

การสร้างแบบจำลอง Credit Scorecard ด้วยเทคนิค Weight of Evidence (WoE) และ Logistic Regression



1. บทนำ (Introduction)

ในภูมิทัศน์ของสถาบันการเงิน การประเมินความเสี่ยงสินเชื่อ (Credit Risk Modeling)

ถือเป็นหัวใจสำคัญของการดำเนินธุรกิจ

การมีแบบจำลองที่สามารถประเมินความน่าจะเป็นในการผิดนัดชำระหนี้ของผู้ขอสินเชื่อได้อย่างเป็นระบบและแม่นยำ ไม่เพียงแต่ช่วยลดความเสี่ยงหนี้เสีย (Non-Performing Loans)

แต่ยังช่วยให้กระบวนการพิจารณาสินเชื่อเป็นไปอย่างรวดเร็ว มีมาตรฐาน และโปร่งใสอีกด้วย

แบบจำลองความเสี่ยงจึงเป็นเครื่องมือเชิงกลยุทธ์ที่ช่วยสร้างความได้เปรียบในการแข่งขันและรักษาเสถียรภาพทางการเงินขององค์กร

เอกสารฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอแนวทางการสร้าง Credit Scorecard อย่างเป็นขั้นตอน

โดยเริ่มต้นตั้งแต่การสำรวจและเตรียมข้อมูล การแปลงคุณลักษณะด้วยเทคนิค Weight of Evidence (WoE)

ซึ่งเป็นวิธีการที่ทรงพลังในการสกัดข้อมูลเชิงลึกเกี่ยวกับความเสี่ยง ไปจนถึงการสร้างแบบจำลอง Logistic

Regression ที่สามารถตีความผลลัพธ์ได้อย่างชัดเจน และท้ายที่สุดคือการพัฒนาแบบจำลองดังกล่าวให้กลายเป็น

Scorecard ที่สามารถนำไปใช้งานได้จริงในทางปฏิบัติ โดยกระบวนการทั้งหมดจะอ้างอิงจากชุดข้อมูล

credit_risk_dataset.csv เพื่อเป็นกรณีศึกษาที่จับต้องได้

เนื้อหาในส่วนถัดไปจะเริ่มต้นด้วยขั้นตอนพื้นฐานที่สำคัญที่สุด คือการสำรวจและเตรียมข้อมูล เพื่อสร้างความเข้าใจในโครงสร้างและความสัมพันธ์ของข้อมูล ก่อนที่จะนำไปสู่การสร้างแบบจำลองในลำดับต่อไป

2. การสำรวจและเตรียมข้อมูล (Data Exploration and Preparation)

ขั้นตอนการสำรวจและเตรียมข้อมูลเปรียบเสมือนการวางรากฐานที่มั่นคงสำหรับการสร้างแบบจำลองที่มีประสิทธิภาพ ในส่วนนี้ เราจะทำการตรวจสอบภาพรวมของข้อมูล ระบุความสัมพันธ์ระหว่างตัวแปรต่างๆ และค้นหาปัญหาที่อาจส่งผลกระทบต่อความแม่นยำของแบบจำลอง เช่น ภาวะร่วมเส้นตรง (Multicollinearity) เพื่อให้มั่นใจว่าข้อมูลมีความพร้อมและเหมาะสมสำหรับการวิเคราะห์ในขั้นตอนถัดไป

ภาพรวมชุดข้อมูล

ชุดข้อมูลที่ใช้ในการวิเคราะห์นี้ประกอบด้วยข้อมูลผู้ขอสินเชื่อจำนวน 32,581 รายการ และมีตัวแปรทั้งหมด 12 ตัวแปร โดยมีรายละเอียดของตัวแปรที่สำคัญดังนี้

Variable	Description
person_age	อายุของผู้ขอสินเชื่อ
person_income	รายได้ต่อปี
person_home_ownership	สถานะการครอบครองที่อยู่อาศัย (เช่น เช่า, ผ่อน, เป็นเจ้าของ)
person_emp_length	อายุงาน (ปี)
loan_intent	วัตถุประสงค์ในการขอสินเชื่อ
loan_grade	เกรดของสินเชื่อ
loan_amnt	วงเงินที่ขอสินเชื่อ
loan_int_rate	อัตราดอกเบี้ย
loan_status	สถานะของสินเชื่อ (0 = ไม่ผิดนัดชำระ, 1 = ผิดนัดชำระ)
loan_percent_income	สัดส่วนวงเงินสินเชื่อต่อรายได้
cb_person_default_on_file	ประวัติการผิดนัดชำระในอดีต
cb_person_cred_hist_length	ระยะเวลาของประวัติสินเชื่อ (ปี)

การวิเคราะห์ความสัมพันธ์และภาวะร่วมเส้นตรง (Correlation and Multicollinearity Analysis)

จากการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณผ่าน Correlation Heatmap

พบประเด็นที่น่าสนใจหลายประการ:

- ความสัมพันธ์เชิงบวกที่สูงสุด: คู่ตัวแปรที่มีความสัมพันธ์เชิงบวกสูงสุดคือ Age และ Cb_Cred_Hist_Length โดยมีค่าสหสัมพันธ์ (r) เท่ากับ 0.86 ซึ่งเป็นความสัมพันธ์ที่สมเหตุสมผลอย่างยิ่งเนื่องจากผู้ขอสินเชื่อที่มีอายุมากย่อมมีแนวโน้มที่จะมีประวัติทางการเงินและประวัติสินเชื่อที่ยาวนานกว่า
- ความสัมพันธ์กับสถานะความเสี่ยง (Loan_Status):
 - Loan_Percent_Income ($r = 0.38$): มีความสัมพันธ์เชิงบวกในระดับปานกลางกับความเสี่ยง ซึ่งหมายความว่ายิ่งสัดส่วนภาระหนี้ต่อรายได้สูงขึ้นเท่าใด โอกาสในการผิดนัดชำระหนี้ก็จะยิ่งสูงขึ้นตามไปด้วย
 - Loan_Int_Rate ($r = 0.34$): อัตราดอกเบี้ยมีความสัมพันธ์เชิงบวกกับความเสี่ยง โดยกลุ่มลูกค้าที่มีความเสี่ยงสูงมักจะได้รับอัตราดอกเบี้ยที่สูงกว่า
 - Income ($r = -0.14$): รายได้มีความสัมพันธ์เชิงลบเล็กน้อยกับความเสี่ยง บ่งชี้ว่าผู้ที่มีรายได้สูงมีแนวโน้มที่จะผิดนัดชำระหนี้้น้อยกว่า

นอกจากการวิเคราะห์ความสัมพันธ์แล้ว เรายังได้ตรวจสอบปัญหาภาวะร่วมเส้นตรง (Multicollinearity)

ด้วยการคำนวณค่า Variance Inflation Factor (VIF) ซึ่งได้ผลลัพธ์ดังตาราง

Feature	VIF
Age	3.941476
Income	1.508461
Emp_Length	1.064903
Loan_Amnt	2.174360
Loan_Int_Rate	1.159783
Loan_Status	1.346153
Loan_Percent_Income	2.367170
Cb_Cred_Hist_Length	3.854312

ผลการตรวจสอบพบว่าค่า VIF ของทุกตัวแปรอยู่ในช่วง 1-5 ซึ่งเป็นเกณฑ์ที่ยอมรับได้และบ่งชี้ว่าชุดข้อมูลนี้ไม่มีปัญหาภาวะร่วมเส้นตรงที่รุนแรง ที่จะส่งผลกระทบต่อเสถียรภาพของแบบจำลอง

หลังจากทำความเข้าใจภาพรวมและความสัมพันธ์ของข้อมูลแล้ว

ขั้นตอนต่อไปคือการแปลงข้อมูลเหล่านี้ให้อยู่ในรูปแบบที่เหมาะสมกับการสร้างแบบจำลอง Logistic Regression ด้วยเทคนิค Weight of Evidence

3. การแปลงคุณลักษณะด้วย Weight of Evidence (WoE)

Weight of Evidence (WoE) เป็นเทคนิคการแปลงคุณลักษณะที่ทรงพลังในการสร้าง Credit Scorecard โดยมีเป้าหมายเพื่อแปลงตัวแปรกลุ่ม (Categorical) และตัวแปรต่อเนื่อง (Continuous) ที่ผ่านการจัดกลุ่ม (Binning) ให้กลายเป็นค่าตัวเลขที่สะท้อนถึงพลังในการจำแนกกลุ่มลูกค้าดี (Non-default) และกลุ่มลูกค้าเสีย (Default) ได้อย่างชัดเจน นอกจากนี้ เรายังใช้ค่า Information Value (IV) ซึ่งคำนวณจาก WoE เพื่อเป็นตัวชี้วัดพลังในการพยากรณ์ของตัวแปรแต่ละตัว

หลักการของ WoE และ Information Value (IV)

ค่า WoE

สำหรับแต่ละกลุ่มของตัวแปรจะถูกคำนวณจากลอการิทึมธรรมชาติของอัตราส่วนระหว่างการกระจายตัวของกลุ่มลูกค้าดีต่อกลุ่มลูกค้าเสีย ดังสูตรต่อไปนี้:

$$\text{WoE} = \ln(\% \text{ Non-default} / \% \text{ Default})$$

ค่า WoE ที่เป็นบวกหมายถึงกลุ่มนั้นมีสัดส่วนลูกค้าดีมากกว่าลูกค้าเสีย (ความเสี่ยงต่ำ) ในขณะที่ค่า WoE ที่เป็นลบหมายถึงกลุ่มนั้นมีสัดส่วนลูกค้าเสียมากกว่า (ความเสี่ยงสูง) ส่วนค่า IV จะเป็นตัวสรุปพลังการพยากรณ์ของตัวแปรทั้งหมด โดยมีเกณฑ์การแปลความหมายดังนี้

ค่า IV	การแปลความหมาย
< 0.02	ไม่มีอำนาจในการพยากรณ์
0.02 - 0.1	อำนาจการพยากรณ์ต่ำ
0.1 - 0.3	อำนาจการพยากรณ์ปานกลาง
0.3 - 0.5	อำนาจการพยากรณ์สูง
> 0.5	สูงผิดปกติ (อาจเกิด Overfitting)

ผลการแปลง WoE และการประเมิน IV ของตัวแปรสำคัญ

Home_Ownership (สถานะการครอบครองที่อยู่อาศัย)

- Information Value (IV): 0.377 (อำนาจการพยากรณ์สูง)

Value	WoE	IV
RENT	-0.50	0.1451
OWN	1.24	0.0827
MORTGAGE	0.66	0.1483

Value	WoE	IV
OTHER	-0.47	0.0008

การตีความ: กลุ่มผู้ขอสินเชื่อที่เป็นเจ้าของบ้าน (OWN) มีค่า WoE เป็นบวกสูงที่สุด (1.24) ซึ่งบ่งชี้ว่าเป็นกลุ่มที่มีความเสี่ยงต่ำที่สุดอย่างชัดเจน ในทางกลับกัน กลุ่มที่ยังเช่าที่อยู่อาศัย (RENT) มีค่า WoE เป็นลบ (-0.50) แสดงถึงความเสี่ยงที่สูงกว่า

income_bin (กลุ่มรายได้)

- Information Value (IV): 0.462 (อำนาจการพยากรณ์สูง)

Value	WoE	IV
(86000.0, 6000000.0]	1.02	0.1513
(63000.0, 86000.0]	0.49	0.0421
(49000.0, 63000.0]	0.23	0.0094
(35000.0, 49000.0]	-0.09	0.0018
(3999.0, 35000.0]	-1.01	0.2571

การตีความ: กลุ่มที่มีรายได้สูงที่สุดมีค่า WoE เป็นบวกมาก (1.02) สะท้อนถึงความเสี่ยงที่ต่ำ ในขณะที่กลุ่มที่มีรายได้น้อยที่สุดมีค่า WoE เป็นลบมาก (-1.01) ซึ่งชี้ให้เห็นถึงความเสี่ยงที่สูงที่สุด

Loan_Percent_Income_bin (สัดส่วนหนี้ต่อรายได้)

- Information Value (IV): 0.707 (สูงผิดปกติ)

Value	WoE	IV
(-0.01, 0.08]	0.77	0.1096
(0.08, 0.12]	0.67	0.0633
(0.12, 0.18]	0.44	0.0391
(0.18, 0.25]	0.16	0.0046
(0.25, 0.83]	-1.39	0.4900

การตีความ: ตัวแปรนี้มีพลังในการพยากรณ์สูงที่สุด โดยกลุ่มผู้ขอสินเชื่อที่มีสัดส่วนหนี้ต่อรายได้สูง (0.25, 0.83] มีค่า WoE เป็นลบมากถึง -1.39 ซึ่งสะท้อนความเสี่ยงสูงสุดอย่างมีนัยสำคัญ ค่า IV ที่สูงกว่า 0.5 อาจเป็นสัญญาณของการรั่วไหลของข้อมูล (Data Leakage) หรือบ่งชี้ว่าตัวแปรนี้มีความสัมพันธ์กับผลลัพธ์ใกล้เคียงเกินไป ซึ่งต้องได้รับการตรวจสอบเพิ่มเติมในการใช้งานจริงเพื่อป้องกันการ Overfitting

Emp_Length_bin (กลุ่มอายุงาน)

- Information Value (IV): 0.061 (อำนาจการพยากรณ์ต่ำ)

Value	WoE	IV
10-40	0.35	0.0124
8-10	0.30	0.0075
6-8	0.19	0.0051
4-6	0.12	0.0023
2-4	-0.07	0.0011
<2	-0.32	0.0245
Missing	-0.50	0.0078
>40	-0.58	0.0000

การตีความ: แม้จะมี IV ต่ำ แต่ยังเห็นแนวโน้มที่ชัดเจนว่ากลุ่มที่มีอายุงานสั้น (<2), กลุ่มที่ไม่มีข้อมูล (Missing) และกลุ่มที่มีอายุงานยาวนานผิดปกติ (>40) มีความเสี่ยงสูงกว่า (WoE เป็นลบ) เมื่อเทียบกับกลุ่มที่มีอายุงานในเกณฑ์ปกติ

หลังจากแปลงตัวแปรที่มีนัยสำคัญให้อยู่ในรูปของค่า WoE เรียบร้อยแล้ว ข้อมูลเหล่านี้ก็พร้อมสำหรับการนำไปสร้างแบบจำลอง Logistic Regression ในขั้นตอนต่อไป

4. การสร้างแบบจำลอง Logistic Regression

เราเลือกใช้แบบจำลอง Logistic Regression เนื่องจากเป็นแบบจำลองมาตรฐานในการสร้าง Credit Scorecard ที่มีความสามารถในการตีความผลลัพธ์ได้ง่ายและให้ผลลัพธ์ที่โปร่งใส ในส่วนนี้ เราจะนำเสนอขั้นตอนการสร้างแบบจำลองโดยใช้ตัวแปรที่ผ่านการแปลงค่า WoE และวิเคราะห์ผลลัพธ์ที่ได้จากแบบจำลอง

การสร้างแบบจำลองและผลลัพธ์

แบบจำลองถูกสร้างขึ้นโดยใช้ตัวแปร WoE ที่มีนัยสำคัญ 5 ตัว ได้แก่ Home_Ownership_woe, Income_woe, Emp_Length_woe, Loan_Amnt_woe, และ Loan_Percent_Income_woe ผลลัพธ์ของแบบจำลองสรุปได้ดังตารางต่อไปนี้

	coef	std err	z	P> z
--	------	---------	---	------

	coef	std err	z	P> z
const	-1.3022	0.018	-73.725	0.000
Home_Ownership_woe	-0.8839	0.029	-30.660	0.000
Income_woe	-0.8017	0.031	-25.537	0.000
Emp_Length_woe	-0.3238	0.072	-4.508	0.000
Loan_Amnt_woe	-1.2027	0.082	-14.690	0.000
Loan_Percent_Income_woe	-0.7270	0.023	-31.427	0.000

การตีความสัมประสิทธิ์ของแบบจำลอง (Coefficient Interpretation)

ค่าสัมประสิทธิ์ (coef) ของตัวแปร WoE ทั้งหมดมีค่าเป็น ลบ ซึ่งเป็นผลลัพธ์ที่คาดหวังและถูกต้องตามหลักการเสมอ เนื่องจากการคำนวณ WoE ของเรากำหนดให้ค่า WoE ที่สูงขึ้นหมายถึงความเสี่ยงที่ต่ำลง (Good > Bad) ดังนั้นในแบบจำลอง Logistic Regression ที่ทำนายการผิดนัดชำระหนี้ (Default = 1) ค่าสัมประสิทธิ์จึงต้องเป็นลบเพื่อลดความน่าจะเป็นของ Default ลง เราสามารถจัดลำดับความสำคัญของปัจจัยที่ขับเคลื่อนความเสี่ยงได้จากขนาดของค่าสัมประสิทธิ์ (โดยไม่คำนึงถึงเครื่องหมายลบ) ดังนี้:

- 1. Loan_Amnt_woe (-1.2027): เป็นปัจจัยที่มีอิทธิพลต่อการพยากรณ์ความเสี่ยงสูงสุดในแบบจำลองนี้ ขนาดของวงเงินสินเชื่อที่จัดกลุ่มแล้วเป็นตัวขับเคลื่อนที่สำคัญที่สุด
- 2. Home_Ownership_woe (-0.8839): สถานะการครอบครองที่อยู่อาศัยเป็นปัจจัยที่มีอิทธิพลสูงเป็นอันดับสอง
- 3. Income_woe (-0.8017): ระดับรายได้เป็นปัจจัยที่มีอิทธิพลสำคัญในอันดับที่สาม
- 4. Loan_Percent_Income_woe (-0.7270): สัดส่วนภาระหนี้ต่อรายได้มีอิทธิพลรองลงมา
- 5. Emp_Length_woe (-0.3238): อายุงานเป็นปัจจัยที่มีอิทธิพลน้อยที่สุดในกลุ่มนี้ แต่ยังคงมีนัยสำคัญทางสถิติ (ค่า $P>|z| < 0.05$) และสมควรถูกรวมไว้ในแบบจำลอง

เมื่อได้แบบจำลองที่ผ่านการตรวจสอบนัยสำคัญทางสถิติและสามารถตีความผลได้แล้ว ขั้นตอนต่อไปคือการประเมินประสิทธิภาพของแบบจำลองในการจำแนกกลุ่มลูกค้า

5. การประเมินประสิทธิภาพของแบบจำลอง (Model Performance Evaluation)

การประเมินประสิทธิภาพของแบบจำลองเป็นขั้นตอนที่ขาดไม่ได้ เพื่อให้มั่นใจว่าแบบจำลองที่สร้างขึ้นมีความสามารถในการจำแนกกลุ่มลูกค้าที่มีความเสี่ยง (Default)

ออกจากกลุ่มลูกค้าที่ดี (Non-default) ได้อย่างแม่นยำ โดยเราจะใช้เมตริกมาตรฐานในอุตสาหกรรม Credit Scoring คือ AUC (Area Under the Curve) และ Gini Coefficient

การวัดผลการจำแนก

From the model evaluation on the test dataset:

- AUC Score: 0.7719
- Gini Coefficient: 0.5437

ค่า Gini Coefficient คำนวณมาจาก $2 * AUC - 1$ ซึ่งเป็นตัวชี้วัดผลการจำแนกที่นิยมใช้กันอย่างแพร่หลาย โดยค่า Gini ที่ 0.5437 (หรือ 54.37%) สามารถจัดอยู่ในเกณฑ์ "ยอดเยี่ยม (Excellent)" สำหรับแบบจำลอง Credit Scoring

ซึ่งแสดงให้เห็นว่าแบบจำลองมีความสามารถในการแยกแยะกลุ่มลูกค้าดีและลูกค้าเสียออกจากกันได้เป็นอย่างดี

การแสดงผลด้วยเส้นโค้ง ROC (ROC Curve)

ประสิทธิภาพของแบบจำลองยังสามารถแสดงผลในรูปแบบของกราฟเส้นโค้ง ROC (Receiver Operating Characteristic) ซึ่งจะพลอตค่า True Positive Rate เทียบกับ False Positive Rate ที่จุดตัด (Threshold) ต่างๆ เส้นโค้งที่เข้าใกล้แกน Y ด้านซ้ายบนมากเท่าไร ยิ่งแสดงถึงประสิทธิภาพในการจำแนกที่ดีขึ้นเท่านั้น และค่า AUC ก็คือพื้นที่ใต้เส้นโค้งนี้นั่นเอง

เมื่อแบบจำลองได้รับการพิสูจน์แล้วว่ามีความน่าเชื่อถือและมีประสิทธิภาพในการจำแนกที่ดีเยี่ยม ก็พร้อมที่จะถูกนำไปพัฒนาต่อให้เป็นเครื่องมือที่สามารถใช้งานได้จริงในรูปแบบของ Credit Scorecard

6. การพัฒนาเป็น Credit Scorecard

ขั้นตอนสุดท้ายของกระบวนการคือการแปลงผลลัพธ์ทางสถิติที่ซับซ้อนจากแบบจำลอง Logistic Regression ให้กลายเป็นระบบคะแนน (Points-based system)

ที่บุคลากรหน้างานสามารถนำไปใช้ในการตัดสินใจได้อย่างง่ายดายและรวดเร็ว Scorecard

นี้จะให้คะแนนสำหรับคุณลักษณะแต่ละอย่างของผู้ขอสินเชื่อ

และคะแนนรวมสุดท้ายจะถูกใช้เป็นตัวชี้วัดระดับความเสี่ยง

หลักการแปลงแบบจำลองเป็นคะแนน

การแปลงแบบจำลองเป็นคะแนนจะอิงตามหลักการของการกำหนดคะแนนเป้าหมาย (Target Score)

อัตราต่อรองเป้าหมาย (Target Odds) และจำนวนคะแนนที่ทำให้อัตราต่อรองเพิ่มขึ้นเป็นสองเท่า (PDO - Points to Double the Odds) ซึ่งในการพัฒนานี้ เราได้ใช้ค่าพารามิเตอร์ดังนี้:

- Factor: 28.85
- Offset: 487.12

จากค่าพารามิเตอร์เหล่านี้ เราสามารถคำนวณคะแนนพื้นฐาน (Base Score) ซึ่งเป็นคะแนนเริ่มต้นสำหรับผู้ขอสินเชื่อทุกคนก่อนที่จะพิจารณาคุณลักษณะอื่นๆ ได้เท่ากับ 525 คะแนน

ตาราง Credit Scorecard ฉบับสมบูรณ์

ตารางด้านล่างคือ Scorecard ที่พัฒนาเสร็จสมบูรณ์ ซึ่งแสดงคะแนนสำหรับคุณลักษณะแต่ละกลุ่ม

Feature	Bin (กลุ่ม/คุณลักษณะ)	Points (คะแนน)
Base Score		525
Home_Ownership	RENT	-13
Home_Ownership	MORTGAGE	17
Home_Ownership	OWN	32
Home_Ownership	OTHER	-12
Income	(3999.0, 35000.0]	-23
Income	(35000.0, 49000.0]	-2
Income	(49000.0, 63000.0]	5
Income	(63000.0, 86000.0]	11
Income	(86000.0, 600000.0]	24
Emp_Length	<2	-3
Emp_Length	2-4	-1
Emp_Length	4-6	1
Emp_Length	6-8	2
Emp_Length	8-10	3
Emp_Length	10-40	3
Emp_Length	>40	-5
Emp_Length	Missing	-5
Loan_Amnt	(499.0, 4400.0]	2
Loan_Amnt	(4400.0, 6750.0]	10
Loan_Amnt	(6750.0, 10000.0]	6

Feature	Bin (กลุ่ม/คุณลักษณะ)	Points (คะแนน)
Loan_Amnt	(10000.0, 14500.0]	0
Loan_Amnt	(14500.0, 35000.0]	-16
Loan_Percent_Income	(-0.01, 0.08]	16
Loan_Percent_Income	(0.08, 0.12]	14
Loan_Percent_Income	(0.12, 0.18]	9
Loan_Percent_Income	(0.18, 0.25]	3
Loan_Percent_Income	(0.25, 0.83]	-29

วิธีการใช้งาน: ผู้ขอสินเชื่อทุกคนจะเริ่มต้นด้วยคะแนนพื้นฐาน 525 คะแนน จากนั้นนำคะแนนจากคุณลักษณะต่างๆ ที่ตรงกับโปรไฟล์ของตนเองมาบวกรวมกัน ตัวอย่างเช่น หากผู้ขอสินเชื่อมีสถานะเป็นเจ้าของบ้าน (OWN) จะได้รับคะแนนบวกเพิ่ม 32 คะแนน แต่ถ้ามีรายได้อยู่ในช่วง (3999.0, 35000.0] จะถูกหักคะแนน -23 คะแนน เป็นต้น คะแนนรวมสุดท้ายคือ Credit Score ที่ใช้ในการประเมินความเสี่ยง

7. บทสรุป (Conclusion)

เอกสารฉบับนี้ได้นำเสนอภาพรวมของกระบวนการสร้าง Credit Scorecard ตั้งแต่การเตรียมข้อมูล การใช้เทคนิค Weight of Evidence (WoE) เพื่อสกัดพลังการพยากรณ์และแปลงตัวแปรให้อยู่ในรูปแบบที่เหมาะสม การสร้างแบบจำลอง Logistic Regression ที่สามารถตีความผลได้อย่างโปร่งใส ไปจนถึงการแปลงผลลัพธ์ทางสถิติให้กลายเป็น Scorecard ที่สามารถนำไปใช้ประกอบการตัดสินใจในทางปฏิบัติได้อย่างมีประสิทธิภาพ

แนวทางที่นำเสนอนี้ช่วยสร้างแบบจำลองที่มีความโปร่งใสสูง (Transparent) ซึ่งง่ายต่อการอธิบายและตรวจสอบ อีกทั้งยังแสดงให้เห็นถึงประสิทธิภาพในการจำแนกที่ยอดเยี่ยม (Gini = 0.5437) ซึ่งเป็นที่ยอมรับในมาตรฐานอุตสาหกรรม

กระบวนการทั้งหมดนี้จึงถือเป็นกรอบการทำงานที่แข็งแกร่งและน่าเชื่อถือสำหรับการพัฒนาระบบประเมินความเสี่ยงสินเชื่อในองค์กร เพื่อสนับสนุนการตัดสินใจที่ขับเคลื่อนด้วยข้อมูลและลดความเสี่ยงที่อาจเกิดขึ้นในอนาคต