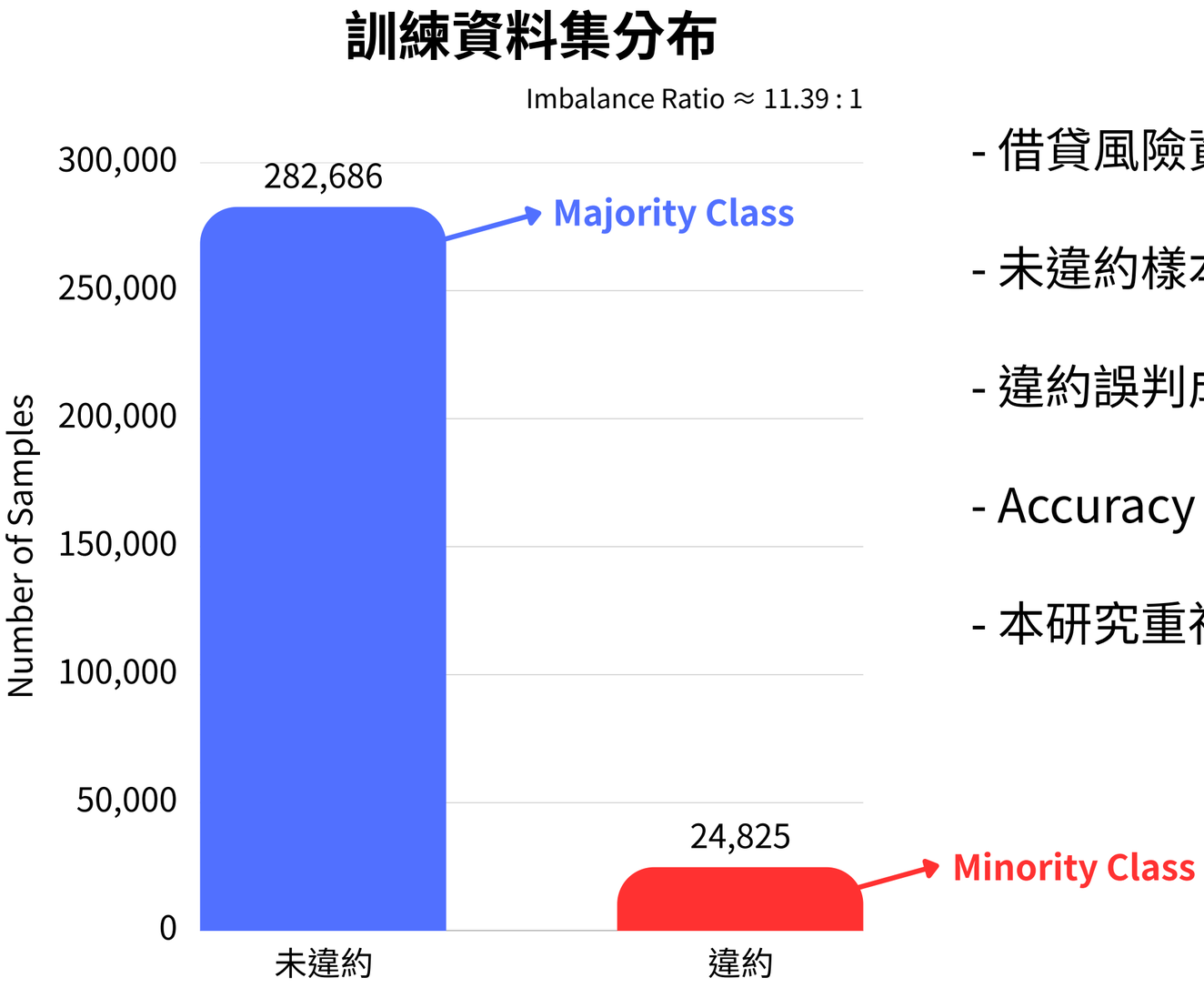
不平衡信用風險預測之建模優化與可解釋性分析

Data-centric Modeling and Explainability Analysis for Imbalanced Credit Risk Prediction

指導教授：蔣榮先 教授

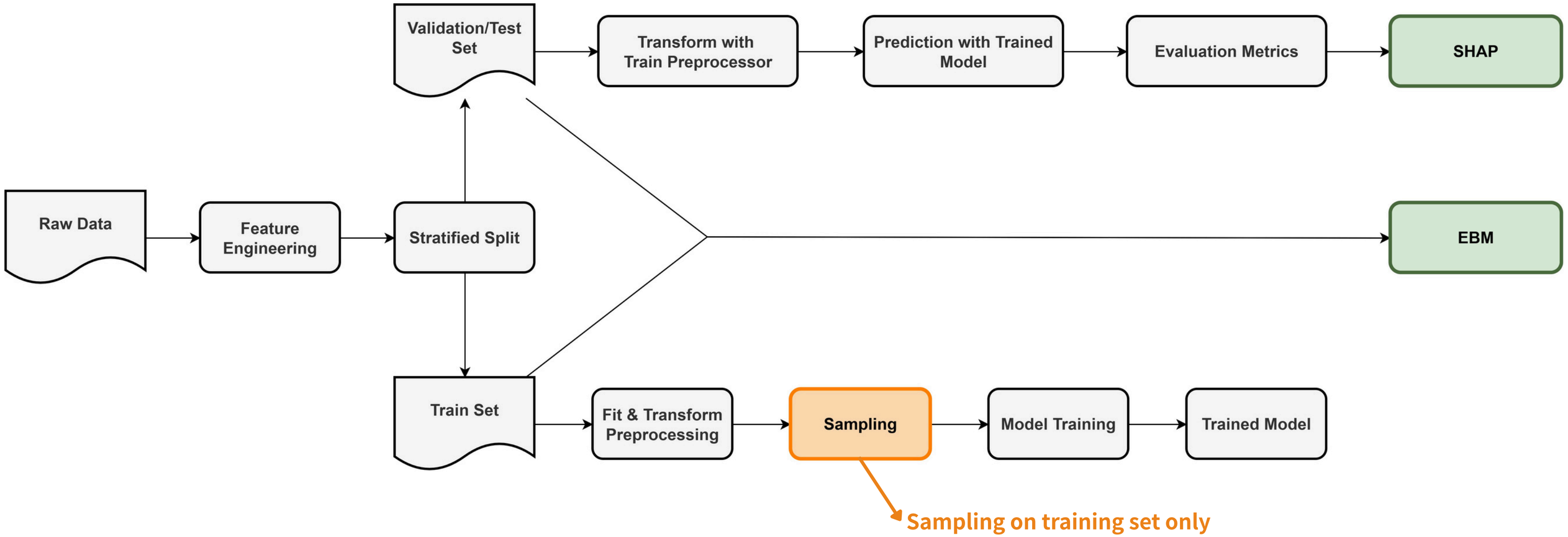
蔡源慶 資訊116 F74126165

研究動機：資料不平衡的影響



- 借貸風險資料中，違約樣本通常較少
- 未違約樣本數約占 91.93%，違約樣本數僅約 8.07%
- 違約誤判成本較高
- Accuracy 可能高估模型表現
- 本研究重視違約樣本的辨識能力

實驗流程圖


Validation/Test Set
Transform with Train Preprocessor
Prediction with Trained Model
Evaluation Metrics
SHAP
Raw Data
Feature Engineering
Stratified Split
EBM
Train Set
Fit & Transform Preprocessing
Sampling
Model Training
Trained Model
Sampling on training set only

方法與實驗設計：三個實驗設計

1 Model Selection

Models

- Logistic Regression
- Random Forest
- XGBoost
- LightGBM
- EBM

2 Imbalance Handling

Methods

- Cost-sensitive learning
- Over-sampling
- Under-sampling
- Hybrid sampling
- Threshold adjustment

3 Explainability Analysis

Methods

- Feature importance
- SHAP analysis
- EBM global explanation

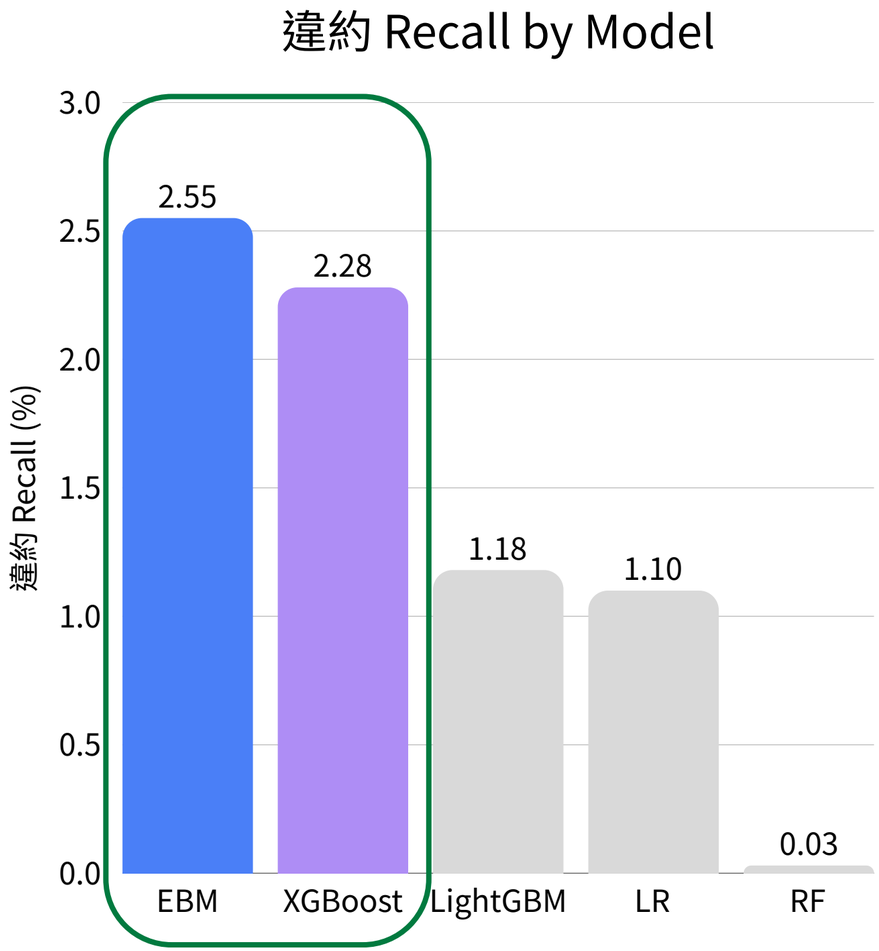
> 本研究先比較模型基準表現，再改善不平衡問題，最後分析模型判斷依據。

實驗一：Baseline Models 整體比較

Task: Binary Classification
Models: LR, RF, XGBoost, LightGBM, EBM
Metrics: Accuracy, 違約 Recall, PR-AUC, FN
Threshold: 0.5

Model	Accuracy	Precision	Recall	F1-score	PR-AUC	FN	FP
Logistic Regression	91.95%	58.57%	1.10%	2.16%	23.63%	3683	29
Random Forest	91.93%	**100.00%**	0.03%	0.05%	23.45%	3723	**0**
XGBoost	**92.01%**	63.91%	2.28%	4.41%	24.95%	3639	48
LightGBM	91.98%	70.97%	1.18%	2.32%	24.53%	3680	18
EBM	91.98%	57.23%	**2.55%**	**4.88%**	**25.16%**	**3629**	71

實驗一：模型傾向預測多數樣本



- 各模型 Accuracy 皆接近 92%
- 違約 Recall 皆低於 3%
- 多數高風險樣本仍被預測為 未違約
- 仍需加入不平衡處理與 threshold adjustment

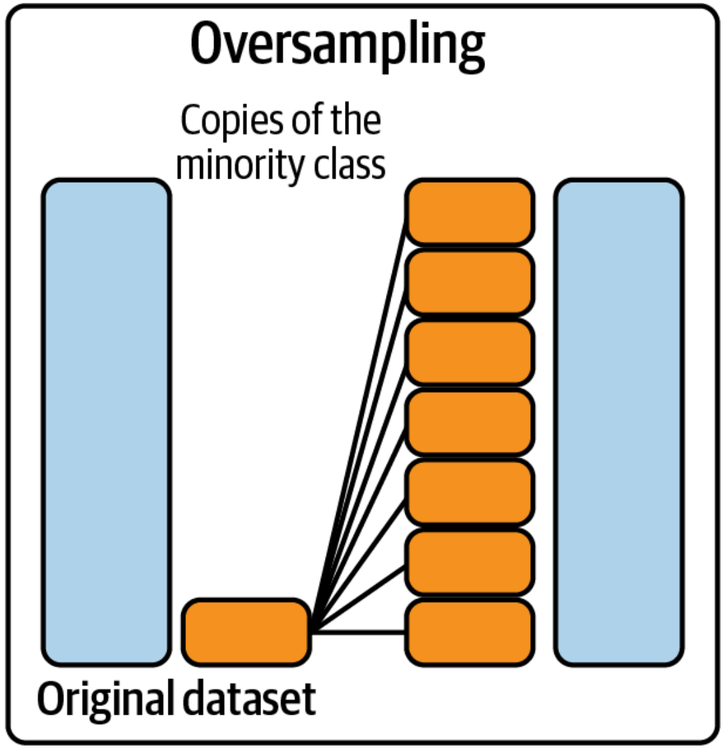
> 後續以 XGBoost 作為主要預測模型，並保留 EBM 作為可解釋性對照模型。

實驗二：以不同策略改善少數類辨識能力

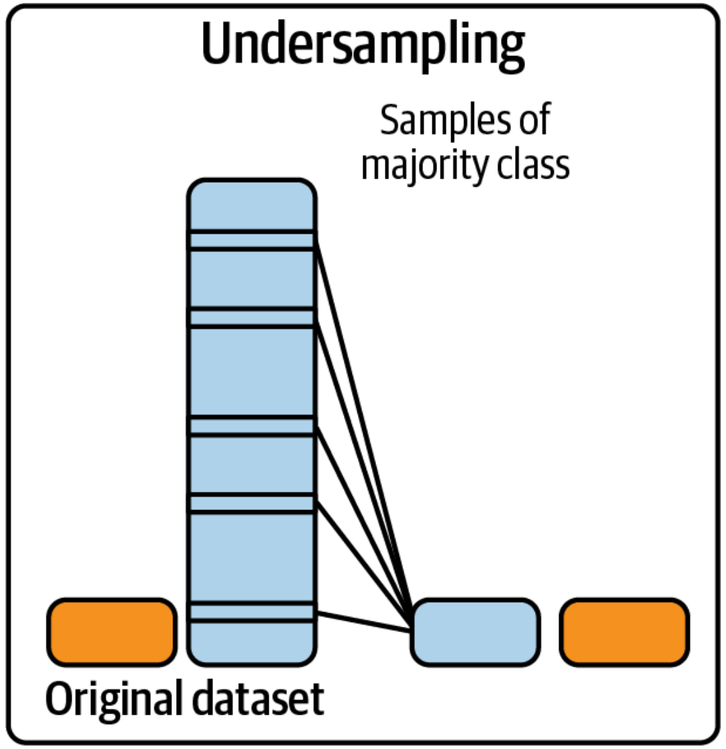
1. Cost-sensitive Learning

- XGBoost :Scale-pos-weight
- FN:higher cost

2. Over-sampling



3. Under-sampling



4. Threshold Adjustment

- Change the classification threshold
- Trade-off: Recall ↑, Precision ↓

實驗二：整理表格

	Method	Recall	Precision	F1-score	PR-AUC	FN	FP	Added FP / Reduced FN
	Baseline	1.80%	**63.60%**	3.50%	24.70%	3656	**39**	–
C	Scale pos weight	60.30%	18.20%	28.00%	24.70%	1477	10071	4.60
O	SMOTE	1.60%	58.70%	3.20%	23.70%	3663	43	N/A
O	RandomOverSampler	67.20%	16.60%	26.60%	**24.80%**	1223	12546	5.14
U	EditedNearestNeighbours	5.70%	49.70%	10.30%	24.70%	3509	216	**1.20**
U	RandomUnderSampler	**67.90%**	16.30%	26.30%	24.60%	**1194**	12981	5.26

> Added FP / Reduced FN → 每減少一個違約誤判增加的正常還款誤判

實驗二：Threshold adjustment 是可行的 FN 降低策略

	Method	Threshold	Accuracy	Precision	Recall	F1-score	PR-AUC	FN	FP	Added FP / Reduced FN
	Baseline	0.500	**91.90%**	**63.60%**	1.80%	3.50%	**24.70%**	3656	**39**	–
U	EditedNearestNeighbours	0.210	85.70%	25.20%	39.50%	**30.80%**	24.50%	2253	4365	**3.08**
U	AllKNN	0.222	85.20%	24.70%	40.50%	30.70%	24.50%	2215	4611	3.17
	Baseline Best F1	0.151	84.80%	24.30%	41.60%	30.70%	**24.70%**	2173	4831	3.23
C	Scale pos weight	0.610	84.10%	23.60%	**43.70%**	30.70%	**24.70%**	2097	5258	3.35
U	TomekLinks	0.150	84.00%	23.60%	**43.70%**	30.60%	24.60%	**2095**	5286	3.36

Threshold adjustment 讓每個方法中降低一個FN所需增加的FP的數值都有明顯下降

實驗二：Threshold adjustment 是較有效的決策策略

結論策略

主要模型

XGBoost

不平衡處理方式

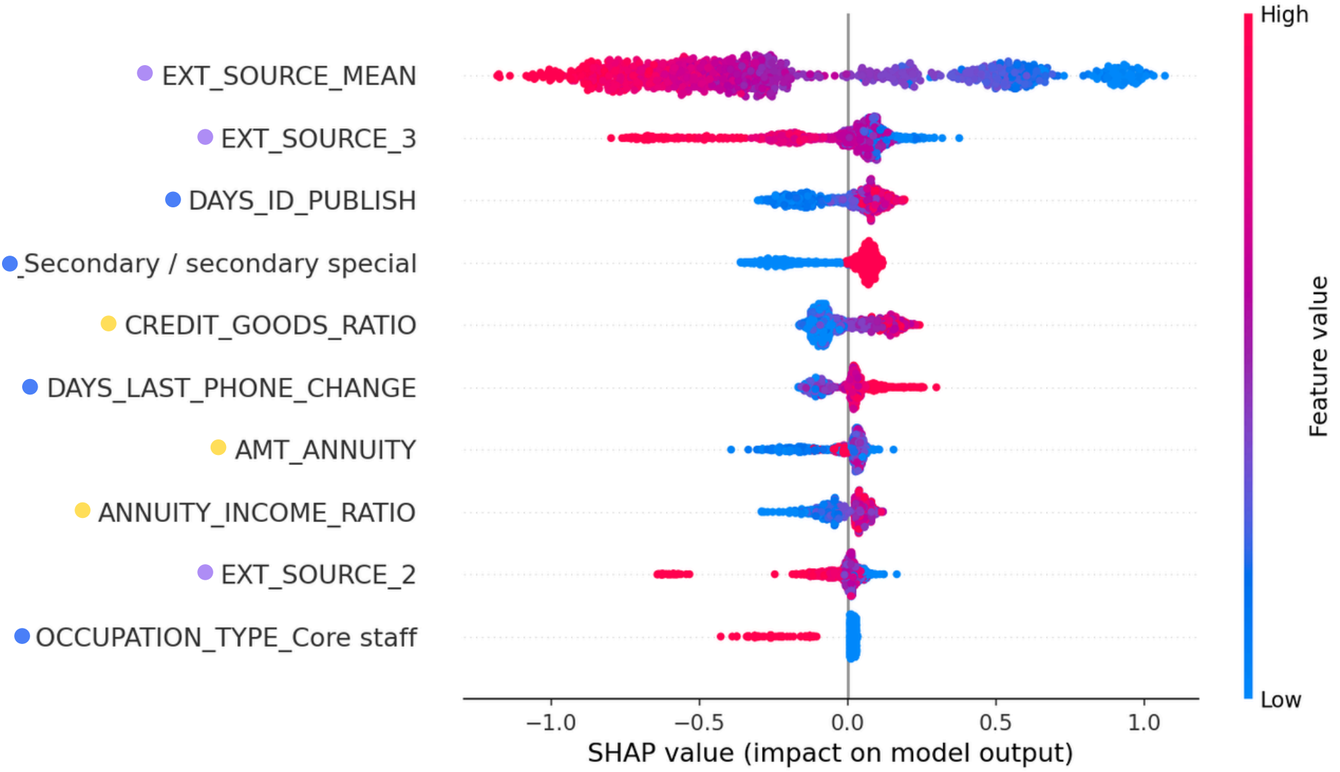
主要：Threshold adjustment
次要：Sampling

原因

Sampling能提升Recall，但FP大幅提升
Threshold Adjustment在實驗的不同方法下都能有效平衡FN和FP

實驗三：外部信用分數是主要影響特徵

Red = high value, Blue = low value
Positive SHAP value → higher default-risk prediction

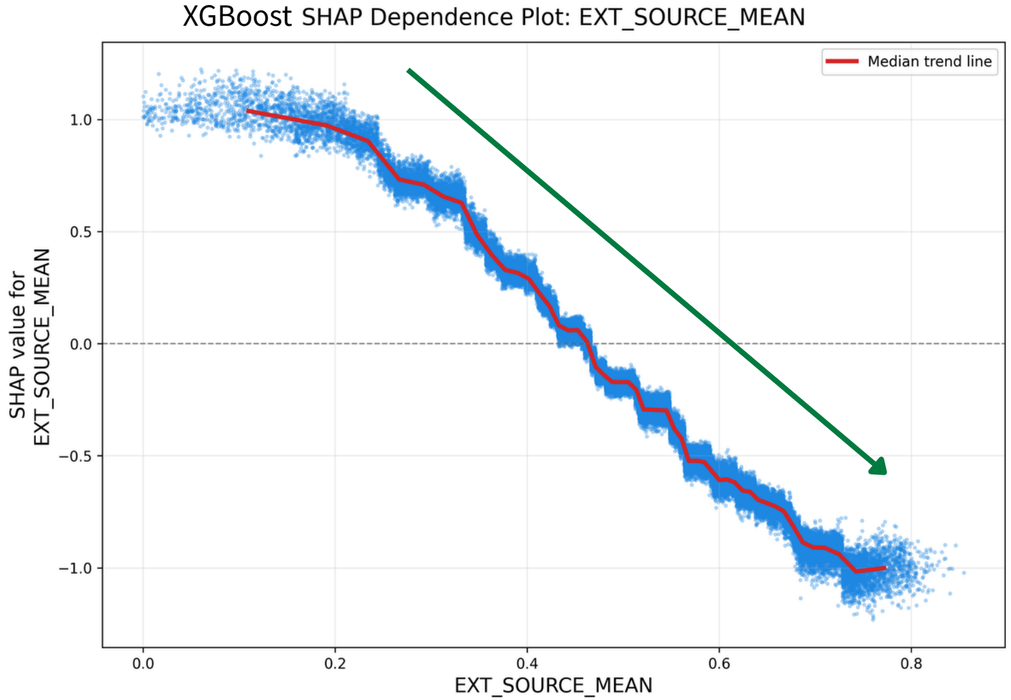


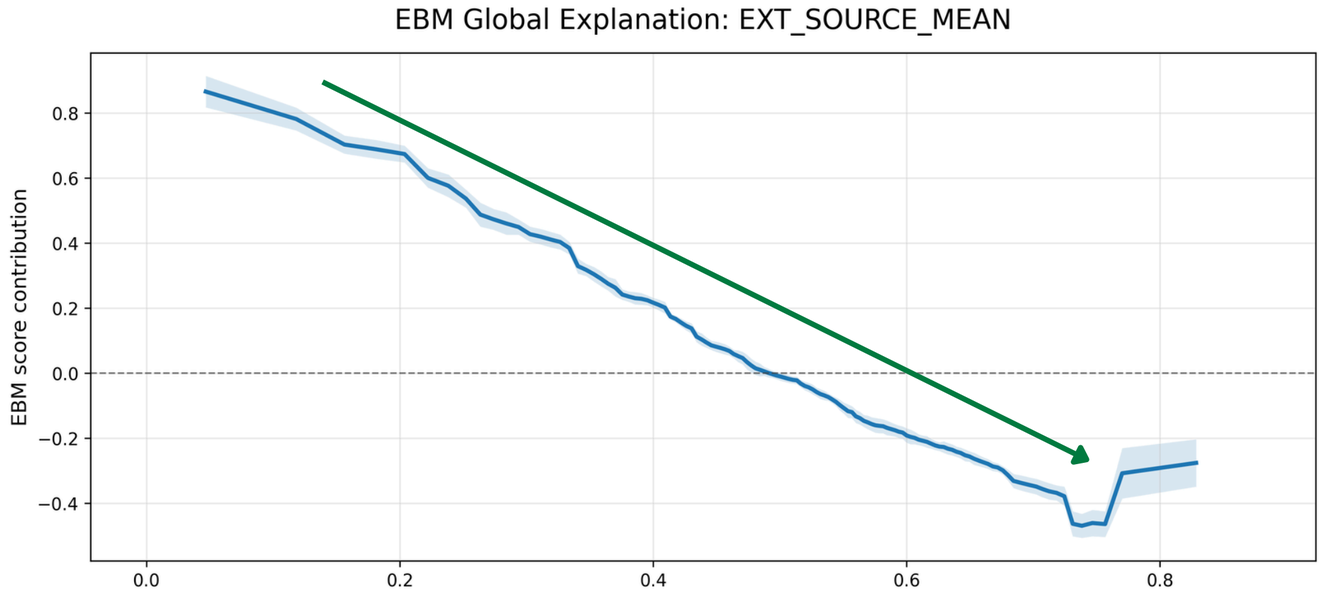
- **1. EXT_SOURCE**
  外部資料來源對申請人信用狀況的綜合評估。
- **2. Repayment burden**
  包含貸款金額、年金支出、收入比例等。
- **3. Applicant profile / stability**
  包含年齡、就業年資、家庭狀況等。

SHAP 顯示模型主要依賴外部信用分數，並以還款負擔與申請人背景訊號輔助風險判斷。

實驗三：兩種可解釋方法呈現一致風險趨勢

以 EXT_SOURCE_MEAN 為例，觀察外部信用分數高低如何影響模型預測的違約風險





信用分數越高→影響越低

可解釋性分析不只指出重要特徵，也能觀察兩個模型風險判斷分別對特徵是否有一致的影響方向

結論

本實驗最後以XGBoost為主要模型，Threshold adjustment作為主要不平衡處理方法，並以SHAP與EBM進行可解釋分析

1. 準確率不等於風險辨識能力
高Accuracy可能只反映模型偏向預測多數類，無法真正代表模型能有效找出違約風險

2. Threshold 是有效的風控決策
調整分類門檻能直接控制 FN / FP trade-off，是本研究中較有效的不平衡處理策略。

3. 可解釋性提升模型可信度
SHAP 與 EBM 皆指出相似風險訊號，讓模型判斷依據更可視化、可檢驗。

未來展望

1. 成本導向決策
結合實際 FN / FP 成本，建立更符合風控需求的決策門檻。

2. 提升模型預測能力
在基礎模型上加入更多特徵工程、參數調整與模型校準，提升整體預測能力，再進一步進行風控決策。

3. 公平性與穩定性驗證
檢查背景特徵是否造成偏誤，並測試模型在不同資料切分下是否穩定。

THANK YOU