

УДК 004.7

*Литвинець Б.В., здобувач  
Шушура О.М., д.т.н., професор*

*Державний університет інформаційно-комунікаційних технологій*

## **МІНІМІЗАЦІЯ ВПЛИВУ ШАХРАЙСТВА ТА ПРОТИДІЯ ЗАШУМЛЕННЮ ДАНИХ У СИСТЕМАХ КРЕДИТНОГО СКОРИНГУ**

Ефективність кредитних скорингових систем багато в чому залежить від якості історичних даних. Швидкий перехід банківського сектору на цифрові технології призвів до збільшення кількості транзакцій, але водночас створив більше можливостей для шахраїв, що викликає проблему зростання шахрайських операцій. Такі випадки вносять шум у дані, що негативно впливає на результати роботи моделі.

Якщо шахрай отримує кредит, використовуючи синтетичні дані, та не повертає кошти, класична модель позначає це як низьку платоспроможність. В результаті алгоритм помилково сприймає поведінку шахрая як ознаки “поганого, але реального позичальника”. Через це відбувається зміщення концепту (Data Drift), що порушує логіку даних і знижує точність прогнозів алгоритмів.

Основна проблема — сильно дисбалансовані класи. Шахрайські операції складають дуже малу частку від усіх транзакцій, часто менше за 0.1%. Якщо подати такий неврівноважений набір даних звичному алгоритму, він ігнорує цю меншість і показує високу загальну точність, але при цьому пропускає більшість шахраїв (високий рівень хибнонегативних результатів).

Щоб алгоритм зміг краще знаходити шахрайські операції, потрібен агресивний підхід до підготовки даних. Ефективно штучно збільшувати кількість прикладів шахрайства, одночасно зменшуючи легітимні транзакції. Методи SMOTE та ADASYN допомагають створювати більше прикладів шахрайства, а Undersampling скорочує кількість нормальних транзакцій.

Щодо вибору моделей, найбільшу стійкість до шуму показують дерева рішень. Особливо градієнтний бустинг, який здатен виявляти складні та приховані зв'язки у поведінці клієнтів. Проте таку модель потрібно ретельно налаштовувати з урахуванням метрик, які добре працюють при дисбалансі даних: ROC-кривої, AUC, Precision-Recall AUC та F-міри.

Для роботи та впровадження таких моделей важлива надійна інфраструктура. Системи типу ClickHouse дозволяють виконувати

складні агрегатні SQL-запити в режимі реального часу. Їх інтеграція із візуалізаційними платформами допомагає моніторити стабільність системи та швидко реагувати на аномалії, що дуже важливо.

Отже, щоб ефективно мінімізувати вплив шахрайства, потрібно поєднувати просунуті методи роботи з незбалансованими даними, сучасні алгоритми та надійну інфраструктуру. Лише такий підхід дозволяє відокремити реальні кредитні ризики від шахрайських загроз і забезпечити стабільність системи.

#### **Список використаних джерел**

1. Agrwal C. C. Outlier Analysis . 2nd ed.2017.
2. Synthetic Identify Fraud in the U.S Payment System A Review of Causes and Contributing Factors.2019.