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

指導教授：蔣榮先 教授　專題組員：蔡源慶

一、研究動機與目的

金融借貸資料常存在明顯的類別不平衡問題，**逾期或還款困難樣本通常遠少於正常還款樣本**。以本專題的資料為例，違約的高風險樣本僅約 8.07%，使得模型容易**傾向預測多數類**。若僅以 Accuracy 作為主要評估指標，即使模型大量預測為正常還款，**也可能取得較高的準確率，卻無法有效辨識真正需要被風控關注的高風險客戶**。

因此，本研究不將信用風險預測視為單純的二元分類任務，而是進一步**關注少數類辨識能力**、False Negative 降低、False Positive 成本，以及模型判斷依據是否能被解釋。**本研究目標包括：比較不同模型在不平衡資料下的基準表現、分析不平衡處理與 threshold tuning 對 FN / FP 成本的影響，並透過 SHAP 與 EBM 檢視模型主要依據哪些風險訊號進行判斷**。

二、研究問題

本研究圍繞三個核心問題進行分析：

Q1：在資料不平衡的信用風險資料中，高 Accuracy 是否能代表模型具備高風險客戶辨識能力？

Q2：不同不平衡處理方法如何影響違約的 Recall、False Negative 與 False Positive 成本？

Q3：模型主要依據哪些特徵進行風險判斷？這些特徵是否具有可解釋的金融意義？

三、實驗設計

本研究以固定資料前處理流程為基礎，設計三個實驗，從模型比較、不平衡處理到可解釋性分析，從不同角度切入模型訓練的過程。

實驗一：Model Selection

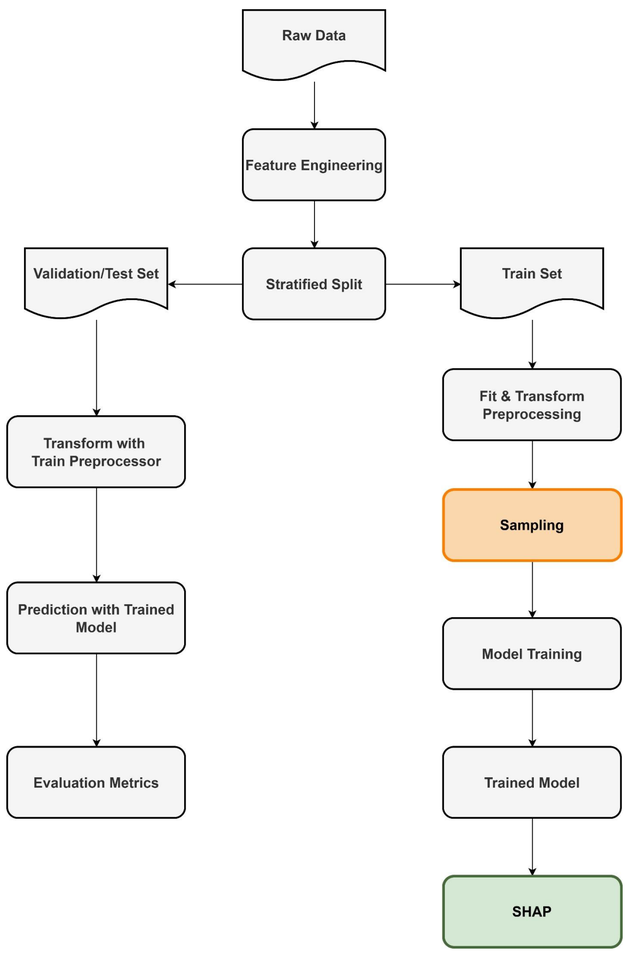
比較 Logistic Regression、Random Forest、XGBoost、LightGBM、EBM 在**原始不平衡資料下的 baseline 表現**。

實驗二： Imbalance Handling

測試 resampling、cost-sensitive learning 與 threshold adjustment，**分析 Recall 提升與 FN / FP 成本取捨**。

實驗三： Explainability Analysis

透過 SHAP 檢視模型主要風險特徵，**觀察模型判斷依據是否具有金融合理性**。



圖一：訓練流程圖

四、實驗結果

實驗一：

比較不同模型的 baseline 表現。結果顯示，各模型 Accuracy 皆接近 92%，但 違約的 Recall 皆低於 3%，代表**模型雖然整體準確率高，卻幾乎無法有效辨識真正高風險客戶**。此結果說明，在不平衡信用風險資料中，Accuracy 容易被多數類樣本主導，後續會以 XGBoost 為主，且評估會更重視違約的 Recall、PR-AUC 與 False Negative。

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

表一：模型 Baseline 比較

實驗二：

本實驗測試不同不平衡處理方法，分析不同方法對違約的 Recall、FN 與 FP有什麼影響。結果顯示，**這些方法能提高模型對高風險樣本的敏感度**，但也會使 FP 增加，**使模型在降低漏判風險的同時，必須承擔較高的誤判成本**。

以 Original model 為例，調整至 best F1 threshold 後，Recall 由 0.018 提升至 0.416，FN 由 3656 降至 2173，但 FP 由 39 增加至 4831。此結果說明，**threshold adjustment雖然不會提升模型能力**，但在能明顯降低 FN 的方法中，Threshold adjustment每降低一個 FN 所需增加的 FP 較少。

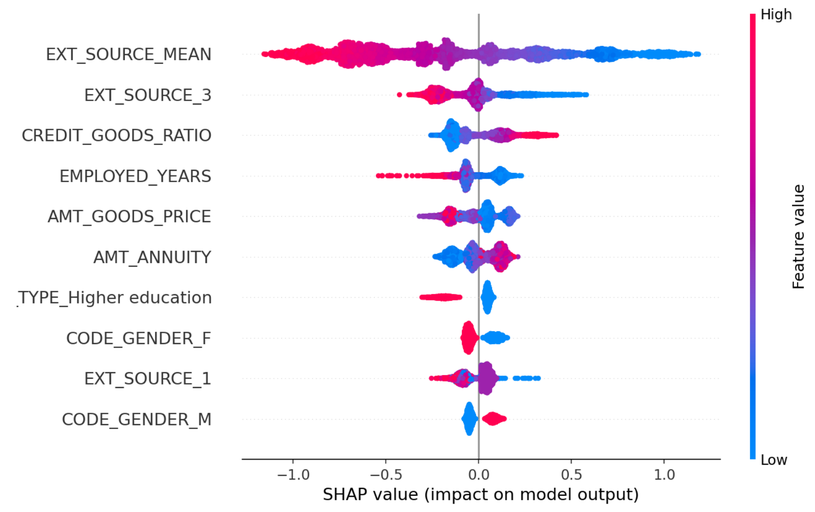
Method	Threshold	Recall	Precision	F1-score	PR-AUC	FN	FP
Original	0.500	1.80%	**63.60%**	3.50%	24.70%	3656	**39**
Original Best F1	0.151	41.60%	24.30%	**30.70%**	24.70%	2173	4831
Scale pos weight	0.500	60.30%	18.20%	28.00%	24.70%	1477	10071
EditedNearestNeighbours	0.500	5.70%	49.70%	10.30%	24.70%	3509	216
SMOTE	0.500	1.60%	58.70%	3.20%	23.70%	3663	43
RandomOverSampler	0.500	67.20%	16.60%	26.60%	**24.80%**	1223	12546
RandomUnderSampler	0.500	**67.90%**	16.30%	26.30%	24.60%	**1194**	12981

表二：不平衡處理比較

實驗三：

本實驗使用 SHAP 分析模型判斷依據，**觀察模型在預測違約風險時主要依賴哪些特徵**。EXT_SOURCE 是外部資料來源對申請人信用風險的評估分數，從結果來看，**模型在預測違約風險時高度依賴這類外部信用評分特徵**；此外，**還款金額**、**貸款金額比例**與**申請人資料穩定性**等特徵，也會影響模型對風險的判斷。

模型不僅能輸出預測結果，也能透過 SHAP 可解釋性分析檢視其判斷依據。藉由**觀察模型主要依賴的特徵與其影響方向**，可以**使模型結果更容易被理解**，並協助判斷模型是否符合金融風控上的合理性。



表三：SHAP影響力分布圖

五、結論

本研究最終選擇 XGBoost 作為主要預測模型，並以 threshold adjustment 作為不平衡處理方法。結果顯示，此方法能依照風控需求調整 Recall、FN 與 FP 的取捨；同時，SHAP 可協助檢視模型主要風險特徵，使預測結果更具可解釋性。

整體而言，信用風險預測不只是分類問題，而是少數類辨識、誤判成本與可解釋性的綜合風控決策問題。未來可加入實際 FN / FP cost matrix、probability calibration，以及公平性與穩定性分析，使模型更接近實務風控需求。

參考文獻

[1] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," JRSS Series A, 1997.

[2] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, 2012.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," KDD, 2016.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," NeurIPS, 2017.