิตบูโร หรือข้อมูลเศรษฐกิจมหภาคเพื่อเพิ่มศักยภาพในการทายผลของแบบจำลอง

5.3.2.2 การทดลองใช้อัลกอริทึมขั้นสูงเพิ่มเติม อาจทดลองใช้อัลกอริทึมอื่น เช่น LightGBM, CatBoost หรือ Deep Learning เพื่อเปรียบเทียบประสิทธิภาพเพิ่มเติม

5.3.3. ข้อจำกัดของงานวิจัย (Limitations)

5.3.3.1 ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลแบบตารางและไม่มีความสัมพันธ์เชิงโครงสร้างที่ชัดเจนระหว่างตัวอย่างข้อมูล ส่งผลให้ GCN ไม่สามารถแสดงศักยภาพได้อย่างเต็มที่

5.3.3.2 กระบวนการเตรียมข้อมูลมีการตัดข้อมูลที่มีค่าขาดหาย (missing values) ออกเป็นจำนวนมาก ซึ่งอาจก่อให้เกิดความเอนเอียงของข้อมูลและอาจส่งผลต่อความสามารถในการนำโมเดลไปประยุกต์ใช้กับข้อมูลจริง

5.3.3.3 การใช้เทคนิค SMOTE เพื่อแก้ปัญหาความไม่สมดุลของข้อมูล แม้ว่าจะช่วยเพิ่มประสิทธิภาพของโมเดล แต่อาจไม่สามารถสะท้อนพฤติกรรมของลูกหนี้ผิดนัดได้อย่างสมบูรณ์

5.3.3.4 จำนวนตัวแปรที่ใช้ในการสร้างโมเดลมีเพียง 11 ตัวแปร ซึ่งอาจยังไม่ครอบคลุมปัจจัยสำคัญอื่น ๆ เช่น อัตราส่วนทางการเงิน หรือพฤติกรรมการใช้จ่าย

5.3.3.5 ข้อมูลที่ใช้ในงานวิจัยนี้มาจากสถาบันการเงินเพียงแห่งเดียว และไม่ได้รวมข้อมูลจากเครดิตบูโร ซึ่งเป็นแหล่งข้อมูลสำคัญที่สะท้อนพฤติกรรมทางการเงินโดยรวมของผู้กู้ ส่งผลให้แบบจำลองไม่สามารถสะท้อนความเสี่ยงด้านเครดิตได้อย่างครบถ้วน และจำกัดความสามารถในการนำไปประยุกต์ใช้กับสถาบันการเงินอื่นที่มีลักษณะลูกค้าและนโยบายสินเชื่อแตกต่างกัน

5.3.3.5 ข้อมูลที่ใช้ครอบคลุมเฉพาะช่วงเวลา พ.ศ. 2565–2568 ซึ่งเป็นช่วงสถานการณ์ COVID-19 และการเปลี่ยนแปลงของอัตราดอกเบี้ยนโยบาย ดังนั้น แบบจำลองที่พัฒนาขึ้นอาจมีข้อจำกัดในการนำไปใช้กับช่วงเวลาอื่นที่มีสภาวะเศรษฐกิจแตกต่าง

บรรณานุกรม

1. กมลชนก รุ่งเจริญ. (2565). การพยากรณ์ลูกหนี้ที่ค้างชำระหนี้ของสินเชื่อเพื่อที่อยู่อาศัยด้วยเทคนิคเหมืองข้อมูล. วิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบสารสนเทศเพื่อการจัดการ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
2. เครือวัลย์ เนตรพนา และ ศิริสรรพ เหล่าหะเกียรติ. (2566). การวิเคราะห์ความเสี่ยงในการผิดนัดชำระของลูกหนี้บัตรเครดิต. วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
3. Chang and et al. (2025). Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers Department of Operations and Information Management, Aston Business School, Aston University, Birmingham B4 7ET, UK
4. Gunnarsson et al. [4] (2022). Deep learning for credit scoring do or don't. European Journal of Operational Research, 295(1), 292–305.
5. Sukharev and et al. (2019). EWS-GCN: Edge Weight-Shared Graph Convolutional Network for Transactional Banking Data. arXiv. 2009.14588.
6. Sun and et al. (2024). Applying Hybrid Graph Neural Networks to Strengthen Credit Risk Analysis. arXiv. 2410.04283.
7. Shi and et al. (2024). Improved Credit Risk Prediction based on an Integrated Graph Representation Learning Approach with Graph Transformation. European Journal of Operational Research, Vol.315(2) : 786-801.
8. Wang and et al. (2021). Temporal-Aware Graph Neural Network for Credit Risk Prediction. SIAM International Conference on Data Mining (SDM), เมษายน 2564
9. Liu Zhang and Fan. (2022) A Two-Stage hybrid Credit Risk Prediction Model based on XGBoost and Graph-based Deep Neural Network. Expert Systems with Applications, Vol.195:116624
10. Albanesi and Vamossy. (2019). Predicting Consumer Default: A Deep Learning Approach. NBER, Working Paper No. 26165 : 1-72
11. ทิพย์ธิดา วงศ์พิพันธ์. (2556). การทำเหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อของบริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน). วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์และการสื่อสาร คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์.
12. ญาณินี สิทธิสาร และ พุธษดี ศิริแสงตระกูล. (2561). โมเดลทำนายความสามารถในการชำระหนี้เงินกู้ของสมาชิกสำนักงานส่งเสริมสวัสดิการและสวัสดิการภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี. การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 19 วันที่ 9 มีนาคม พ.ศ.2561

13. ขวัญนภา พิมพ์ชารีย์. (2564). การใช้เหมืองข้อมูลสำหรับพิจารณาการให้สินเชื่อสำหรับธนาคาร. วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม.
14. ลัลธริดา จันทร์สุข. (2561). การพยากรณ์ความล้มเหลวและภาวะกดดันทางการเงินของบริษัทในตลาดหลักทรัพย์แห่งประเทศไทยโดยแบบจำลอง CAMPBELL. วิทยาศาสตรมหาบัณฑิต สาขาวิชาการบริหารการเงิน คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์.
15. นพมาศ บุญปั้น. (2564). การประยุกต์เหมืองข้อมูลเพื่อพยากรณ์ปัจจัยที่ทำให้เกิดหนี้ค้างภายใต้การให้บริการด้านสินเชื่อโครงการสินเชื่อดิจิทัล. วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์และเทคโนโลยีการเงิน คณะวิศวกรรมศาสตร์ มหาวิทยาลัยหอการค้าไทย.
16. รังรองรัตน์ อินทร์พยุง. (2021). แพลตฟอร์มการกู้ยืมระหว่างบุคคลกับบุคคลผ่านช่องทางออนไลน์จากข้อมูลทางเลือกแบบทดสอบบุคลิกภาพและแบบประเมินเจตคติต่อเงิน. วิทยาศาสตรมหาบัณฑิต สาขาวิชาธุรกิจเทคโนโลยีและการจัดการนวัตกรรม สหสาขาวิชาธุรกิจเทคโนโลยีและการจัดการนวัตกรรม จุฬาลงกรณ์มหาวิทยาลัย.
17. Zhang and et al. (2021) Research on Personal Credit Scoring Model based on Multi-Source data. Journal of Physics Conference Series, Vol.1437(1) 012053 : 1-14.
18. Bae and Kim. (2015). A Personal Credit Rating Prediction Model Using Data Mining in Smart Ubiquitous Environments. International Journal of Distributed Sensor Networks, Vol.11:179060
19. Ampountolas and et al. (2021) A Machine Learning Approach for Micro-Credit Scoring. Risks 2021, Vol.9(3):50 : 1–20.
20. Minakawa and et al. (2021) Firm Default Prediction by GNN with Gravity-Model Informed Neighbor Sampling. The Review of Socionetwork Strategies (Springer), Vol.18(2) : 303–328.
21. ชลลดา ม่วงธนัง, สุรศักดิ์ มังสิงห์และ นิเวศ จิระวิชิตชัย. (2564). การพัฒนาแบบจำลองการให้คะแนนสินเชื่อโดยใช้เทคนิคเหมืองข้อมูล. วารสารวิทยาศาสตร์และเทคโนโลยี หัวเฉียวเฉลิมพระเกียรติ. ปีที่ 7 ฉบับที่ 1 (2564): มกราคม - มิถุนายน 2564
22. ปารดา ศัสตุระ. (2565). การประยุกต์เหมืองข้อมูลในการพยากรณ์สถานะ NPLs ของลูกค้าสินเชื่อ กรณีศึกษา ธนาคารออมสิน สาขามหาวิทยาลัยเชียงใหม่. การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา ครั้งที่ 17 วันพุธที่ 10 สิงหาคม พ.ศ.2565.

ภาคผนวก

หนังสือขอความอนุเคราะห์เก็บรวบรวมข้อมูล



ที่ อว ๗๑๐๗/๐๔๔

คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
๑๕๑๘ ถนนประชาราษฎร์ ๑ แขวงวงศ์สว่าง
เขตบางซื่อ กรุงเทพฯ ๑๐๘๐๐

๒๐ มกราคม ๒๕๖๙

เรื่อง ขอความอนุเคราะห์เก็บรวบรวมข้อมูล

เรียน ผู้อำนวยการฝ่ายบริหารความเสี่ยงเครดิต
ธนาคารออมสินสำนักงานใหญ่

ด้วยนายนิธิคุณ สุนทรานุรักษ์ รหัสประจำตัวนักศึกษา ๖๗-๐๗๐๒๑๘- ๕๖๐๖-๐ นักศึกษาระดับบัณฑิตศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบสารสนเทศเพื่อการจัดการ ภาควิชาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ได้จัดทำการค้นคว้าอิสระ เรื่อง “การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเหมืองข้อมูล” โดยมีผู้ช่วยศาสตราจารย์ ดร.วัชรีวรรณ จิตต์สกุลเป็นประธานคณะกรรมการที่ปรึกษาการค้นคว้าอิสระ ในการนี้นักศึกษามีความประสงค์ขอความอนุเคราะห์เก็บรวบรวมข้อมูลลูกหนี้ของสินเชื่อบุคคลประกอบด้วย เพื่อประกอบการทำการค้นคว้าอิสระ ดังต่อไปนี้

๑. จำนวนเงินขอกู้
๒. จำนวนเงินคงเหลือ
๓. จุดประสงค์การกู้
๔. อายุ
๕. สถานะสมรส
๖. อาชีพ
๗. ระดับการศึกษา
๘. รายได้
๙. ค่าใช้จ่าย
๑๐. สถานะของที่อยู่อาศัยในปัจจุบัน
๑๑. ระยะเวลาในการทำงาน
๑๒. จำนวนงวดในการผ่อนชำระ
๑๓. ลักษณะสถานประกอบการ
๑๔. สถานะการกู้ร่วม

จึงเรียนมาเพื่อโปรดพิจารณาให้ความอนุเคราะห์ จักขอบพระคุณยิ่ง

ขอแสดงความนับถือ

(ผู้ช่วยศาสตราจารย์ ดร.ศักดิ์ชาย ตั้งวรรณวิทย์)
รองคณบดีฝ่ายวิชาการและวิจัย
ปฏิบัติการแทนคณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล

ภาควิชาการจัดการเทคโนโลยีสารสนเทศ
โทรศัพท์ ๐ ๒๕๕๕ ๒๗๒๒
www.itd.kmutnb.ac.th

ประวัติผู้เขียน

ชื่อ	นิธิคุณ สุนทรานุรักษ์
ชื่อการค้นคว้าอิสระ	การพยากรณ์ลูกหนี้ค้างชำระของสินเชื่อบุคคลด้วยเทคนิคเหมืองข้อมูล
สาขาวิชา	ระบบสารสนเทศเพื่อการจัดการ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ประวัติ	ปริญญาตรี บริหารธุรกิจบัณฑิต สาขาบัญชี มหาลัยอัสสัมชัญ

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ