

ФІНАНСИ, ГРОШОВИЙ ОБІГ І КРЕДИТ

УДК 330.131.7

DOI: 10.36919/2312-7872.3.2024.129

О. М. Солошенко,

кандидат економічних наук,

ПЗНВ «Європейський університет»

Бульвар Академіка Вернадського 16-В, Київ 03115

E-mail: soloshenko_s@ukr.net

ORCID: 0000-0002-1329-0989

ІННОВАЦІЙНІ ПРАКТИКИ ТА ІННОВАЦІЙНИЙ ПОТЕНЦІАЛ МЕТОДОЛОГІЇ КРЕДИТНОГО СКОРИНГУ КОМПАНІЙ МІКРОКРЕДИТУВАННЯ

Метою статті є дослідження сучасних інноваційних практик, що застосовуються в кредитному скорингу компаніями мікрокредитування, а також проведення ретельної оцінки майбутніх перспектив та інноваційного потенціалу методології кредитного скорингу загалом.

Наукові результати дослідження і прогнозування інноваційних трендів методології кредитного скорингу для мікрофінансових організацій було отримано за допомогою наступних методів: аналіз окремих компонентів сучасних інноваційних процесів побудови моделей кредитного скорингу, синтез окремих тенденцій розвитку та їх системне узагальнення в високорівневі закономірності та тренди, а також системний метод технологічного передбачення щодо майбутніх перспектив інноваційного розвитку методології кредитного скорингу у сфері мікрокредитування.

На основі сучасних трендів методології кредитного скорингу в сфері мікрокредитування доведено, що інтеграція сучасних напрацювань зі сфери машинного навчання у сферу кредитного скорингу має великий вплив на функціонування, і, як наслідок, на фінансові результати роботи компаній мікрокредитування.

Ідентифіковано ключові етапи побудови моделей кредитного скорингу спряжені з особливостями мікрокредитування, котрі підлягають найбільшому впливу інновацій та розвитку технологій, а також ідентифіковано відповідні тенденції. Таким чином виділено три основні етапи моделювання кредитних ризиків, які наразі мають найбільший інноваційний потенціал:

- 1) означення цільової змінної моделювання;
- 2) побудова базової моделі кредитного скорингу на прийнятих заявках (тільки на повних даних);
- 3) аналіз та включення у фінальну модель відхилених заявок з метою забезпечення незміщеності фінальної моделі (включаючи неповні дані).

Спрогнозовано подальші тренди розвитку методології моделювання кредитних ризиків у сфері мікрокредитування. Ідентифіковано пов'язану нагальну проблематику спричинену використанням інноваційних методів бінарної класифікації з області машинного навчання. Запропоновано ключові ідеї зокрема по вирішенню проблеми аналізу та включення відхилених заявок у фінальну модель для випадків використання новітніх методів машинного навчання дискретної природи, що є одним з ключових елементів перспектив подальших досліджень.

Ключові слова: кредитний скоринг, кредитний ризик-менеджмент, ризики мікрокредитування, машинне навчання.

Вступ та постановка проблеми. Кредитний ризик-менеджмент, зокрема методологія кредитного скорингу, пройшли довготривалий шлях своєї еволюції, який ще далекий до свого завершення, хоч і досяг вже своєї зрілості або плато продуктивності в рамках термінології циклу зрілості технологій [1]. Моделі і методи оцінювання кредитоспроможності фізичних осіб

включають у себе дуже багато апробованих на реальних задачах моделей та методів бінарної класифікації для ймовірнісного моделювання подій дефолту (виходу в прострочення заборгованості) [2]. Тим не менше, класичні банківські інституції є більш регульованими як зовнішньо національними регулюючими органами так і внутрішніми нормативними документами, що стосуються як і методології управління кредитними ризиками, так і тісно пов'язаними вимогами щодо резервування потенційних втрат власного кредитного портфелю. З іншого боку, мікрофінансові організації, які займаються мікрокредитуванням, є традиційно менш регульованими як зовнішніми регулюючими органами так і внутрішніми нормативними документами, що дозволяє впроваджувати інноваційні підходи значно швидше, а короткі терміни мікропозик обумовлюють більш динамічну і мінливу ситуацію щодо прострочень, що, в свою чергу, також дозволяє провадити більш оперативний моніторинг впливу змін [3]. Ключовою багатогранною проблематикою сучасного кредитного ризик-менеджменту у галузі мікрокредитування є детальна оцінка інноваційного потенціалу методології кредитного скорингу як складової частини стратегії управління кредитними ризиками, а також технологічне передбачення (як інструмент системного аналізу [4]) подальшого інноваційного розвитку методології кредитного скорингу у сфері мікрокредитування з наданням технологічних передбачень по кожному окремому аспекту.

Аналіз останніх джерел і публікацій. Особливості ризик-менеджменту саме в сфері мікрокредитування у вітчизняних публікаціях детально розглянуті проф. Камінським А.Б. [3], де виділено характерні ознаки та відмінності від класичних банківських послуг такі як характерні невеликі суми позик, короткі строки позик, високі відсоткові ставки, сегмент клієнтів з високим середнім ступенем ризику як загальний показник сфери мікрокредитування і, відповідно, високий відсоток неповернення позик. У зарубіжній літературі найбільш фундаментально поняття та особливості мікрокредитування розглядає Майкл А. Стегман [5].

Методологія кредитного скорингу як моделювання ймовірності майбутньої події виходу в певну кількість днів прострочення заборгованості в залежності від вхідних даних клієнта описана у багатьох зарубіжних базових працях на дану тему [6, 7]. Деталі методів моделювання представлені у вітчизняних роботах [1, 2] і, зокрема, включають у себе логістичну регресію, дерева рішень, нейронні мережі і т.д. Основоположні та сучасні праці [2, 7] також описують корисні розширені етапи моделювання і коригування моделі шляхом аналізу та включення відхилення заяв у модель, де проблемою є відсутність спостережуваності цільової поведінкової змінної, але маючи спостережуваність вхідних змінних.

Невирішені частини проблеми. До ключових невирішених проблем в області перспектив кредитного скорингу для мікрофінансових організацій належать:

- оцінка інноваційного потенціалу кредитного скорингу для мікрофінансових організацій в розрізі певних етапів організації системи кредитного скорингу;
- технологічне передбачення щодо інноваційних практик кредитного скорингу у сфері мікрокредитування в розрізі деяких етапів організації системи кредитного скорингу з урахуванням сучасних популярних трендів машинного навчання, глибокого навчання та штучного інтелекту.

Мета статті. Дослідження інноваційних практик, що застосовуються в кредитному скорингу компаніями мікрокредитування, та проведення оцінки інноваційного потенціалу та перспектив методології кредитного скорингу в цілому і в розрізі етапів моделювання ідентифікованих як найбільш чутливих до впливу використання інноваційних підходів і технологій.

Основні результати дослідження. Щодо моделювання та використання систем кредитного скорингу слід виділити етап означення цільової змінної, етап моделювання на прийнятих заявках та етап аналізу та включення відхилення заяв для повторного моделювання.

Щодо етапу означення цільової змінної, то традиційним типом цільової змінної є бінарний індикаторний тип виходу на певну кількість днів прострочення заборгованості [2, 6, 7]. Потенційним альтернативним підходом можна назвати мультикласову цільову змінну, що відповідає різним ступеням ризику, і абсолютну числову змінну, що відповідає задачам класичної регресії. Остання ідея не виправдала себе з точки моделювання, однак, можна передбачити, що будуть спроби повернутися саме до числових значень пов'язаних з прибутковістю або збитковістю в якості цільової змінної, або принаймні приведення до класів прибутковості та збитковості по клієнтам. Мультикласова класифікація у даному випадку на практиці зазвичай добре вирішується

таким методом машинного навчання як випадковий ліс (random forest) у вигляді саме для мультикласової класифікації [8]. Стратегічно також можна спрогнозувати всеохоплюючий перехід до класів пожиттєвої цінності клієнта (customer lifetime value) з урахуванням утримання клієнтів і повторних конверсій, тобто на практиці повернення до більш реального проникнення маркетингу в ризик-менеджмент і навпаки [9, 10], незважаючи на технічні складнощі, що досить легко вирішуються динамічністю та короткостроковістю ринку мікрокредитування, що дозволяє спостерігати досить багато послідовних конверсій у короткостроковій і середньостроковій перспективах. Таким чином можна змоделювати певну обмежену часом апроксимацію пожиттєвої цінності клієнта з урахуванням як повторних конверсій, так і втрат не лише на етапі потенційного прострочення першої позики.

Другим важливим аспектом моделювання кредитного скорингу, що має глибокий інноваційний потенціал, є, власне, побудова первинної скорингової моделі на прийнятих заявках (на історіях виданих позик). Відносно низький ступінь внутрішньої та зовнішньої регламентованості та швидкий зворотний зв'язок показників кредитного портфелю зі змінами, що апробуються, забезпечують глибокий інноваційний потенціал щодо базових методів моделювання. Дані обставини дозволяють щонайменше апробувати всі новітні методи бінарної класифікації (в термінах класичного підходу до цільової змінної), котрі добре працюють саме на табличних даних (tabular data), враховуючи специфіку вхідних змінних кредитного скорингу. Зазвичай на практиці найсильнішими реалізаціями бінарних класифікаторів є варіації машин градієнтного бустингу такі як LightGBM та XGBoost, що мають загальну назву GBDT [11], тому мікрофінансові організації активно користуються даними інструментами для побудови найбільш точних моделей з усіх існуючих варіантів. Також одним з потужних мета-методів є ансамблевий стекинг (stacking ensemble), що по суті агрегує множину інших конкретизованих методів [12]. Технологічним передбаченням щодо аспекту моделювання на прийнятих заявках з класичним підходом щодо цільової змінної можна назвати прогнозований тренд майбутньої популярності експериментів з використанням спеціалізованих архітектур глибоких нейронних мереж таких як TabPFN, TabNet [11] у рамках процедури стекингу разом з ефективними машинами градієнтного бустингу, експериментів з використанням принципу вбудовування слів (embeddings) для категоріальних змінних, експериментів з ефективною обробкою текстових вхідних полів з використанням великих мовних моделей [12], що безумовно матиме свій вплив і на моделювання кредитного скорингу у найближчому майбутньому при наявності у компанії з мікрокредитування масивів даних з відкритими питаннями текстової аналітики, наприклад, при автоматичному отриманні і опрацюванні даних з соціальних медіа та з відкритих джерел по клієнтах. Ще одне технологічне передбачення може стосуватися вже згаданого першого аспекту моделювання пов'язаного з означенням типу цільової змінної, коли з часом питання бінарної класифікації може бути поступово витіснено питанням мультикласової класифікації або навіть регресії з використанням тих же новітніх підходів. Також слід окремо відзначити перспективу апробацій глибокого навчання з підкріпленням (reinforcement learning) [13] саме для задач кредитного скорингу, де простором дій (action space) може бути не лише тривіальне бінарне рішення щодо видачі кредиту, а власне присвоєння класу рівня ризику потенційному клієнту з яким напругу пов'язана стратегія управління кредитними ризиками, а функцією цінності може виступати як бінарний або мультиноміальний індикатор, так і абсолютні значення втрат або прибутків регресійного типу, як було зазначено вище у перспективах означення цільової. Ще одним потенційним напрямком досліджень є спроба адаптації традиційно табличних даних до даних у вигляді послідовностей з метою застосування рекурентних нейронних мереж по типу LSTM [13].

Третім, дуже важливим, науково-методологічним і практичним аспектом побудови якісних та незміщених моделей кредитного скорингу є аналіз та включення відхилених заявок у фінальну модель [2, 7, 14]. Інноваційний потенціал даного напрямку є традиційно глибоким [7] оскільки передбачає низку різних концепцій та підходів щодо вирішення проблеми зміщеності первинної моделі побудованої лише на прийнятих заявках. Новітні методи моделювання на множині прийнятих заявок розкривають великий інноваційний потенціал у напрямку аналізу та включення відхилених заявок для перебудови моделей кредитного скорингу задля забезпечення їх незміщеності. Варто відзначити, що у сучасній літературі, у тематичних форумах і на конференціях ще дуже мало висвітлені особливості застосування мета-методів аналізу та включення відхилених заявок з урахуванням типів моделей, що застосовуються для первинного

моделювання тільки на прийнятих заявках. Зазвичай при використанні новітніх методів бінарної класифікації обмежуються тільки моделями на прийнятих заявках без проведення аналізу та включення відхилених заявок, що є великим недоліком. Тим не менше, інноваційний потенціал якраз і включає опис деталізацій застосування зокрема тих же мета-методів ітеративної класифікації (iterative classification) або нечіткого доповнення (fuzzy augmentation) [7, 14] з використанням вже конкретизованих базових моделей застосованих на множині прийнятих заявок таких як реалізації машин градієнтного бустингу, реалізації випадкових лісів, реалізацій моделей глибокого навчання і т.д. З одного боку, механізми зважування елементів вибірки багатьох методів дозволяють спростити процедури включення відхилених заявок, а, з іншого боку, дискретна природа наразі найбільш ефективних методів моделювання створює також певні проблеми щодо гладкості збіжності процедур включення відхилених заявок і це потребує ретельних експериментальних досліджень і коригування підходів щодо метрик збіжності для методів, що базуються зокрема на деревах рішень. Загалом, саме включення відхилених заявок у процес моделювання має невичерпний інноваційний потенціал у плані деталізацій комбінування як мета-методів так і базових методів моделювання, а ключовими ідеями щодо вирішення проблем можна запропонувати підбір показових метрик збіжності, зважування відхилених заявок і підбір найбільш відповідних мета-методів.

Висновки та перспективи подальших досліджень. Характеристичні особливості сфери мікрокредитування такі як динамічність та дуже швидкий зворотній зв'язок щодо змін стратегій і політик управління кредитними ризиками дозволяють швидко впроваджувати та тестувати зокрема інноваційні підходи щодо побудови моделей кредитного скорингу. Дана обставина також разом з низькою зовнішньою та внутрішньою регульованістю дозволяє саме мікрофінансовим організаціям ефективно і неперервно синхронізуватися з останніми досягненнями в області машинного навчання, глибокого навчання та штучного інтелекту, тому, відповідно, інноваційний потенціал кредитного скорингу тісно пов'язаний з розвитком перелічених областей як показано у даному дослідженні. Високоєфективні методи моделювання дозволяють краще ідентифікувати потенційних клієнтів з високим рівнем ризику, що дозволяє мінімізувати втрати кращим чином, і, відповідно, покращити фінансові результати роботи компанії.

Перспективи подальших досліджень включають перегляд здійснення наведених технологічних передбачень через певний проміжок часу, а також пропозиції по інтеграції наведених інноваційних можливостей методів машинного навчання в процес вирішення спеціалізованих задач кредитного скорингу. Зокрема, однією з найбільш перспективних та досі мало висвітлених тем є аналіз відхилених заявок та їх включення у фінальну модель (reject inference) саме для сучасних ефективних методів машинного навчання по типу реалізацій машин градієнтного бустингу. Тому перспективною є розробка відповідних інтеграцій та їх апробація на реальних даних у порівнянні з класичними моделями.

ЛІТЕРАТУРА

1. Терентьев О.М. Модели і методи побудови та аналізу байєсівських мереж для інтелектуального аналізу даних: дис. кандидата техн. наук: 05.13.06 / Терентьев Олександр Миколайович. — К., 2009. — 258 с.
2. Солошенко О.М. Модели і методи оцінювання кредитоспроможності фізичних осіб: дис. кандидата техн. наук: 01.05.04 / Солошенко Олександр Миколайович. — К., 2016. — 180 с.
3. Камінський А.Б. Характеристичні особливості ризик-менеджменту в сегменті мікрокредитування / А.Б. Камінський // Наукові записки НаУКМА. Економічні науки. — 2016. — Т. 1, № 1. — С. 80–85.
4. Згуровський М.З. Основи системного аналізу: підручник [для студ. вищ. навч. закл., які навч. за напрямками «Системний аналіз», «Прикладна математика», «Інформатика», «Комп'ютерні науки», «Комп'ютерна інженерія», «Системна інженерія»] / М.З. Згуровський, Н.Д. Панкратова. — К.: Видавнична група «BVH», 2007. — 544 с.: іл.
5. Stegman M. Payday Lending / M. Stegman // The Journal of Economic Perspectives. — 2007. — Vol. 21, № 1. — Pp. 169–190.
6. Thomas L.C. Credit Scoring and its applications: Monograph / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. — Philadelphia: SIAM, 2002. — 248 p.
7. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring / Naeem Siddiqi. — Hoboken: John Wiley & Sons, Inc., 2006. — 196 p.

8. Prinzie A. Random Forests for multiclass classification: Random MultiNomial Logit / A. Prinzie, D. Van den Poel // *Expert Systems with Applications*. — 2008. — Vol. 34, № 3. — Pp. 1721–1732.
9. Ryals L. Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value / Lynette Ryals, Simon Knox // *European Journal of Marketing*. — 2005. — № 39. — Pp. 456–472.
10. Kostiv M. Customer Lifetime Value For Credit Limit Optimization: Master Thesis / Markiyan Kostiv. — Lviv, 2019. — 23 p.: il.
11. When Do Neural Nets Outperform Boosted Trees on Tabular Data? [Електронний ресурс] / [McElfresh Duncan, Khandagale Sujay, Valverde Jonathan et al.]. — Ithaca: Cornell University, 2024. — 39 p. — Режим доступу: <https://arxiv.org/pdf/2305.02997>.
12. Gaius A. A Stacking-Based Ensemble Approach with Embeddings from Language Models for Depression Detection from Social Media Text / Akwa Gaius, Ronald Waweru Mwangi, Antony Ngunyi // *Journal of Data Analysis and Information Processing*. — 2023. — № 11. — Pp. 420–453.
13. Ravichandiran S. Hands-On Reinforcement Learning with Python: Master Reinforcement and Deep Reinforcement Learning Using OpenAI Gym and TensorFlow / Sudharsan Ravichandiran. — Birmingham: Packt Publishing, 2018. — 318 p.
14. Солошенко О.М. Вдосконалення методу ітеративної класифікації з включення відхилення заявок у кредитному скорингу / О.М. Солошенко // *Наук. вісті НТУУ «КПІ»*. — 2014. — № 5. — С. 63–69.

REFERENCES

1. Terentyev O. Models and methods of construction and analysis of the Bayesian networks for the intellectual data analysis: Thesis in fulfillment of the requirements for the degree of candidate of engineering sciences: 05.13.06 / Terentyev Oleksandr. — Kyiv, 2009. — 258 p. [in Ukrainian].
2. Soloshenko O. Models and methods for creditworthiness assessment of private individuals: A dissertation submitted in fulfillment of the requirements for the degree of candidate of engineering sciences: 01.05.04 / Soloshenko Oleksandr. — Kyiv, 2016. — 180 p. [in Ukrainian].
3. Kaminskyi A. Characteristic singularities of risk management in the microcrediting segment / A. Kaminskyi // *Scientific Papers NaUKMA. Economics*. — 2016. — Vol. 1, № 1. — Pp. 80–85 [in Ukrainian].
4. Zgurovsky M. Foundation of system analysis: textbook [for students of higher education institutions specializing in the fields of «System Analysis», «Applied Mathematics», «Informatics», «Computer Science», «Computer Engineering», «System Engineering»] / M. Zgurovsky, N. Pankratova. — Kyiv: «BHV» Publishing Group, 2007. — 544 p.: il. [in Ukrainian].
5. Stegman M. Payday Lending / M. Stegman // *The Journal of Economic Perspectives*. — 2007. — Vol. 21, № 1. — Pp. 169–190.
6. Thomas L.C. Credit Scoring and its applications: Monograph / Lyn C. Thomas, David B. Edelman, Jonathan N. Crook. — Philadelphia: SIAM, 2002. — 248 p.
7. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring / Naeem Siddiqi. — Hoboken: John Wiley & Sons, Inc., 2006. — 196 p.
8. Prinzie A. Random Forests for multiclass classification: Random MultiNomial Logit / A. Prinzie, D. Van den Poel // *Expert Systems with Applications*. — 2008. — Vol. 34, № 3. — Pp. 1721–1732.
9. Ryals L. Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value / Lynette Ryals, Simon Knox // *European Journal of Marketing*. — 2005. — № 39. — Pp. 456–472.
10. Kostiv M. Customer Lifetime Value For Credit Limit Optimization: Master Thesis / Markiyan Kostiv. — Lviv, 2019. — 23 p.: il.
11. When Do Neural Nets Outperform Boosted Trees on Tabular Data? [Electronic source] / [McElfresh Duncan, Khandagale Sujay, Valverde Jonathan et al.]. — Ithaca: Cornell University, 2024. — 39 p. — Access mode: <https://arxiv.org/pdf/2305.02997>.
12. Gaius A. A Stacking-Based Ensemble Approach with Embeddings from Language Models for Depression Detection from Social Media Text / Akwa Gaius, Ronald Waweru Mwangi, Antony Ngunyi // *Journal of Data Analysis and Information Processing*. — 2023. — № 11. — Pp. 420–453.

13. Ravichandiran S. Hands-On Reinforcement Learning with Python: Master Reinforcement and Deep Reinforcement Learning Using OpenAI Gym and TensorFlow / Sudharsan Ravichandiran. — Birmingham: Packt Publishing, 2018. — 318 p.

14. Soloshenko O. Adjustment of the Iterative Reclassification Method for Including the Rejected Applications into the Credit Scoring / O. Soloshenko // Research Bulletin of the National Technical University of Ukraine «Kyiv Polytechnic Institute». — 2014. — № 5. — Pp. 63–69 [in Ukrainian].

Soloshenko O. M. INNOVATIVE PRACTICES AND INNOVATIVE POTENTIAL OF THE CREDIT SCORING METHODOLOGY OF MICROCREDIT COMPANIES

The objective of the research is to discover the modern innovative practices used by microfinance companies for credit scoring modeling and to make an appropriate assessment of the future perspectives and innovation potential of credit scoring methodologies in general.

The results of the research and the forecasts of the innovation trends of credit scoring methodologies for microfinance companies are obtained using the following methods: analysis of the separate components of the modern innovative processes of credit scoring modeling, synthesis of the separate development tendencies with following system generalization to high-level dependencies and trends, and system method of technology foresight as for the future perspectives of innovative development of credit scoring methodologies for microfinance companies.

Based on the modern trends of credit scoring methodologies for payday lending industry, it was proven that integration of the modern machine learning contributions with credit scoring methodologies has a significant impact on operation of microfinance companies and on corresponding financial results as well.

The key stages of building credit scorecards which are related with payday lending specific features and which are influenced by the largest innovation and technology development impact are successfully identified also identifying the corresponding tendencies. That means that three credit risk modeling stages are highlighted having the largest contemporary innovation potential:

1) modeled target variable definition;

2) building the primary credit scoring model based on approved applications only (complete data only);

3) building the final model using reject inference approach to additionally include rejected applications for making the final model unbiased (including incomplete data).

Expected development trends of credit risk modeling methodologies are forecasted for payday lending industry. The related contemporary problematic issues are also identified resulting from usage of innovative binary classification methods taken from machine learning contributions. The key ideas for solving the reject inference issues for the final model are proposed for the use-cases for the modern machine learning methods with a discrete background and these ideas reflect one of the key elements of the prospects for further research.

Key words: credit scoring, credit risk management, payday lending risk, machine learning.