

Проблема зміщення співвідношення класів при встановленні порогу відсікання для бінарного класифікатора у задачах кредитного скорингу

Солошенко О.М., *аспірант*; Бідюк П.І., *професор*
Національний технічний університет України
«Київський політехнічний інститут», м. Київ

Метою дослідження є формалізація та наведення формули розрахунку зміщеної долі елементів вибірки нульового класу (bad rate) при встановленні порогу відсікання (cut-off rate) за допомогою статистики Колмогорова-Смирнова [1]. Актуальність дослідження полягає у широкому застосуванні бінарних класифікаторів у задачах кредитного скорингу, зокрема у задачах аналізу відхилених заявок (reject inference) на отримання кредиту [2, 3] з метою включення їх до навчальної вибірки моделі.

Класична формула обчислення статистики Колмогорова-Смирнова має вигляд [1, 2]:

$$K = \max |F(x | y = 0) - F(x | y = 1)|, \quad (1)$$

де K – статистика Колмогорова-Смирнова, F – кумулятивна функція умовного розподілу, x – прогнозована ймовірність належності до одиничного класу (тому $x \in [0; 1]$), y – цільова змінна.

Формула оптимального порогу відсікання (cut-off rate) [1]:

$$c = \operatorname{argmax} |F(x | y = 0) - F(x | y = 1)|, \quad (2)$$

де c – оптимальний поріг відсікання.

Пропонована формула зміщеної ймовірності у термінах (1) та (2):

$$p(y' = 0) = p(y = 0) F(c | y = 0) + (1 - p(y = 0)) F(c | y = 1),$$

де $p(y = 0)$ – початкова доля нульових значень, $p(y' = 0)$ – кінцева.

1. Э. Мэйз, *Руководство по кредитному скорингу* (Минск: Гревцов Паблишер: 2008).
2. Н. Сиддики, *Скоринговые карты для оценки кредитных рисков* (Москва: Манн, Иванов и Фербер: 2014).
3. О.М. Солошенко, *Наук. вісті НТУУ «КПІ»* 5, 63 (2014).