

УДК 004.85:[336.77:336.717.5]

**ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ
ДЛЯ КРЕДИТНОГО СКОРІНГУ У ВИПАДКУ
КОРОТКОСТРОКОВИХ ПОЗИК**

Троян Н. А.

Науковий керівник – д-р техн. наук, проф. Кіріченко Л.О.
Харківський національний університет радіоелектроніки, каф. ПМ
м. Харків, Україна

тел. +38 (057) 702-14-36, email: nadiia.troian@nure.ua

The goal of the study is to determine a client's default risk level in the short term. This indicator, in turn, determines what kinds of loans the client can get. In this work, two machine learning methods, logistic regression and random forest, were used to process credit-based data and determine a client's default risk level in the short term. Logistic regression and random forest methods were chosen due to their effectiveness in handling large datasets and identifying relevant variables. The results obtained from the research can be used in banks and financial companies to improve the decision-making process and the efficiency of lending.

Кредитний скоринг – це важлива задача для банків та фінансових установ, яка дозволяє оцінити ризики невиконання кредитів. Кредитний скоринг базується на аналізі історії кредитних операцій клієнта. До уваги беруться такі фактори, як терміни повернення попередніх позик, їх кількість, зміна місця роботи, наявність невчасно виконаних платежів, кількість утриманців тощо. За допомогою цих показників можна обчислити рівень ризику, який пов'язаний з кредитуванням конкретного клієнта. Чим вище рівень ризику ми отримуємо, тим нижче буде ймовірність отримання банком назад наданих в кредит коштів. Рівень ризику може слугувати для визначення відсоткової ставки, встановлення граничної суми кредиту, строку повернення і інших умов кредитування для кожного окремого взятого клієнта.

Видача кредитів завжди пов'язана з ризиками, наприклад, із ризиком неповернення кредиту одержувачем. Отже, метою кредитного скорингу у випадку короткострокових позик є забезпечення фінансової стабільності банку та підвищення ефективності роботи банку шляхом зменшення ризику невиконання позики. Якісна робота з обчисленням кредитного скорингу допомагає банку надавати послугу усім типам клієнтів із максимальною користю для бізнесу, але задача аналізу великого обсягу даних часто пов'язана із витратою великої кількості часу та людського ресурсу. Для ефективного вирішення завдання кредитного скорингу в теперішній час використовуються методи машинного навчання. Ці методи дають змогу автоматично витягувати інформацію з великих обсягів даних, а також будувати моделі, які здатні класифікувати позичальників на основі різних

критеріїв, такі як кредитна історія, доходи тощо. Завдяки машинному навчанню можливо знизити ризики, пов'язані з видачою кредитів, і збільшити точність прогнозування неповернення кредиту [1].

В роботі розглянуто набір даних для дослідження «Give me some credit» з платформи Kaggle [2]. В ньому містяться дані про загальний баланс за кредитними картками та особистими кредитними лініями, кількість разів, коли позичальник прострочував оплату, реальний місячний дохід, кількість відкритих позик, кількість утриманців в сім'ї та подібна інформація, на основі якої можна зробити висновок про репутацію особи в найближчому майбутньому. Для кредитного скорінгу застосовуються різноманітні алгоритми машинного навчання, такі як логістична регресія, дерева рішень, випадковий ліс, нейронні мережі та інші. Метод логістичної регресії широко використовується для вирішення задач класифікації, зокрема для кредитного скорінгу. Це обумовлено тим, що логістична регресія може працювати з багатофакторними даними, що дозволяє враховувати багато параметрів кредитної історії під час прийняття рішення щодо кредитування та його умов. Цей метод може бути добре відкалібрований та показувати точні результати, що дозволяє віднести кредитних клієнтів до категорій з різними рівнями ризику та встановити потрібні кредитні лінії для кожної категорії.

Метод «випадковий ліс» об'єднує багато дерев рішень та зводить до мінімуму ефект перенавчання, може обробляти велику кількість факторів та не потребує попередньої обробки даних, що робить його ефективним для аналізу великих обсягів даних. Автоматичне визначення важливості факторів та врахування їх у моделі дозволяє виявити взаємозв'язки між різними факторами та їх вплив на рішення про надання кредиту. Головним завданням роботи є побудова моделі машинного навчання двома методами: логістичної регресії та випадкового лісу; підбір параметрів для прогнозування ймовірності того, що клієнт зазнає фінансових труднощів протягом наступних двох років і не зможе виконати боргові зобов'язання перед позичальником.

Список використаних джерел:

1. Sousa, M. de M. & Figueiredo. Credit Analysis Using Data Mining: Application in the Case of a Credit Union. *Journal of Information Systems and Technology Management*, 11 (2), 379–396.

2. *Give Me Some Credit*: 2011 Competition Data
<https://www.kaggle.com/datasets/brycecf/give-me-some-credit-dataset>