УДК 336.711/.717+336.77:330.567.22]:004.85(477)
JEL G32, C38
ORCID ID https://orcid.org/0009-0008-7017-0175
DOI https://doi.org/10.17721/tppe.2025.50.7

Красовицький Д. О., аспірант
КНУ імені Тараса Шевченка
daniel.krass2112@knu.ua

АНАЛІЗ ВПЛИВУ ПОКАЗНИКІВ, НАЯВНИХ У КРЕДИТНОМУ РЕЄСТРІ, НА ЙМОВІРНІСТЬ ДЕФОЛТУ ПОЗИЧАЛЬНИКІВ-ФІЗИЧНИХ ОСІБ

У статті досліджено ефективність використання показників, наявних у Кредитному реєстрі Національного банку України, для оцінки ймовірності дефолту позичальників-фізичних осіб. Актуальність роботи зумовлена необхідністю підвищення точності прогнозування кредитних ризиків у банківському секторі України, особливо в умовах війни, що триває. Метою дослідження є оцінка прогностичної здатності змінних із Кредитного реєстру, зокрема віку, кількості днів прострочки, показника боргового навантаження (DSTI) та інших характеристик, а також визначення їхнього впливу на ймовірність дефолту. Методологія роботи базується на використанні багатофакторного регресійного аналізу, логістичної регресії та сучасних методів машинного навчання. Для оцінки важливості змінних застосовувалися статистичні підходи Information Value (IV) та Variable Importance (VI), які дозволяють ідентифікувати ключові детермінанти кредитного ризику. Результати дослідження підтвердили, що дані Кредитного реєстру є надійним джерелом для оцінки ймовірності дефолту. Найбільший вплив на дефолтність позичальників має показник боргового навантаження та кількість днів прострочки, тоді як вплив віку та доходів є менш вираженим. Застосування методу IV показало, що більшість змінних мають помірну або високу прогностичну цінність, що підкреслює їхню значущість для побудови моделей. Отримані результати свідчать про значний потенціал використання даних Кредитного реєстру для оптимізації процесів управління кредитними ризиками. Моделі продемонстрували хорошу класифікаційну якість, що свідчить про прийнятність показників Кредитного реєстру для моделювання ймовірності дефолту позичальників-фізичних осіб. Перспективи подальших досліджень пов'язані з розширенням аналізу впливу макроекономічних факторів, таких як ВВП чи інфляція, на ймовірність дефолту. Крім того, інтеграція методів керованого та некерованого машинного навчання дозволить сегментувати позичальників за рівнем ризику, що сприятиме точнішому управлінню кредитними портфелями.

Ключові слова: *машинне навчання, кредитний ризик, ймовірність дефолту, кредитний реєстр*

Постановка проблеми. Повномасштабне вторгнення рф спричинило підвищену увагу до фінансової стійкості. На тлі цього Національний банк активно працює над оптимізацією процесів оцінки кредитних ризиків. При цьому нові джерела даних можуть покращити процес прогнозуванні ймовірності дефолту позичальників. Одним з таких джерел може бути Кредитний реєстр, до якого протягом останніх п’яти років подаються дані про якість кредитного портфеля фізичних та юридичних осіб. Проте економетричні та математичні аспекти використання цих показників потребують додаткового дослідження. Постає питання, чи здатні наявні дані забезпечити точність оцінки ризиків? Яким чином різні характеристики позичальників впливають на ймовірність дефолту? Ці питання визначають сутність проблеми, яка потребує аналізу та наукового обґрунтування.

Аналіз останніх публікацій. Питання моделювання впливу різних показників на ймовірність дефолту позичальників є предметом досліджень як зарубіжних, так і українських дослідників. Р. Турксон та ін. використали метод Variable Importance для того, щоб відібрати, які змінні впливають на ймовірність дефолту, а потім дослідили їхній вплив за допомогою різних методів машинного навчання [1]. Ф. Зігріст і Н. Лойенбергер також використали методи машинного навчання для знаходження зв'язку між різними змінними та дефолтом [2]. Більше того, автори використали метод SHAP для оцінки значущості цього впливу. Українські дослідники В. Філатов та А. Камінський досліджували вплив різних змінних на ймовірність дефолту позичальників-фізичних осіб на основі даних Кредитного реєстру [3]. Вони з'ясували на основі логістичної регресії, що кредитний ризик, ставка за кредитом та дохід позичальника мають статистично значимий нелінійний вплив на ймовірність дефолту. Ф. Доко [4] з'ясував на основі методу Information Value, що кількість днів прострочки є вирішальним показником для дефолту. Е. Думітреску [5] використали результати дерева рішень на основі показників кредитного реєстру, у якості незалежних змінних в логістичній регресії. Робота О. Ляшенко була сконцентрована не на пояснювальній здатності окремих змінних, а наскільки в цілому їхня сукупність прогнозувала дефолт [6]. Ф. Шень зауважив, що пояснювальна здатність змінних може викривлюватись дисбалансом у залежній змінній [7]. В. Бао використав одночасно методи керованого та некерованого машинного навчання з метою покращення прогностичної здатності [8].

Невирішеними аспектами проблеми є економетрична оцінка ефективності використання показників із Кредитного реєстру для прогнозування ймовірності дефолту. Зокрема, залишається відкритим питання, чи здатна економетрична модель адекватно врахувати нелінійний вплив ключових характеристик позичальників, таких як рівень доходу, вік, та показники обслуговування боргу. Також невідомо, чи забезпечує поєднання різних методів моделювання, наприклад, регресійного аналізу та машинного навчання (керованого чи некерованого), більш точну якість прогнозу дефолту.

Метою статті є дослідження ефективності використання показників, наявних у Кредитному реєстрі, для оцінки ймовірності дефолту. У рамках цього дослідження планується визначити, чи допомагають ці показники визначити дефолтних позичальників, чи можна виміряти вплив на ймовірність дефолту, та чи можна вчасно відслідкувати міграцію позичальника в дефолт.

Методологія дослідження ґрунтується на використанні економіко-математичних інструментів для аналізу взаємозалежностей між показниками, наявними у Кредитному реєстрі, та ймовірністю дефолту позичальників. У роботі застосовуються методи багатофакторного регресійного аналізу, моделювання за допомогою логістичних функцій та машинного навчання для ідентифікації ключових детермінантів кредитного ризику.

Результати дослідження. Показники, що містяться в Кредитному реєстрі, та правила їхнього заповнення визначені в документі «Технічні умови подання банками до Національного банку України звіту про кредитні операції банку з боржником-фізичною/юридичною особою та ведення Кредитного реєстру Національного банку України». Структура реєстру передбачає дві підсистеми: для фізичних осіб і для юридичних осіб. Підсистема для фізичних осіб додатково поділена на три частини: інформація про боржника, дані про кредитну операцію та відомості про забезпечення

кредиту. Унікальними ідентифікаторами, що забезпечують зв'язок між цими елементами, є код клієнта (який пов’язує боржника з кредитною операцією) та код кредиту (який пов’язує кредитну операцію із забезпеченням).

Дані в Кредитному реєстрі оновлюються з місячною частотою станом на 1 число кожного місяця. Однак така частота є надлишковою для оцінки кредитного ризику, адже для класифікації кредиту як непрацюючого необхідно 90 днів (3 місяці). Відтак у межах даного дослідження використані квартальні дані, починаючи з 1 квітня 2020 року — дати, з якої реєстр містить якісні та верифіковані дані.

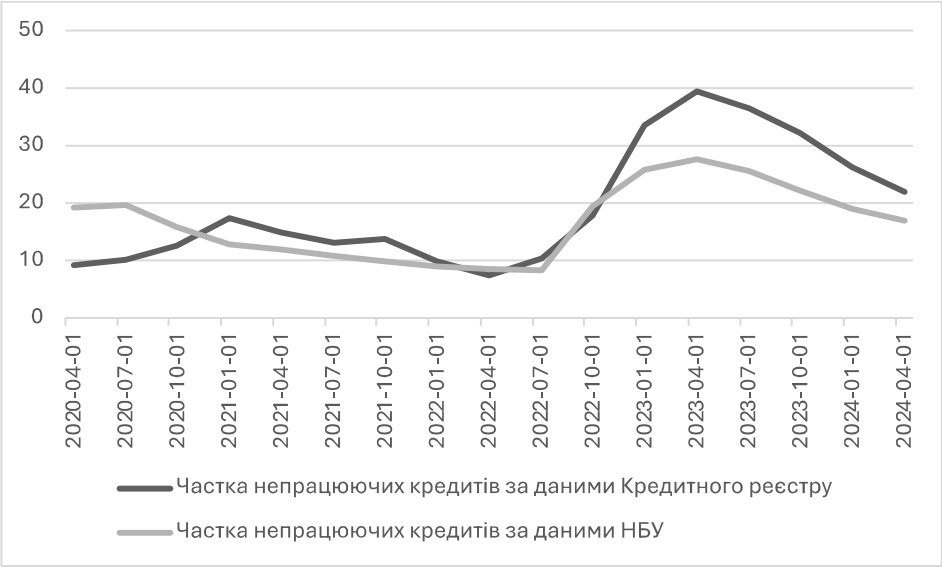


Рис. 1. Динаміка частки непрацюючих кредитів за даними Кредитного реєстру та НБУ
Джерело: розраховано автором на основі [9].

На рис. 1 представлено порівняння динаміки частки непрацюючих кредитів фізосіб з Кредитного реєстру та файлу статистичної звітності на сайті НБУ. Аналіз свідчить, що дані Кредитного реєстру адекватно відображають загальні тенденції розвитку кредитного портфеля домогосподарств. У деякі періоди значення з різних джерел є доволі близькими, що дозволяє зробити висновок про прийнятність використання даних реєстру для моделювання.

Інформація про боржника у Кредитному реєстрі містить лише дві характеристики, що дозволяють оцінити його кредитоспроможність – вік та інформація про доходи. Доходи позичальника поділяються на підтверджені та непідтверджені. Підтверджений дохід відображає середньомісячний дохід боржника за останні шість місяців, верифікований документами, виданими третьою стороною, такими як довідка з місця роботи, довідка про доходи тощо. Будь-які інші доходи зазначаються як непідтверджені. Важливо, що підтверджений дохід враховується лише для особи, на яку оформлено кредит. Наприклад, якщо іпотека оформлена на одного з подружжя, доходи іншого зазначаються як непідтверджені.

Вік позичальника є важливим фактором, що може суттєво впливати на ймовірність дефолту. Дослідження Р. Турксон та ін. вказує, що вік позитивно корелює з ризиком неплатоспроможності: з віком зростає ймовірність втрати платоспроможності [1]. Це пояснюється тим, що люди старшого віку часто стикаються з труднощами у пошуку роботи

у разі втрати доходу, що збільшує їх фінансову уразливість. У розвинених країнах, зокрема в Словаччині, деякі механізми стримування кредитного ризику передбачають вікову градацію, що також свідчить про важливість віку в управлінні кредитними ризиками.

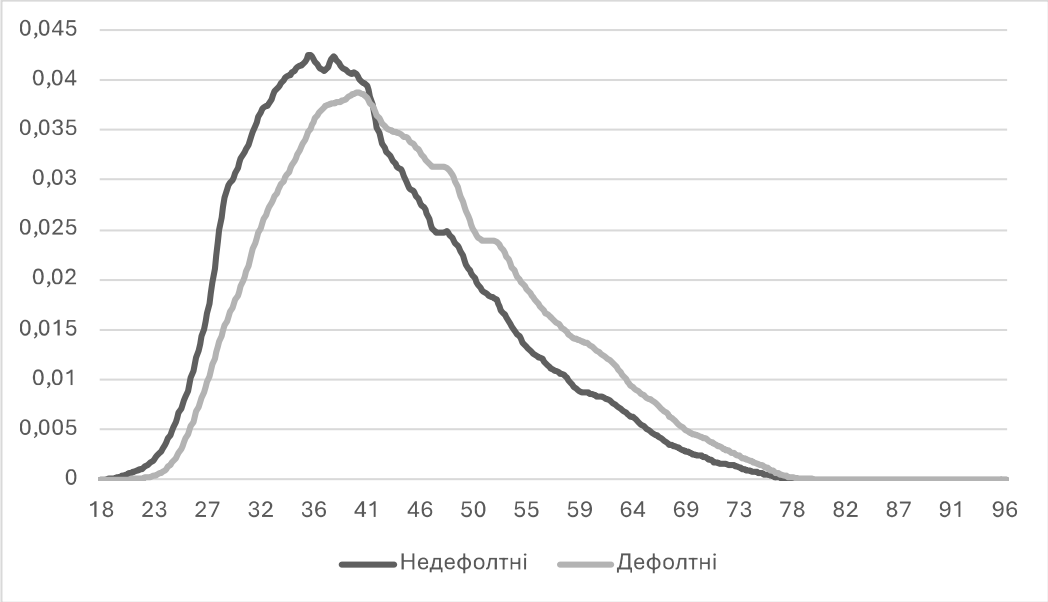


Рис. 2. Розподіл позичальників за віком в розрізі дефолтні/недефолтні
Джерело: розроблено автором на основі [9].

Частково це підтверджується статистикою Кредитного реєстру. На рис. 2 помітно, що розподіл неплатоспроможних позичальників за віком зміщений вправо відносно платоспроможних. Водночас необхідне додаткове дослідження за допомогою IV чи ROC кривої для визначення значущості цієї різниці.

Інформація про кредитну операцію охоплює широкий спектр показників, що характеризують умови та стан конкретного кредиту на визначену дату. До таких показників належать: клас ризику, причина дефолту, сума кредитного ризику оцінена відповідно до вимог Постанови 351 «Про затвердження Положення про визначення банками України розміру кредитного ризику за активними банківськими операціями», процентна ставка, сума заборгованості (прострочена та непрострочена) за основним боргом і процентами, сума наданого кредиту, вид кредиту, дати укладання та погашення кредитної угоди, кількість днів прострочки [10]. Ці показники є ключовими для аналізу ризиків кредитних операцій та їх впливу на ймовірність дефолту позичальника.

Інформація про заставу за кредитом містить не так багато полів, що можна використовувати з метою оцінки ризику. По-перше, близько 80% суми чистого гривневого кредитного портфеля фізичних осіб станом на 1 грудня 2024 року не належать ані до автокредитів, ані до іпотечних. Це означає, що переважна більшість кредитів не мають застав взагалі. По-друге, тип застави за її наявності є похідним від типу кредиту. Відповідно найпоширенішими видами є автомобілі та нерухомість. По-третє, числових показників в цій базі є лише два – вартість застави за угодою, яка не міняється з часом, та вартість застави згідно з останньою оцінкою, яка може змінюватись. Банк зацікавлений в щорічній оцінці вартості забезпечення, оскільки вона визначатиме фінансові втрати в разі

дефолту і конфіскації забезпечення на користь банку. Зв'язок між вартістю застави та рівнем ризику є прямим: чим вище відношення кредиту до вартості забезпечення, тим більш ризиковим для банку є такий кредит. Центральні та комерційні банки можуть регулювати це співвідношення залежно від типів кредиту. Такий інструмент називається кредит-до-застави (англ. Loan-to-Value, далі LTV). Наприклад, за державною програмою «єОселя» цей показник на момент видачі кредиту не має перевищувати 80% [11].

Аналізувати вплив цих показників на ймовірність дефолту можна різними способами. Першим з них є Information value (IV) – це статистичний показник, який використовується для оцінки предиктивної потужності змінних у задачах класифікації. IV допомагає визначити, наскільки добре змінна відрізняється між бінарними класами, наприклад, між позичальниками що стали неплатоспроможними та не стали такими. Цей метод був зокрема використаний в дослідженні В. Філатова та А. Камінського [3].

IV базується на концепції Weight of Evidence (WOE), яка вимірює силу кожної змінної в розрізні двох класів. IV використовується для виявлення найважливіших змінних для побудови моделі і відбору змінних. Фактично цей метод є попереднім етапом перед побудовою моделі.

Розрахунок відбувається за наступними кроками:

1. Розділення змінну на категорії або інтервали (біни).
2. Для кожного біна розраховується кількість спостережень кожного класу (good і bad).
3. Обчислення Weight of Evidence (WOE):

$$WOE_i = \ln\left(\frac{\frac{Good_i}{Total\ good}}{\frac{Bad_i}{Total\ bad}}\right) \quad (1)$$

де WOE_i– розрахунковий рівень Weight of Evidence для біна *i*;
$Good_i$ – кількість спостережень позитивного класу для біна *i*;
Bad_i – кількість спостережень негативного класу для біна *i*;
$Total\ good$ – загальна кількість спостережень позитивного класу;
$Total\ bad$ – загальна кількість спостережень негативного класу;

4. Обчислення IV

$$IV = \sum_i WOE_i * \left(\frac{Good_i}{Total\ good} - \frac{Bad_i}{Total\ bad}\right) \quad (2)$$

Таблиця 1.

Діапазони значень IV та пояснення до них

Діапазон значень IV	Пояснення
[0;0.02)	Немає прогностичної цінності
[0.02;0.1)	Слабка прогностична цінність
[0.1;0.3)	Помірна прогностична цінність
[0.3;0.5)	Висока прогностична цінність
[0.5;+∞]	Дуже висока прогностична цінність (рідко зустрічається в реальних даних)

Джерело: розроблено на основі [12]

Головною перевагою цього методу є те, що за його допомогою можна відібрати змінні, що в подальшому будуть використовуватись в моделі, що дозволить їй швидше працювати. Проте, результати можуть бути чутливими до способу бінінгу змінної. Крім того, цей метод підходить для задач з бінарними залежними змінними, і не підходить для більш складних моделей. IV також не показує напрям впливу незалежних змінних на залежну, а радше лише їхню роздільну здатність.

Результати використання цього методу із відповідними значеннями IV продемонстровані на рис 3.

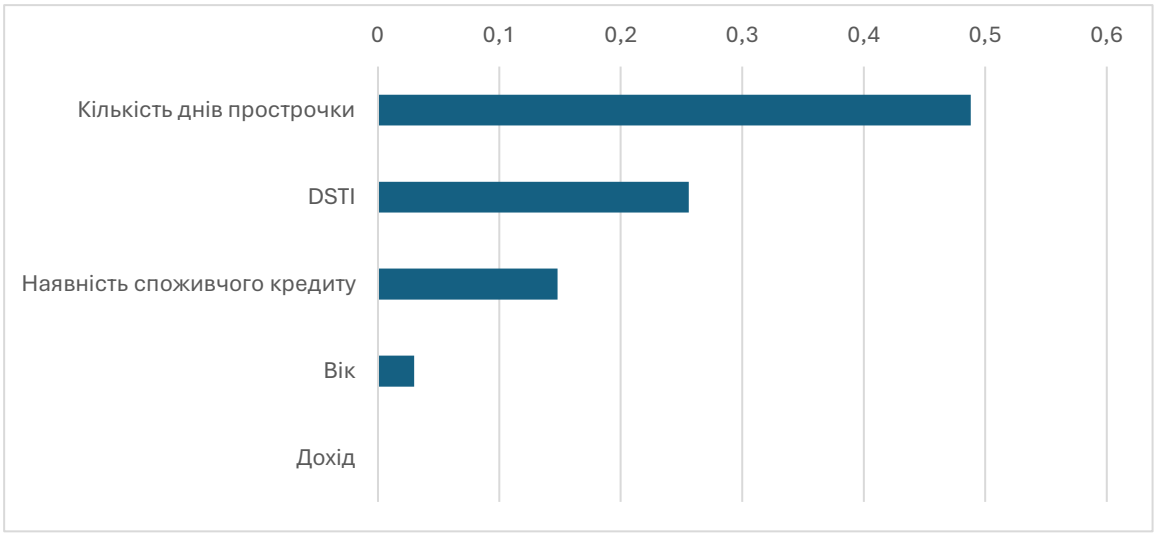


Рис. 3. Значенням IV для окремих змінних

Джерело: розраховано автором на основі Кредитного реєстру

.

Результати свідчать, що найбільший вплив на неплатоспроможність позичальників має кількість днів прострочки за кредитом. DSTI (показник боргового навантаження) має помірну прогностичну здатність. Несподівано вік не зважаючи на візуально помітну різницю на рис. 2, не має суттєвого впливу на ймовірність дефолту. Доходи позичальника також самі по собі не визначають чи стане він неплатоспроможним.

Іншим методом оцінки сили впливу є «важливість змінних» (англ. Variable importance, VI). Це комплекс методів заснований на оцінці впливу незалежних змінних на ймовірність приналежності до певного класу безпосередньо після оцінки економетричної моделі чи моделі машинного навчання. Цей спосіб можна застосовувати до всіх моделей машинного навчання, проте інтерпретація результатів буде відмінною.

У моделях логістичної регресії VI базується на величині оцінених коефіцієнтів. Більші коефіцієнти вказують на сильніший вплив на змінну відповіді. Крім того, в моделі логістичної регресії є можливість оцінити наскільки та чи інша змінна статистично значима. Для таких моделей цей метод є простим, якщо використовуються лінійні моделі, однак додавання interaction terms або поліномів може призвести до «запозичення» важливості між змінними.

У моделях дерев рішень і випадкового лісу VI розраховується на основі зменшення індексу Джині при кожному розподілі. Чим більшим за величиною є зниження індексу Джині, тим важливішою є змінна. Аналіз VI також можна застосувати до нейронних мереж.

Відповідно до Г. Гарсона важливість змінної обчислюється як сума всіх зважених зв'язків між вузлами [13].

Р. Турксон та ін. використали VI для того, щоб з'ясувати, які змінні мають найсильніший вплив на ймовірність дефолту [1]. Їхні результати засвідчили, що вік, сума на депозитному рахунку на момент видачі кредиту та статус першого платежу (прострочений чи ні) мають найбільший та статистично значимий вплив. Значення VI за цими змінними були найбільшими, а результати моделювання використовуючи їх та повний набір доступних змінних не відрізнялись суттєво.

Побудуємо моделі логістичну, дерево рішень, випадковий ліс та нейронну мережу, де у якості залежної змінної буде бінарна змінна дефолт/недефолт, а інші змінні будуть незалежними. Якість моделі оцінено за допомогою F-score.

Таблиця 2.

Нормалізовані оцінки VI на різних моделях

Змінна	**Логістична модель**	**Дерево рішень**	**Випадковий ліс**	**Нейронна мережа**
Вік	56.47	11.77	69.26	0
Наявність споживчого кредиту	92.42	71.2	0	66.06
Дохід	0	0	55.24	54.01
Кількість днів прострочки	95.12	86.94	41.55	100
DSTI	100	100	100	37.82
F-score	0.63	0.65	0.74	0.68

Джерело: розраховано автором на основі Кредитного реєстру

Вибірку для дослідження було попередньо очищено від пропущених та екстремальних значень. Наприклад вік клієнта мав бути від 18 до 100 років, DSTI не більше 300%, а доходи додатніми та більше 0. Кількість спостережень у фінальній вибірці склала 1293974. Результати розрахунків методом VI представлено у таблиці 2. Показники було нормалізовано від 0 до 100 з метою порівнюваності між моделями. За 100 приймалось максимальне значення поміж усіх оцінок, за 0 - мінімальне. Це не означає, що ефект від тої чи іншої змінної дорівнює 0, лише, що він є найменшим. Можна зробити висновок, що DSTI, має найбільший вплив на ймовірність дефолту, оскільки він показав найвищі результати поміж трьох моделей з чотирьох. На другому місці кількість днів прострочки, що також консистентно із попереднім методом. Доходи та вік, хоча в окремих моделях і мали середні показники VI врешті-решт мають найменший вплив на ймовірність дефолту позичальників фізичних осіб. Якість моделей оцінена за F-score, що вимірюється від 0 до 1, демонструє доволі близькі та високі результати, що говорить про хорошу класифікаційну якість моделі та можливість розрахунку ймовірностей дефолту для позичальників фізичних осіб.

Висновки. У процесі дослідження було виявлено, що показники, наявні в Кредитному реєстрі, такі як вік, наявність споживчого кредиту, DSTI та інші показники, здатні відображати прогнозувати ймовірність дефолту позичальників-фізичних осіб. Зокрема,

встановлено двома методами, що найвагомішим фактором, який впливає на ймовірність дефолту, є показник боргового навантаження (DSTI), тоді як вплив таких змінних, як вік і доходи, є менш вираженим.

Економетричний аналіз підтвердив релевантність застосування логістичної регресії та методів машинного навчання для оцінки ризиків, хоча їхні результати дещо різняться за тим, які змінні важливі для прогнозу неплатоспроможності клієнтів. Отримані результати також підкреслили ефективність використання таких інструментів, як Information Value (IV) і Variable Importance (VI), для ідентифікації змінних із найбільшим впливом на категорійну бінарну змінну дефолт/недефолт. Аналіз даних із Кредитного реєстру показав, що більшість ключових змінних характеризується помірною або високою прогностичною цінністю.

Перспективи подальших досліджень полягають у розширенні використання інтегрованих підходів для моделювання ризиків, таких як поєднання некерованих та керованих методів машинного навчання, а також у детальному аналізі впливу макроекономічних факторів на ймовірність дефолту. Крім того, існує можливість аналізу додаткових факторів таких як кредитна історія, процентна ставка за кредитами та вартість застави.

Література

1. Турксон Р. Машинне навчання для прогнозування кредитоспроможності банківських клієнтів / Р. Турксон, Е. Баагіре, Г. Веня // Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR). – 2016. [Електронний ресурс] – Режим доступу до ресурсу: https://ieeexplore.ieee.org/document/7585216/authors#authors.
2. Зігрист Ф., Лойенбергер Н. Машинне навчання для оцінки ризику дефолту корпоративних позичальників: багатоперіодний прогноз, кореляція слабкості, кредитні портфелі та ймовірності крайових ризиків // Європейський журнал операційних досліджень. – 2023. – Т. 305, № 3. – С. 1390–1406. [Електронний ресурс] – Режим доступу до ресурсу: https://ideas.repec.org/a/eee/ejores/v305y2023i3p1390-1406.html
3. Філатов В., Камінський А. Застосування скорингового підходу до моніторингової функції Кредитного реєстру центрального банку // Наукові записки НаУКМА. Економіка. – 2021. – Т. 6, № 1. – С. 73–83. – doi: 10.18523/2519-4739.2021.6.1.73-83.
4. Доко Ф., Калайджиський С., Мішковський І. Модель кредитного ризику на основі даних Кредитного реєстру центрального банку // Журнал управління ризиками та фінансами. – 2021. – Т. 14, № 3. – С. 138. – doi: 10.3390/jrfm14030138.
5. Думітреску Е., Салліван Х., Хурлін К., Токпаві С. Машинне навчання для кредитного скорингу: вдосконалення логістичної регресії за допомогою нелінійних ефектів дерев рішень // Європейський журнал операційних досліджень. – 2022. – Т. 297, № 3. – doi: 10.1016/j.ejor.2021.06.053.
6. Ляшенко О., Кравець Т., Костовецький Є. Машинне навчання та методи балансування даних для прогнозування банкрутства // *Економіка*. – 2023. – Т. 102, № 2. – С. 28–46. – doi: 10.15388.
7. Шень Ф., Чжао С., Лі Ч., Лі К., Мен Ч. Новітня ансамблева модель класифікації на основі нейронних мереж та техніки оптимізації класифікатора для оцінки дисбалансованих кредитних ризиків // *Physica A: Statistical Mechanics and Its Applications*. – 2019. – Т. 526. – Стаття ID: 121073. – doi: 10.1016/j.physa.2019.121073.
8. Бао В., Нін Л., Конг Ю. Інтеграція некерованих і керованих алгоритмів машинного навчання для оцінки кредитного ризику // *Expert Systems with Applications*. – 2019. – Т. 128. – С. 301–315. – ISSN: 0957-4174. – doi: 10.1016/j.eswa.2019.02.033.
9. Національний банк України. 2024. Статистика фінансового сектору. [Електронний ресурс] – Режим доступу до ресурсу: https://bank.gov.ua/ua/statistic/sector-financial

10. Національний банк України. (2016). Постанова №351 від 30.06.2016 «Про затвердження Положення про визначення банками України розміру кредитного ризику за активними банківськими операціями». [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/v0351500-16#Text

11. Укрфінжитло. [Електронний ресурс] – Режим доступу до ресурсу: https://ukrfinzhytlo.in.ua/

12. TIBCO. Information Value and Weight of Evidence Analysis. [Електронний ресурс] – Режим доступу до ресурсу: https://docs.tibco.com/pub/sfire-dsc/6.5.0/doc/html/TIB_sfire-dsc_user-guide/GUID-07A78308-525A-406F-8221-9281F4E9D7CF.html

13. Гарсон Г.Д. Інтерпретація ваг зв'язків у нейронних мережах // AI Expert. – 1991. – Т. 6, № 4. – С. 46–51. – doi: 10.5555/129449.129452.

References

1. Turkson R.Mashinne navchannya dlya prognozuvanna kreditospromozhnosti bankivskih klientiv / R. Turkson, E. Bagire, G. Venya // Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR). – 2016. [Elektronnuy resurs] – Rezhim dostupu do resusrsu: https://ieeexplore.ieee.org/document/7585216/authors#authors.

2. Zigrist F., Leunberger N. Mashinne navchannya dlya ozinki ryzyku defoltu korporatuvnykh pozychalnikiv: bahatoperiodnyi prohnoz, koreliatsiia slabkosti, kredytni portfeli ta ymovirnosti kraiovykh ryzykiv // Yevropeiskyi zhurnal operatsiinykh doslidzhen. – 2023. – T. 305, № 3. – S. 1390–1406. [Elektronnyi resurs] – Rezhym dostupu do resursu: https://ideas.repec.org/a/eee/ejores/v305y2023i3p1390-1406.html

3. Filatov V., Kaminskyi A. Zastosuvannia skorynhovoho pidkhodu do monitorynhovoi funktsii Kredytnoho reiestru tsentralnoho banku // Naukovi zapysky NaUKMA. Ekonomika. – 2021. – T. 6, № 1. – S. 73–83. – doi: 10.18523/2519-4739.2021.6.1.73-83.

4. Doko F., Kalaidzhyskyi S., Mishkovskyi I. Model kredytnoho ryzyku na osnovi danykh Kredytnoho reiestru tsentralnoho banku // Zhurnal upravlinnia ryzykamy ta finansamy. – 2021. – T. 14, № 3. – S. 138. – DOI: 10.3390/jrfm14030138.

5. Dumitresku E., Sallivan Kh., Khurlin K., Tokpavi S. Mashynne navchannia dlia kredytnoho skorynhu: vdoskonalennia lohistychnoi rehresii za dopomohoiu neliniinykh efektiv derev rishen // Yevropeiskyi zhurnal operatsiinykh doslidzhen. – 2022. – T. 297, № 3. – DOI: 10.1016/j.ejor.2021.06.053.

6. Liashenko O., Kravets T., Kostovetskyi Ye. Mashynne navchannia ta metody balansuvannia danykh dlia prohnozuvannia bankrutstva // Ekonomika. – 2023. – T. 102, № 2. – S. 28–46. – DOI: 10.15388.

7. Shen F., Chzhao S., Li Ch., Li K., Men Ch. Novitnia ansambleva model klasyfikatsii na osnovi neironnykh merezh ta tekhniky optymizatsii klasyfikatora dlia otsinky dysbalansovanykh kredytnykh ryzykiv // Physica A: Statistical Mechanics and Its Applications. – 2019. – T. 526. – Stattia ID: 121073. – doi: 10.1016/j.physa.2019.121073.

8. Bao V., Nin L., Konh Yu. Intehratsiia nekerovanykh i kerovanykh alhorytmiv mashynnoho navchannia dlia otsinky kredytnoho ryzyku // Expert Systems with Applications. – 2019. – T. 128. – S. 301–315. – ISSN: 0957-4174. – doi: 10.1016/j.eswa.2019.02.033.

9. Natsionalnyi bank Ukrainy. 2024. Statystyka finansovoho sektoru. [Elektronnyi resurs] – Rezhym dostupu do resursu: https://bank.gov.ua/ua/statistic/sector-financial

10. Natsionalnyi bank Ukrainy. (2016). Postanova №351 vid 30.06.2016 «Pro zatverdzhennia Polozhennia pro vyznachennia bankamy Ukrainy rozmiru kredytnoho ryzyku za aktyvnymy bankivskymy operatsiiamy». [Elektronnyi resurs] – Rezhym dostupu do resursu: https://zakon.rada.gov.ua/laws/show/v0351500-16#Text

11. Ukrfinzitlo. [Elektronnyi resurs] – Rezhym dostupu do resursu: https://ukrfinzhytlo.in.ua/

12. TIBCO. Information Value and Weight of Evidence Analysis. [Elektronnyi resurs] – Режим доступу до ресурсу: https://docs.tibco.com/pub/sfire-dsc/6.5.0/doc/html/TIB_sfire-dsc_user-guide/GUID-07A78308-525A-406F-8221-9281F4E9D7CF.html

13. Harson H.D. Interpretatsiia vah zviazkiv u neironnykh merezhakh // AI Expert. – 1991. – T. 6, № 4. – S. 46–51. – doi: 10.5555/129449.129452.

Krasovytskyi D. O., PhD student
Taras Shevchenko National University of Kyiv
daniel.krass2112@knu.ua

ANALYSIS OF THE IMPACT OF INDICATORS AVAILABLE IN THE CREDIT REGISTRY ON THE PROBABILITY OF DEFAULT OF INDIVIDUAL BORROWERS

The article examines the effectiveness of using the indicators available in the Credit Registry of the National Bank of Ukraine to assess the probability of default among retail borrowers. The study's relevance is driven by the need to improve the accuracy of credit risk forecasting in Ukraine's banking sector, especially during the ongoing war. The study aims to evaluate the predictive power of variables from the Credit Registry, including age, the number of days overdue, the debt-to-service-income ratio (DSTI), and other characteristics, as well as to determine their impact on default probability. The methodology uses multifactor regression analysis, logistic regression, and modern machine learning methods. This study emphasizes the benefits of combining traditional econometric models with machine learning approaches to improve model robustness and predictive performance. Statistical approaches such as Information Value (IV) and Variable Importance (VI) were applied to identify the key determinants of credit risk. The results confirm that data from the Credit Registry are a reliable source for assessing default probability. The most significant predictors of borrower default are the DSTI and the number of overdue days, while the influence of age and income is less pronounced. The application of the IV method revealed that most variables have moderate to high predictive value, emphasizing their importance in building models. The findings highlight the considerable potential of using Credit Registry data to optimize credit risk management processes. The models demonstrated high classification quality, indicating the suitability of the Credit Registry's indicators for modeling the default probability of retail borrowers. Future research prospects include expanding the analysis to assess the impact of macroeconomic factors, such as GDP growth or inflation, on default probability. The approach can be further enhanced through real-time data updates and the incorporation of systemic risk indicators to increase resilience in times of economic stress. Additionally, integrating supervised and unsupervised machine learning methods could enable the segmentation of borrowers by risk level, facilitating more precise management of credit portfolios.

Keywords: *machine learning, credit risk, default probability, credit registry*