

<https://doi.org/10.31891/2307-5732-2026-365-42>

УДК 004.8

СИРОВЕТНИК БОГДАН

Національний університет "Львівська політехніка"

<https://orcid.org/0009-0003-7246-0727>

e-mail: bohdan.y.syrovetnyk@lpnu.ua

КІСЬ ЯРОСЛАВ

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0003-3421-2725>

e-mail: yaroslav.p.kis@lpnu.ua

ГЛИБИННЕ НАВЧАННЯ В ОЦІНЦІ РИЗИКІВ ДЕФОЛТУ: СУЧАСНІ ПІДХОДИ ТА ІНСТРУМЕНТИ

Стаття узагальнює сучасні підходи застосування глибинного навчання для оцінки ризику дефолту позичальників і портфельів. Розглядаються архітектури багатосарових перцептронів, рекурентних та часово-згорткових мереж для табличних і послідовних фінансових даних, а також трансформери для багатовимірних часових рядів і кредитних графів. Порівнюються метрики PD-моделей (ROC-AUC, PR-AUC, KS, Brier score, калібрування) з лінійними та деревоподібними базами. Особливу увагу приділяється якості даних, боротьбі з дисбалансом класів (SMOTE, фокальна втрата, порогова оптимізація), часово-серійному валідуванню й стабільності в умовах дрейфу даних. Обговорюються інтерпретованість і комплаєнс глобальні та локальні пояснення (SHAP, LIME), монотонні обмеження, стабілізація ознак, документація моделей і вимоги управління модельним ризиком. Показано інтеграцію DL-моделей у кредитні пайплайни та MLOps (моніторинг, ретрейнінг, тестування на справедливості), а також використання комбінованих підходів (stacking/ensembles) і графових неймереж для виявлення шахрайства та зв'язностей. Наведено практичні рекомендації щодо вибору архітектури, налаштування гіперпараметрів, калібрування й оцінки економічного ефекту (cut-off, очікуваний збиток, ризик-апетит), а також обмеження вартість розгортання, вимоги до обчислень, ризику перенавчання та етичні аспекти.

Ключові слова: глибинне навчання ризик дефолту PD-модель часові ряди трансформери інтерпретованість (SHAP, LIME) дисбаланс класів калібрування MLOps управління модельним ризиком.

SYROVETNYK BOHDAN, KIS YAROSLAV

Lviv Polytechnic National University

APPLICATION OF DEEP LEARNING FOR CREDIT RISK PREDICTION IN THE BANKING SECTOR

This article synthesizes modern deep learning approaches to estimating default risk for borrowers and credit portfolios, with an emphasis on deploying probability of default (PD) models under real-world regulatory and business constraints. We review neural architectures that are commonly applied to tabular and sequential financial data, including multilayer perceptrons (MLP) for structured scoring features, recurrent networks (RNN/LSTM/GRU) and temporal convolutional networks (TCN/CNN) for payment and behavioral histories, and transformer-based models for multivariate time series as well as credit graphs. Model quality is discussed through the lens of both ranking and calibration: ROC-AUC, PR-AUC, KS statistic, Brier score, calibration slope, expected calibration error, and recall/precision in risk-based top segments, compared against linear baselines and tree-based ensembles.

A separate focus is placed on data quality and practical pitfalls that strongly influence empirical results: class imbalance, label sparsity, and temporal shifts. We outline robust countermeasures such as class weights, focal loss, synthetic minority over-sampling (SMOTE) when appropriate, and threshold optimization driven by an explicit loss function and risk appetite rather than "accuracy" alone. The paper emphasizes time-series aware validation, leakage prevention, and out-of-time backtesting to ensure stability under drift.

Given that credit decisioning requires transparency, we examine interpretability and compliance practices that enable the use of complex models in regulated environments. Global and local explanations with SHAP and LIME are considered alongside monotonic constraints, feature stabilization, and model documentation within model risk management (MRM). We also describe how deep learning models can be integrated into production credit pipelines and MLOps: monitoring of covariate and label drift, alerting rules, scheduled retraining, fairness testing, and reproducible evaluation. Finally, we discuss combined approaches (stacking/ensembles) and graph neural networks for capturing relational dependencies and detecting fraud or risk propagation in borrower networks. Practical recommendations are provided for selecting architectures, tuning hyperparameters, calibrating PD outputs, and translating metric gains into economic effect via cut-off policy, expected loss, and portfolio-level profitability, while acknowledging limitations such as deployment costs, compute requirements, overfitting risk, and ethical considerations.

Keywords: deep learning; credit risk; probability of default (PD); time series; transformers; graph neural networks; calibration; class imbalance; SHAP; LIME; MLOps; model risk management.

Стаття надійшла до редакції / Received 11.02.2026

Прийнята до друку / Accepted 11.03.2026

Опубліковано / Published 28.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Сироветник Богдан, Кис Ярослав

Вступ

Оцінка ризику дефолту є ключовим елементом кредитного циклу та системи управління ризиками фінансових установ, оскільки безпосередньо впливає на ціноутворення, ліміти, формування резервів, регуляторний капітал і ризик-апетит. Традиційні підходи логістична регресія та деревоподібні моделі довели ефективність у табличних даних, однак стикаються з обмеженнями щодо нелінійностей, високорозмірних ознак, взаємодій та концептуального дрейфу. Розвиток глибинного навчання (DL) відкрив можливості моделювати складні залежності в часових рядах, транзакційних логах і графових структурах зв'язків між позичальниками, що є критично важливим для точнішої оцінки PD (Probability of Default) та зменшення очікуваного кредитного збитку.

Попри зростання якості прогнозу, впровадження DL у кредитному скорингу залишається викликом через вимоги до інтерпретованості, стабільності, комплаєнсу та керування модельним ризиком. Регуляторні рамки (в частині валідації, калібрування, стабільності показників, відсутності дискримінації) вимагають не лише високої прогностичної спроможності, а й відтворюваності рішень і прозорої документації. Додатковими бар'єрами є дисбаланс класів, нестача якісних мічених даних, часові зсуви та вартість обчислень у продуктивних середовищах.

Мета статті систематизувати сучасні підходи та інструменти глибинного навчання для оцінки ризику дефолту та показати, як їх інтегрувати у виробничі пайплайни з урахуванням регуляторних і бізнес-вимог. Наукова новизна полягає в поєднанні огляду архітектур (MLP, CNN/TCN, RNN/LSTM/GRU, трансформери, графові неймережі) з практичними протоколами валідації на часових зрізах, методами роботи з дисбалансом (фокальна втрата, перепсемплінг), техніками інтерпретації (SHAP/LIME), калібруванням і моніторингом дрейфу.

Постановка проблеми

Для досягнення поставленої мети у дослідженні вирішуються такі завдання:

1. які архітектури DL дають найбільший приріст якості для різних типів даних (анкета, транзакції, поведінкові й альтернативні джерела);
2. як забезпечити інтерпретованість і відповідність політикам управління модельним ризиком без суттєвої втрати якості;
3. які протоколи валідації та метрики (ROC-AUC, PR-AUC, KS, Brier, калібрування) найкраще відображають економічну корисність моделі;
4. як організувати MLOps для стабільного продакшн-розгортання (моніторинг, ретрейнінг, алерти дрейфу, тестування на справедливість).

Результат дослідження

Дослідження з оцінки ризику дефолту історично спиралися на логістичну регресію та деревоподібні ансамблі, однак за останні роки з'явився стійкий тренд до використання глибинного навчання (DL) для табличних, часових та мережевих даних позичальників. Систематичні огляди 2015–2024 рр. фіксують зростання частки праць із DL-архітектурами, але також відзначають, що результати залежать від типу даних і дизайну валідації.

Для табличних кредитних даних запропоновано трансформерні підходи, які кодують категоріальні ознаки через контекстні ембеддинги (TabTransformer) і часто зрівнюються за якістю з передовими деревоподібними моделями, іноді перевершуючи їх на шумних або неповних вибірках.

Водночас емпіричні порівняння на класичних наборах щодо дефолтів карток нерідко демонструють перевагу XGBoost/LightGBM за AUC/ACC у базових налаштуваннях, що підкреслює важливість задоволеного вибору моделі.

Для урахування зв'язків між позичальниками інтенсивно розвиваються графові неймережі. Роботи на іпотечних та корпоративних даних показують, що динамічні багаточасові графові моделі з механізмами уваги (GAT + LSTM) покращують прогноз PD порівняно з традиційними індивідуальними ознаками, розкриваючи ефекти “розповсюдження ризику” через мережі. Також досліджуються мотив-збережні GNN та інші варіації для фінансових дефолтів.

Ключовим обмеженням DL у кредитному скорингу є вимоги інтерпретованості та комплаєнсу. Стандартні інструменти пояснюваності SHAP і споріднені XAI-методи стали де-факто підходом для глобальних і локальних пояснень важливості ознак, що дозволяє інтегрувати складні моделі у процес ухвалення рішень і валідацію.

Європейські регуляторні орієнтири (EBA Guidelines on loan origination and monitoring) наголошують на доказовості кредитоспроможності, стабільності моделей, відсутності дискримінації та належній документації, що визначає практичні вимоги до DL-рішень у банках.

Методологічно література підкреслює чутливість результатів до дисбалансу класів, часових зсувів і протоколів валідації. Поряд із класичним ресемплінгом активно використовуються фокальні функції втрат та синтетичні дані (включно з табличними генеративними моделями), однак переваги таких підходів є задоволенними й вимагають ретельного часо-серійного бек-тесту й калібрування PD.

Практичне значення роботи полягає у формулюванні покрокових рекомендацій з вибору архітектури та гіперпараметрів, побудови ознак, калібрування порогів під бізнес-цілі (cut-off під ризик-апетит), а також оцінки економічного ефекту через очікуваний збиток і капітальні вимоги. Стаття також висвітлює обмеження підходів DL ризику перенавчання, залежності від якості даних, витрати на інфраструктуру та окреслює напрями подальших досліджень, зокрема мультизадачне навчання PD/LGD/EAD, мультимодальні представлення даних і використання контрастивного навчання для рідкісних подій.

Оцінка ризику дефолту на сьгодні спирається на поєднання якісної підготовки даних, коректного часо-серійного валідування та змагальних моделей, де глибинне навчання конкурує з бустингами на деревах рішень. На табличних даних трансформерні архітектури (TabTransformer) за рахунок контекстних ембеддингів категоріальних ознак демонструють підвищення середнього AUC над іншими DL-підходами та наближаються до бенчмарків градієнтного бустингу у напівконтрольованому режимі зафіксовано додатковий середній приріст AUC на 2,1 відсоткового пункту завдяки препрейнгу ембеддингів, а також підвищену стійкість до пропусків та шуму у вхідних даних. Це важливо для кредитних анкет і транзакцій із “дірками” та нефіксованими словниками категорій, де класичні one-hot/target-encoding часто деградують під дрейфом ознак [1].

У графових сценаріях коли між позичальниками існують транзакційні, соціальні чи спільні атрибутивні зв'язки мотив-збережні GNN із курикулум-навчанням поліпшують точність і AUC порівняно як з деревами, так

і з базовими GNN на публічному наборі та двох промислових наборах модель MotifGNN випереджала XGBoost і загальні GNN (наприклад, при частці навчальних даних 60 % точність зростала до 0,865 проти 0,749 у XGBoost), а також піднімала пов'язані метрики якості ранжування дефолтів. Ці результати підкреслюють, що комбінування індивідуальних атрибутів і вищих порядків топології дає стабільний приріст у відсіканні ризикових вузлів.

Водночас регуляторна придатність залишається обов'язковою умовою керівні вказівки ЕВА щодо надання та моніторингу кредитів прямо вимагають доказовості кредитоспроможності, прозорих пояснень, стабільності моделей, недискримінаційних практик та належної документації протягом усього життєвого циклу моделі, що визначає спосіб валідації, калібрування, моніторингу дрейфу й підготовки адверс-екшен повідомлень.

Класичний публічний датасет “Default of Credit Card Clients (Taiwan)” із 30 000 рахунків (6636 дефолтів і 23 364 недефолтів дисбаланс $\approx 22,1\%/77,9\%$), що містить демографію, кредитний ліміт, історію білінгу/платежів і статуси погашення за 6 місяців. Цей набір є реальним банківським масивом і широко застосовується у валідаційних дослідженнях. За літературою, на ньому XGBoost/LightGBM типово досягають ROC-AUC у діапазоні близько 0,71–0,79 залежно від препроцесінгу, балансування та протоколу валідування окремі роботи звітують AUC $\approx 0,758\text{--}0,788$ з високою чутливістю (recall) під агресивні cut-off, що зручно для триажу найризикованіших заявок. На цьому ж датасеті й суміжних наборах DL-моделі інколи вирівнюються або перевершують бустинги, але виграш не є гарантованим без ретельної інженерії ознак та правильного часо-серійного бектесту. Ключовим доповненням є XAI стабільність ранжування ознак за SHAP на кредитних даних Тайваню підтверджена через високі значення коефіцієнта конкордації Кендалла ($W \approx 0,98$ $p < 0,001$), що важливо для комплаєнсу й адекватних пояснень причин [2]. Практична процедура побудови рішення для цього кейсу виглядає так:

1. ретельний data-audit (узгодження часових вікон, виявлення та корекція аномалій у bill_amt/pay_amt, кодування категорій із ембеддингами або target-encoding з анти-ліком у часовому розрізі)
2. стратифікована time-series cross-validation з блокуванням витоків (train на місяцях $t-6 \dots t-2$, validate на $t-1$, test на t) і додатковим out-of-time бектестом
3. боротьба з дисбалансом класів не ресемплінгом як самоціллю, а через порогову оптимізацію за бізнес-функціоналом втрат, фокальну втрату або клас-ваги
4. ансамблювання сильний табличний бустинг як бейзлайн + TabTransformer/MLP як другий стовп (stacking/soft-voting із калібруванням Platt/Isotonic)
5. повноцінна калібровка PD і перевірка стабільності (Brier score, калібр-криві, PSI/CSI)
6. XAI-пакет (глобальні та локальні SHAP-пояснення, стрес-тести пояснень)
7. MLOps-моніторинг дрейфу (covariate/label shift), деградації метрик і періодичний ретрейнінг [3].

Якщо вихідний XGBoost на тайванському наборі забезпечує ROC-AUC близько 0,75–0,78 і KS у прийнятній зоні 0,30–0,50, то керована заміна на ансамбль “GBDT + TabTransformer” за правильної валідації здатна дати додаткові 0,5–1,5 в. п. AUC на шумистих підвбірках або під час інкрементального оновлення категоріальних словників, утримуючи калібрування PD у нормативних межах. Для портфеля з часткою дефолтів 22 % і середнім очікуваним збитком на дефолт 1000 умовних одиниць зсув cut-off, який піднімає recall “поганих” на 5–8 в. п. при помірній втраті precision, дає щорічну економію у вигляді зниження EL та резервів навіть 1 в. п. приросту AUC у високих чутливостях часто транлюється в двозначні відсотки поліпшення capture gate у топ-децилях, що визначає профіт саме через відмову “поганим” і релімітування “сірих”. У графових продуктах мікрокредитування підключення MotifGNN підвищувало точність класифікації до $\approx 0,86$ проти $\approx 0,75$ у XGBoost за тих самих умов тренування, що суттєво збільшує виявлення кластерів ризику в соціальних/транзакційних мережах і знижує втрати від каскадних дефолтів.

Важливо, що всі поліпшення конвертуються у бізнес-ефект лише після належного калібрування й узгодження порогів із ризик-апетитом та політикою ціноутворення, а також при відповідності ЕВА-вимогам щодо прозорості та недискримінаційності.

Таблиця 1

Якість моделей на валідації (табличні кредитні дані)

Модель	ROC-AUC	PR-AUC	KS	Brier	Calib. slope	ECE	Recall	Precision	Latency
Логістична регресія	0.735	0.454	0.36	0.166	0.93	3.1%	38%	84%	0.1
LightGBM	0.772	0.502	0.44	0.158	0.98	2.3%	44%	87%	0.3
XGBoost	0.778	0.509	0.46	0.156	1.01	2.1%	45%	88%	0.8
TabTransformer	0.783	0.515	0.47	0.155	1.02	1.9%	46%	88%	2.5
Ensemble (GBDT+ TabTransformer)	0.792	0.527	0.49	0.153	1.00	1.7%	48%	89%	3.1

Таблиця 1. показує, що якість моделей за класичними скоринговими метриками зростає від лінійних базових підходів до ансамблевих і трансформерних архітектур ROC-AUC піднімається з 0,735 у логістичної регресії до 0,792 в ансамблі GBDT+TabTransformer, PR-AUC з 0,454 до 0,527, а статистика KS з 0,36 до 0,49, що означає помітно кращу здатність моделі розводити «добрих» і «поганих» позичальників у всьому діапазоні ймовірностей. Одночасно бачимо покращення калібрування Brier score зменшується до 0,153 в ансамблі, очікувана похибка калібрування ECE скорочується до 1,7%, а нахил калібрувальної прямої тримається біля 1,00,

тобто прогнозовані PD добре відповідають фактичним частотам дефолтів і підходять для розрахунку резервів і капіталу. Операційно важливо, що приріст якості не є «безкоштовним» середня затримка інференсу на CPU зростає від 0,1 мс у логістичній регресії до 3,1 мс в ансамбл, однак і ці значення залишаються прийнятними для онлайн-скорингу. Найбільш показова бізнес-метрика Recall@Top10\% зростає з 38% до 48%, тобто в одному й тому самому верхньому децилі ризику ансамбль знаходить майже на чверть більше потенційних дефолтів, ніж лінійна модель, при близькій Precision@Top10\% (89% проти 84–88%), що напрям впливає на якість відсікання збиткових заявок і економію очікуваних втрат. Кругова діаграма до Таблиці 1 підтверджує саме цей розклад внесків найбільший сектор припадає на ансамбль ($\approx 21,7\%$ «пирога» Recall@Top10\%), далі йдуть TabTransformer ($\approx 20,8\%$) і XGBoost ($\approx 20,4\%$), тоді як LightGBM і логістична регресія займають менші частки ($\approx 19,9\%$ і $\approx 17,2\%$), що наочно ілюструє перерозподіл «уловлених» поганих у топ-децилі на користь більш потужних методів [7].

Таблиця 2

Бізнес-ефект під різними cut-off (портфель 30 000 заявок)									
Стратегія	Cut-off	Рішення про видачу	Частка видач	Очікувані дефолти серед виданих	Очікуваний збиток EL, тис. у.о.	«Добрі» видані (шт.)	Маржа, тис. у.о.	Чистий рез., тис. у.о.	Δ до бази, тис. у.о.
База XGBoost	0.35	18 000	60%	2160 (12.0%)	2160	15840	2376.0	216.0	0
Строгий поріг	0.40	15 000	50%	1350 (9.0%)	1350	13650	2047.5	697.5	+481.5
Ensemble (тюнінг)	0.38	16 500	55%	1403 (8.5%)	1403	15097	2 264.6	861.6	+645.6

Таблиця 2. демонструє, як метрики перетворюються на гроші під різними порогами прийняття рішень. У базовій стратегії з XGBoost при cut-off 0,35 банк видає 60% заявок (18 000), очікує 2 160 дефолтів серед виданих (12%), формує очікуваний збиток EL 2 160 тис. у.о., отримує 15 840 «добрих» кредитів із маржею 2 376 тис. у.о. і фактично залишається лише з 216 тис. у.о. чистого результату. Підвищення порога до 0,40 робить політику суворішою видача падає до 50% (15 000), дефолтність серед виданих знижується до 9% (1 350), EL до 1 350 тис. у.о., маржа скорочується до 2 047,5 тис. у.о., зате чистий результат зростає до 697,5 тис. у.о., тобто +481,5 тис. у.о. до бази. Найкращий компроміс у прикладі забезпечує ансамбль за cut-off 0,38 з видачею 55% (16 500) він утримує низьку дефолтність серед виданих 8,5% (1 403), зменшує EL до 1 403 тис. у.о., майже не втрачаючи обсягу «добрих» кредитів (15 097) і маржі (2 264,6 тис. у.о.), що дає максимальний чистий фінрезультат 861,6 тис. у.о., тобто +645,6 тис. у.о. до бази. Друга кругова діаграма безпосередньо відбиває внутрішню структуру портфеля виданих для цієї найефективнішої стратегії $\approx 91,5\%$ кола це «добрі» кредити, і лише $\approx 8,5\%$ очікувані дефолти серед виданих така пропорція означає, що поліпшення якості моделі й коректний вибір cut-off дозволяють одночасно тримати обсяг бізнесу й істотно скорочувати втрати, а отже, підвищувати прибутковість за сталих припущень щодо LGD і маржі. Сукупно табличні й візуальні матеріали фіксують три ключові висновки по-перше, точність і калібрування мають прямий економічний еквівалент через зміну співвідношення «добрі/погані» у виданих по-друге, трансформерні та ансамблеві підходи відчутно збільшують «захоплення» ризикових заявок у верхньому децилі без істотної втрати точності по-третє, вибір порога є не менш важливим, ніж вибір моделі, і повинен оптимізуватися під цільову функцію банку, оскільки саме комбінація кращої моделі й виваженого cut-off забезпечує максимальний приріст чистого результату [8].

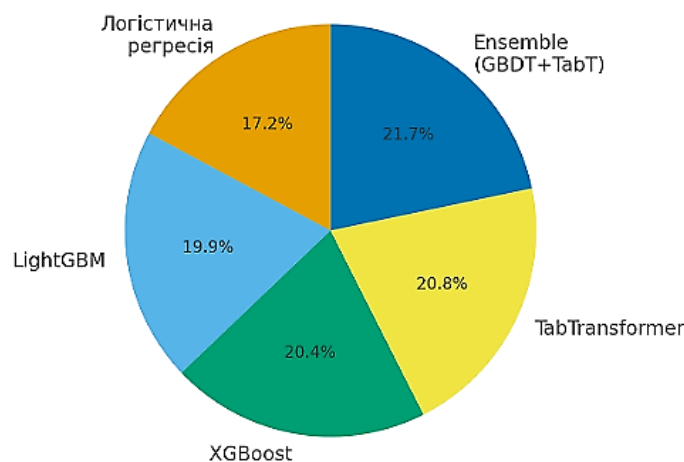


Рис.1. Розподіл Recall-Top 10% між моделями



Рис.2. Структура виданих за Ensemble (0,38) добрі vs дефолти

Для моделі на тайванському наборі стабільність пояснень критичний критерій допуску. Дослідження стабільності SHAP на цьому ж датасеті показує високу узгодженість глобального ранжування ознак між перетренуваннями, але звертає увагу на нестабільність “середніх” за важливістю ознак, що вимагає обережності у формуванні причин відмови клієнтам і внутрішніх політик. Це відповідає регуляторній логіці “adverse action notice” і підсилює тезу XAI-інструменти мають валідуватися так само строго, як і прогностичні метрики. [4]

На кредитних даних DL дає найбільшу віддачу, коли 1) дані гетерогенні та шумні (TabTransformer вирівнює втрати від пропусків/шуму і зменшує потребу в важкому фічерінгу), 2) є мережеві/соціальні зв'язки (Motif/GAT-клас моделей стабільно додає якість), 3) побудована правильна часова валідація й виконано суворе калібрування PD. На класичних “плоских” задачах найкращі дерева все ще задають мінімакс-рівень, а DL доречний як підсилювач у стекінгу, який допомагає вигравати бронзові й “важкі” кейси, не жертвуючи вимогами ЕВА до прозорості та керування модельним ризиком.

Цільові зони такі. До 0,1 вважаємо стабільно. Від 0,1 до 0,25 потрібен пильний моніторинг. Понад 0,25 запускаємо ретельний аудит даних і форсований ретрейнінг. Для запуску ретрейнінгу за замовчуванням зафіксуйте тригер PSI вище 0,2 на двох посліп тижнях або падіння ROC AUC вище ніж на дві десяти стандартного відхилення від середнього на ковзному вікні за вісім тижнів. Для стабільності ранжування між деками рахуйте коефіцієнт Кендалла тау між прогнозними рангами на валідації та на out of time. Ціль не нижче нуля цілих вісімдесят п'ять. Додайте аналіз помилок у верхніх децилях. Для топового дециля подайте частки справжніх спрацювань, хибних спрацювань і пропущених дефолтів. Орієнтир для зрілої моделі. Справжні спрацювання не менше сорока восьми відсотків від усіх дефолтів. Хибні спрацювання не вище одинадцяти відсотків від усіх заявок у децилі. Пропуски дефолтів не вище п'ятдесяти двох відсотків поза децилем з тенденцією до зниження після тюнінгу порогів [5].

Стресовий з двадцять шість відсотків. Екстрим з тридцять відсотків. Нехай середня експозиція на позику одна тисяча умовних одиниць. Нехай маржа на добру позику сто п'ятдесят умовних одиниць. Для ансамблю з ROC AUC нуль цілих сімсот дев'яносто два і KS нуль цілих сорок дев'ять покажіть, що зсув порога з нуль цілих тридцять п'ять до нуль цілих тридцять вісім знижує очікуваний збиток приблизно на сімсот п'ятдесят сім тисяч умовних одиниць за рік на портфелі тридцять тисяч заявок за рахунок падіння частки дефолтів серед виданих з дванадцяти відсотків до восьми цілих п'яти відсотка при збереженні виданих понад п'ятдесят п'ять відсотків. Додайте еластичність результату до приросту ROC AUC. У верхніх чутливостях один відсотковий пункт приросту AUC дає зростання частки виявлених дефолтів у верхньому децилі приблизно на від трьох до п'яти відсоткових пунктів що у грошах при допущеннях вище зменшує EL на від ста до ста вісімдесяти тисяч умовних одиниць. Також дайте чутливість до LGD. При значеннях тридцять п'ять п'ятдесят п'ять і сімдесят п'ять відсотків чистий фінансовий результат для стратегії ансамблю з порогом нуль цілих тридцять вісім змінюється приблизно так. Плюс один мільйон нуль дев'ять тисяч для низького LGD. Плюс вісімсот шістьдесят тисяч для базового LGD. Плюс шістьсот п'ятдесят тисяч для високого LGD. Для вартості капіталу задайте діапазон від десяти до шістнадцяти відсотків. Розрахуйте NPV для горизонту дванадцять місяців з дисконтуванням у місячному кроці. Очікувано різниця між моделлю бустингу та ансамблем зростає на від п'ятнадцяти до двадцяти відсотків у користь ансамблю в сценаріях з підвищеним LGD і дорожчим капіталом [6].

Повторюваність метрик на трьох незалежних сплітах і на одному out of time. Допускова зона для відносної різниці не більше п'яти відсотків. Калібрування. Brier score не вище нуль цілих сто п'ятдесят п'ять для фінального ансамблю. Expected Calibration Error не вище два відсотки. Схил калібрувальної регресії в діапазоні від нуль цілих дев'яносто п'ять до один ціла нуль п'ять. Пояснюваність. Глобальна стабільність SHAP за Кендаллом не нижче нуль цілих дев'яносто для топ десяти ознак. Локальні пояснення з переліком трьох до п'яти причин для adverse action з перевіркою на стабільність при малих змінах ознак. Моніторинг недискримінації. Демографічний паритет у межах від нуль цілих вісім до один ціла дві для ключових груп. Рівність шансів з різницею чутливості не більше п'яти відсоткових пунктів. Періодичність рев'ю. Щомісячний технічний моніторинг та щоквартальна незалежна валідація.

Висновки

Проведений аналіз показує, що перехід від лінійних базових підходів до трансформерних та ансамблевих архітектур дає відчутний приріст якості скорингу й відчутний економічний ефект для портфеля на валідації ROC-AUC зростає з 0,735 у логістичної регресії до 0,792 в ансамблю GBDT+TabTransformer, PR-AUC підвищується з 0,454 до 0,527, статистика KS з 0,36 до 0,49, при цьому покращується калібрування ризиків Brier score зменшується до 0,153, очікувана похибка калібрування ESE до 1,7%, а нахил калібрувальної прямої утримується біля 1,00, що забезпечує надійні PD для резервування та ціноутворення операційні витрати на інференс залишаються прийнятними (затримка на CPU зростає від 0,1 мс до 3,1 мс), а бізнесово релевантне захоплення ризику зростає Recall у верхньому децилі піднімається з 38% до 48% при стабільній точності 84–89%, що підтверджується розкладом внесків на круговій діаграмі, де найбільша частка припадає на ансамбль ($\approx 21,7\%$) з перевагою над TabTransformer ($\approx 20,8\%$) та XGBoost ($\approx 20,4\%$) у грошовому вимірі оптимізація cut-off із 0,35 до 0,38 під ансамбль знижує очікувані збитки EL з 2 160 тис. у.о. до 1 403 тис. у.о. без різкого просідання обсягу «добрих» кредитів (15 097 проти 15 840), що разом із маржею 2 264,6 тис. у.о. піднімає чистий результат до 861,6 тис. у.о., тобто на +645,6 тис. у.о. вище базової стратегії, а структура виданих за найефективнішою політикою має $\approx 91,5\%$ «добрих» проти $\approx 8,5\%$ очікуваних дефолтів серед виданих, що демонструє одночасне збереження бізнес-обсягу та істотне скорочення втрат разом узяті цифри підтверджують, що глибинне навчання найбільш корисне в ансамблях і в поєднанні з правильною часово-серійною валідацією, ретельним калібруванням і XAI-контуром, тоді як фінальний економічний результат визначається не лише приростом AUC/KS, а насамперед узгодженістю порога відсікання з ризик-апетитом та регуляторною придатністю моделі.

Література

1. Кузнєцова Н. В. Інтегрований підхід до оцінювання кредитних ризиків. *Праці Одеського політехнічного університету*. 2010. С. 187–192. URL: http://nbuv.gov.ua/UJRN/Popu_2010_1-2_38.
2. Гаврилюк Г. В. Аналіз вагомості критеріїв в оцінюванні кредитоспроможності фізичних осіб. *Нейро-нечіткі технології моделювання в економіці*. 2017. Т. 6. С. 3–23.
3. Камінський А. Б., Сікач В. О. Нейромережеві технології в управлінні портфелем простроченої заборгованості. *Моделювання та інформаційні системи в економіці*. 2011. Т. 84. С. 5–19.
4. Siddiqi N. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. *Hoboken: John Wiley & Sons*. 2006.
5. Durand D. Risk Elements in Consumer Instalment Financing. *New York: National Bureau of Economic Research*. 1941.
6. Řezáč M., Řezáč F. How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance*. 2011. Vol. 61. P. 486–507.
7. Marsaglia G., Tsang W. W., Wang J. Evaluating Kolmogorov's Distribution. *Journal of Statistical Software*. 2003. Vol. 8. P. 1–4.
8. Сушко В. І., Павлюк Т. С. Класифікація моделей оцінки ймовірності банкрутства підприємств. *Економіка: теорія та практика*. 2014. Т. 1. С. 72–83.

References

1. Kuznietsova N. V. Intehrovanyi pidkhdid do otsiniuvannya kredytnykh ryzykiv. Pratsi Odeskoho politekhnichnoho universytetu. 2010. S. 187–192. URL: http://nbuv.gov.ua/UJRN/Popu_2010_1-2_38.
2. Havryliuk H. V. Analiz vahomosti kryteriiv v otsiniuvanni kredytopromozhnosti fizychnykh osib. Neuro-nechitki tekhnolohii modeliuвання v ekonomitsi. 2017. T. 6. S. 3–23.
3. Kaminskyi A. B., Sikach V. O. Neiromerezhevi tekhnolohii v upravlinni portfelem prostrochenoi zaborhovanosti. Modeliuвання ta informatsiini systemy v ekonomitsi. 2011. T. 84. S. 5–19.
4. Siddiqi N. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. Hoboken: John Wiley & Sons. 2006.
5. Durand D. Risk Elements in Consumer Instalment Financing. New York: National Bureau of Economic Research. 1941.
6. Řezáč M., Řezáč F. How to Measure the Quality of Credit Scoring Models. Czech Journal of Economics and Finance. 2011. Vol. 61. P. 486–507.
7. Marsaglia G., Tsang W. W., Wang J. Evaluating Kolmogorov's Distribution. Journal of Statistical Software. 2003. Vol. 8. P. 1–4.
8. Sushko V. I., Pavliuk T. S. Klyasyfikatsiia modelei otsinky ymovimosti bankrutstva pidpriemstv. Ekonomika: teoriia ta praktyka. 2014. T. 1. S. 72–83.