

УДК 004.42:336.77

КРЕДИТНИЙ СКОРИНГ ДЛЯ РОЗДРІБНИХ ПОЗИК БАНКУ

Максим Бездітко

maksym.bezditko@nure.ua

Ст. викл кафедри ПІ Гліб Терещенко

Харківський національний університет радіоелектроніки каф. ПІ
м. Харків, Україна

This paper describes a credit scoring system for retail bank lending. The system predicts the probability of borrower default within twelve months. It supports loan approval, rejection, manual review, and credit limit recommendation. The solution includes a Python backend, machine learning models and the React frontend. Historical data are processed through cleaning, imputation, encoding, and scaling. Default prediction is based on logistic regression, random forest, and boosting methods. The approach improves the speed, consistency, and transparency of credit risk assessment.

Кредитний скоринг – це метод класифікації позичальників для оцінки їх кредитоспроможності та рівня кредитного ризику. Проєктована система повертатиме ймовірність дефолту протягом дванадцяти місяців і рекомендуватиме рішення щодо видачі кредиту та встановлення кредитного ліміту. Бал скорингу безпосередньо впливає на рішення банку про видачу або відмову у кредиті [1]. Наприклад, згідно зі світовими практиками, application score формує основу для рішення: якщо значення перевищує певний поріг, кредит схвалюють, інакше – відхиляють.

Система реалізується на сучасних технологіях: серверна частина створюється на Python з використанням бібліотек для обробки даних і машинного навчання. Для побудови скорингової моделі доцільно застосовувати логістичну регресію, ансамблеві методи, зокрема Random Forest, а також бустингові моделі, оскільки вони забезпечують високу точність прогнозу дефолту [2]. Сервісний інтерфейс реалізується на FastAPI, що дає змогу зручно розгортати модель у прикладній системі. Клієнтська частина на NodeJS з React забезпечує зручний вебінтерфейс для кредитних спеціалістів. Для збереження вхідних даних заявок і результатів скорингу використовується реляційна база даних PostgreSQL.

Для аналізу даних і побудови моделей застосовуються популярні бібліотеки. Програмний інтерфейс сервісу забезпечує розрахунок ймовірності дефолту за однією заявкою. Клієнтська частина призначена для введення даних, надсилання запиту до сервера та відображення отриманого результату. Система керування базами даних зберігає дві основні таблиці – вхідні параметри заявок і результати скорингу, зокрема ймовірність дефолту, скоринговий бал і фінальне рішення.

Першим етапом є модуль підготовки даних. Історичні дані клієнтів завантажуються з файлів CSV і проходять попередню обробку: заповнення

пропусків, видалення аномальних значень, кодування категоріальних ознак і масштабування числових показників. Для цього використовується послідовний конвеєр обробки даних, який починається із завантаження сирих записів, оцінювання їх якості та виконання етапів очищення й перетворення. Після попередньої обробки дані поділяються на навчальну і тестову вибірки, наприклад у співвідношенні 70:30, а потім зберігаються для повторного використання.

Основними кроками цього етапу є завантаження історії позичальників із CSV, очищення даних, заповнення пропусків, вилучення аномалій, кодування категоріальних змінних, масштабування числових ознак, поділ на навчальну та тестову вибірки і збереження підготовлених наборів. Як приклад готового набору даних можна використати Statlog German Credit Data, який містить класифікацію позичальників на надійних і проблемних.

Після підготовки даних виконується навчання моделі прогнозування дефолту. Логістична регресія використовується як базовий варіант завдяки простоті й зрозумілості результатів, тоді як сучасні методи машинного навчання, наприклад XGBoost або LightGBM, часто забезпечують вищу точність. Хоча прості статистичні моделі й надалі широко застосовуються у скорингових задачах, сучасні методи дають змогу виявляти складніші залежності між ознаками [3].

Етап побудови моделі включає навчання логістичної регресії як початкового варіанта, побудову ансамблевих і бустингових моделей, добір гіперпараметрів за допомогою перехресної перевірки, оцінювання моделей за метриками ROC AUC, коефіцієнтом Джині та показниками калібрування ймовірностей [4, 5]. Після цього обирається найкраща модель, зберігається у файл і формується короткий звіт з основними показниками якості. Такий підхід дає змогу підвищити надійність скорингової моделі та відповідає сучасним практикам у сфері оцінювання кредитного ризику.

Серверна частина реалізує програмний інтерфейс, який під час надходження однієї заявки обчислює ймовірність дефолту та скоринговий бал. Бізнес-логіка системи ґрунтується на порогових значеннях скорингового балу. Якщо бал вищий за верхній поріг, система повертає рішення про видачу кредиту, якщо нижчий за нижній поріг – відмову, а проміжні значення передаються на ручний розгляд. Додатково система обчислює рекомендований кредитний ліміт залежно від скорингового балу та фінансових характеристик клієнта. У відповіді користувач отримує структурований результат із коротким поясненням.

Клієнтська частина на React надає кредитному працівнику зручний інтерфейс для роботи із заявками. Форма введення нової заявки містить усі необхідні поля з перевіркою правильності введення. Після надсилання даних на сервер користувач бачить результат скорингу – ймовірність

дефолту, скоринговий бал, рекомендоване рішення і кредитний ліміт. Також реалізується таблиця історії заявок із можливістю фільтрації та сортування за датою або рішенням. Окрема сторінка відображає основні показники якості моделі, зокрема ROC AUC, коефіцієнт Джині та розподіли балів, що дає змогу оперативно контролювати якість скорингової системи.

Запропонована система кредитного скорингу забезпечує повний цикл обробки – від підготовки даних і побудови моделі до інтеграції з банківськими сервісами. Використання сучасних методів машинного навчання підвищує точність прогнозування дефолту, а автоматизація скорочує час ухвалення рішення та знижує ризик неповернення кредитів. Водночас система забезпечує прозорість процесу, оскільки результати журналюються, правила ухвалення рішень задаються наперед, а історія всіх розрахунків зберігається для подальшого аналізу.

Список використаних джерел:

1. Кредитний скоринг // Вікіпедія : вільна енциклопедія. URL: https://uk.wikipedia.org/wiki/Кредитний_скоринг (дата звернення: 22.03.2026).
2. Rida A. Machine and Deep Learning for Credit Scoring: A compliant approach. 2019. URL: <https://arxiv.org/pdf/2412.20225> (дата звернення: 22.03.2026).
3. Goriunov D., Venzhyk K. Loan Default Prediction in Ukrainian Retail Banking. EERC Working Paper. 2013. № 13/07e. URL: <https://ideas.repec.org/p/eer/wpalle/13-07e.html> (дата звернення: 22.03.2026).
4. Credit Scoring Approaches Guidelines. Washington DC : The World Bank, 2020. URL: <https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf> (дата звернення: 22.03.2026).
5. Kyrychenko I., Tereshchenko G., Storage of Big Data in Intelligent Systems: monograph 2026 URL: <https://publish.nure.ua/catalog/book/1187> (дата звернення: 22.03.2026).