**ПЕРЕЛІК ПОСИЛАНЬ**

1. Корнієнко В.І. Ідентифікація і прогнозування трафіку в телекомунікаційних системах / В.І. Корнієнко, Л.В. Будкова // Системи обробки інформації – 2011. – № 8(98). – С. 208-211. – ISSN 1681-7710.

2. Будкова Л.В. Моделювання самоподібного трафіку в інформаційних телекомунікаційних мережах / Л.В. Будкова, В.І. Корнієнко // Вісник Кременчуцького національного університету імені Михайла Остроградського – 2013. – № 4(81). – С. 101-108.

3. Беллами Дж. Цифровая телефония: пер. с англ. / Дж. Беллами. – М.: Эко-Тренд, 2004. – 640 с.

УДК 519.876.5

Т.А. Желдак[1], Ю.С. Соловей[1]
[1]Національний технічний університет «Дніпровська політехніка», Дніпро, Україна

# БІЗНЕС-АНАЛІЗ ДОДАТКА ДЛЯ ІДЕНТИФІКАЦІЇ КІБЕРЗЛОЧИНІВ У БАНКІВСЬКІЙ СФЕРІ

**Анотація.** Описано процес розробки алгоритму класифікації транзакції з виявленням підозрілих на шахрайство, ефективнішого за існуючі. Підвищення ефективності досягається шляхом використання сучасних алгоритмів автоматизованої класифікації великих даних та автоматизованих методів обробки вимог до програмного забезпечення.

***Ключові слова**: банківські картки, транзакції, класифікація, шахрайство, точність, навчання, бустінг, класифікатор, додаток, ефективність, вимоги, проектування.*

**Вступ.** Кіберпростір створює неймовірні можливості, розширює свободу, стимулює розвиток інновацій та збагачує суспільство. Однак, паралельно з позитивними тенденціями, набуває розвитку і кіберзлочинність, що, зрозуміло, завдає значної шкоди інтересам наших громадян та держави в цілому. З метою протидії такій кримінальній протиправності в Національній поліції функціонує підрозділ кіберполіції.

Дана робота присвячена дослідженню та оптимізації діяльності управління протидії кіберзлочинам в Дніпропетровській області, яка в свою чергу входить до складу апарату центрального органу управління поліції, а також впровадженню новітніх технологій для щоденного використання в розслідуванні злочинів та запобіганні шахрайських дій в банківському секторі.

Метою роботи установи є захист населення від кіберзлочинів, запобігання неправомірних дій та розслідування і викриття вже скоєних злочинів. Основні задачі даного відділку досить широкий: поліцїанти займаються боротьбою з вірусами, DdoS-атаками, спамом, шахрайством з банківськими системами і крадіжкою особистих даних.

Об'єктом дослідження є правозахисна діяльність управління протидії кіберзлочинам в Дніпропетровській області ДКП НПУ, а саме виявлення неправомірних дій в банківській сфері.

Предметом дослідження є використання методів бізнес-аналізу для розробки додатка ідентифікації кіберзлочинів у банківській сфері управлінням протидії кіберзлочинам в Дніпропетровській області ДКП НПУ.

**Метою роботи** є підвищення ефективності роботи кіберполіціантів при розпізнаванні шахрайських дій та подальшого пошуку злочинців, можливого прогнозування неправомірних дій в банківських системах за допомогою розробленої Antifraud detection-системи.

**Постановка задачі.** Для досягнення поставленої мети в роботі визначені та вирішені наступні задачі дослідження:

- дослідити методи, за допомогою яких можна побудувати класифікатор банківських транзакцій;
- розробка та побудова класифікатора на незбалансованих даних;
- аналіз отриманих результатів та вибір оптимального методу;
- проведення бізнес-аналізу додатка для ідентифікації кіберзлочинів у банківській сфері, а саме збір та обробка вимог до додатка, розробка логічної схеми БД, моделювання елементів алгоритму роботи, а також результати застосування та оцінка ефективності додатка.

**Основний зміст роботи.** Для розв'язання поставлених задач в роботі запропоновано використовувати наступні методи дослідження:

- методи Undersampling та Oversampling для збалансування незбалансованої вибірки [1];
- методи XGBoost, AdaBoost та градієнтного бустингу та ансамблі відповідних моделей для побудови класифікатора банківських транзакцій з метою виявлення шахрайської діяльності [2];
- різноманітні техніки виконання бізнес-аналізу додатка, тощо.

Авторами була розроблена програмна реалізація машинного навчання, а саме покрокова реалізація алгоритмів XGBoost, AdaBoost, градієнтного бустингу та ансамблів відповідних алгоритмів [3]. В якості вхідних даних було взято вибірку з відкритого джерела, так як доступ до персональних даних, скачування і подальше використання і вивчення для осіб, що не мають спеціального дозволу, заборонено чинним законодавством. Набір даних містить транзакції, здійснені кредитними картками у вересні 2013 року українськими власниками карток, що відбулися за два дні. В ході аналізу вхідних даних було виявлено, що між даними існує великий дисбаланс, адже кількість шахрайських транзакцій сягала менше 1% від усієї вибірки.

Було проведено чотири експериментальних навчань за різних умов, задля виявлення оптимальних значень. В результаті було отримано наступні результати:

Експеримент №1. Побудова класифікатора на незбалансованій вибірці:

- Метод градієнтного бустингу: TN = 94 778, FP = 36, FN = 45, TP = 77;
- Метод XGBoost: TN = 94 810, FP = 4, FN = 37, TP = 85;
- Метод AdaBoost: TN = 94 802, FP = 12, FN = 38, TP = 84.

Експеримент №2. Побудова класифікатора на вибірці, поділеній на 5 рівних частин, з метою зменшення дисбалансу в вибірці :

- Метод XGBoost: TN = 94 756, FP = 58, FN = 28, TP = 94, а оцінки точності моделі складають : recall = 0.77, precision = 0.62, accuracy = 1.00;
- Метод AdaBoost: TN = 94 380, FP = 434, FN = 27, TP = 95 а оцінки точності моделі складають : recall = 0.81, precision = 0.01, accuracy = 0.92.

Експеримент №3. Побудова класифікатора на збалансованій вибірці методом Undersampling (NNK):

- Метод XGBoost: TN = 93 640, FP = 1174, FN = 24, TP = 98, а оцінки точності моделі складають : recall = 0.80, precision = 0.08, accuracy = 0.99;
- Метод AdaBoost: TN = 80 423, FP = 14 391, FN = 15, TP = 107 а оцінки точності моделі складають : recall = 0.88, precision = 0.01, accuracy = 0.85.

Експеримент №4. Побудова класифікатора на збалансованій вибірці методом Oversampling (Random):

- Метод XGBoost: TN = 94 656, FP = 158, FN = 23, TP = 99, а оцінки точності моделі складають : recall = 0.81, precision = 0.39, accuracy = 1.00;
- Метод AdaBoost: TN = 92 825, FP = 1 989, FN = 23, TP = 99 а оцінки точності моделі складають : recall = 0.81, precision = 0.05, accuracy = 0.98.

Аналізуючи результати машинного навчання варто відзначити, що метод градієнтного бустингу виявися не оптимальним, адже давав низькі за точністю результати і час виконання навчання був дуже довгим в порівнянні з іншими двома методами, тому його було виключено з подальших розрахунків. Також в ході реалізації двох методів було застосовано методи семплінгу, що посприяли покращенню результатів машинного навчання. Спираючись на результати експериментів можна вважати експеримент № 3 і №4 результативними, проте вибір конкретного методу бустингу і семплінгу може варіюватися.

Алгоритм XGBoost має більш вищий показник точності класифікатора (більш точно відокремлює «хороші» транзакції), а також має більш швидкий час виконання алгоритму навчання, а у алгоритму AdaBoost кращий показник глибини класифікатора (більш точно відокремлює «погані» транзакції), зумовлено тим, що алгоритм AdaBoost навчається на «поганих» результатах попереднього навчання. Тобто з кожним кроком все зменшується кількість поганих результатів навчання. Проте за рахунок цього достатньо велика кількість, можливо хороших транзакцій потрапляють до «підозрілих».

В такому випадку можна відштовхуватися від побажань замовника подібних класифікаторів або ж доречно використовувати ці два алгоритми разом. Вони, так би мовити, будуть збалансовувати результати навчання одне одного.

В нашому випадку замовником є державна правоохоронна установа, метою якої є виявити якомога більше підозрілих транзакцій, так як вони в подальшому можуть складати більшу загрозу. Тому можна вважати, що в даному випадку оцінка глибини є більш опорною, ніж точність моделі.

В практичному розділі було проведено бізнес-аналіз додатку протидії кіберзлочинам в банківській сфері. Тобто це можна вважати вже цілісним

проектом по впровадженню Antifraud detection-системи, адже розобки самого функціоналу недостатньо. Для ефективного його використання необхідно розробити зручний функціонал, який буде відповідати усім вимогам користувача, та буде містити увесь необхідний функціонал.

Особливості побудованої Antifraud detection-системи:

- Обробка подій у реальному часі.
- Система відповідає найжорсткішим вимогам до швидкості – тисячі подій обробляються в режимі real-time.
- Максимум виявленого шахрайства при мінімумі хибних спрацьовувань.
- Одночасне застосування експертних правил та самонавчальних моделей.
- Відкрита технологія керування для коригування політик та моделей.
- Рішення побудоване на відкритій технології управління.
- Зручний інтерфейс для швидкого прийняття рішень.
- Можливість створення та наповнення даними різних бізнес-об'єктів.

Усе це стало можливим завдяки методам бізнес-аналізу [4], що дозволили виконати збір необхідних вимог до додатку шляхом проведення інтерв'ю користувачів та опитувань. Також вагому роль відіграло визначення потенційних стейкхолдерів (зацікавлених осіб) продукту, адже це дозволило побудувати всебічну модель додатку, яка враховує усі вимоги, очікування та інтереси та протистояти обмеженням, які були визначені у Stakeholder map.

Також важливим кроком на етапі розробки додатку було побудова інформаційної архітектури додатку, а також відтворення user flow [5], що допомогло зрозуміти чи відбулися покращення у шляху юзера по досягненні однієї і тієї ж цілі в порівнянні з шляхом до впровадження додатку.

Практична цінність отриманих результатів полягає у отриманні готового додатку класифікатора банківських транзакцій, який дозволятиме визначати до якого класу відноситься та чи інша транзакція, тобто чи є вона «хороша», або «погана», блокувати їх, а також з використанням внутрішніх та зовнішніх банків даних та зручного функціоналу відшукувати злочинців, швидко отримувати необхідні дані про них та попереджати подальші неправомірні злочини. Тестування додатку вже за три місяці показало, що раніше в середньому на збір необхідної інформації по справі йшло 9-10 днів, наразі це займає чотири дні, це з урахуванням того факту, що зазвичай шахраї не обходяться одним злочином, а мають велику кількість жертв, і тому саме збір доказів по всім випадкам подовжує час обробки та збору інформації по справі.

Економічний ефект від отриманих результатів очікується позитивним завдяки тому, що зменшується щорічний збиток від фінансового шахрайства в сфері онлайн платежів: впровадження на три місяці готового функціоналу допомогло нам заощадити майже 40 млн. грн у 2022 році і становить 151 млн. грн, а прогнозоване значення на 2023 рік становить 99 млн грн, що на 35% менше ніж у 2022 році та на майже 50% менше ніж у 2020 році.

Соціальний ефект від результатів роботи очікується позитивним через скорочення кількості злочинів в фінансовій сфері, захист населення від

неправомірних транзакцій, тобто призводить до підвищення довіри населення до органів Національної поліції України. Складність впровадження такої системи полягає у тому, що необхідно залучати команду спеціалістів, що мають реалізувати впровадження і забезпечити подальше супроводження продукту, проводити аналізування результатів класифікатора, задля виявлення поламок функціоналу і швидкого їх усунення, адже така система потребує знань та навичок в багатох сферах інформаційних технологій, якими не може володіти кіберполіціянт. З іншого ж підтримка з боку держави, яке є одним зі стейкхолдерів та залучення інших меценатів мають допомогти в зборі необхідних коштів та персоналу, адже це значно покращуватиме положення захисту населення в банківській сфері і підвищуватиме довіру до НПУ та влади.

**Наукова новизна** розробки полягає у застосуванні сучасних алгоритмів класифікації транзакцій для підвищення ефективності роботи кіберполіціантів при розпізнаванні шахрайських дій та подальшого пошуку злочинців, можливого прогнозування неправомірних дій в банківських системах.

**Висновки.** В результаті було у отримано готовий додаток класифікатора банківських транзакцій, який дозволяє визначати до якого класу відноситься та чи інша транзакція, тобто чи є вона «хороша», або «погана», блокувати їх, а також з використанням внутрішніх та зовнішніх банків даних та зручного функціоналу відшукувати злочинців, швидко отримувати необхідні дані про них та попереджати подальші неправомірні злочини. Це має призвести до скорочення кількості злочинів в фінансовій сфері, захист населення від неправомірних транзакцій, відтак до підвищення довіри населення до органів Національної поліції України.

**ПЕРЕЛІК ПОСИЛАНЬ**

1. Паклін, М.Б. Побудова класифікаторів на незбалансованих вибірках на прикладі кредитного скорингу / М.Б. Паклін, С.В. Уланов, С.В. Царьков // Штучний інтелект. – 2010. В. 3. – с. 528-534.

2. Schapire R.E. The Boosting Approach to Machine Learning: An Overview [Архівовано 20 вересня 2020 у Wayback Machine.], MSRI (Mathematical Sciences Research Institute) Workshop on Nonlinear Estimation and Classification. – 2003. – 23 p.

3. Zhou, Zh.H. On the doubt about margin explanation of boosting / Wei Gao, Zhi-Hua Zhou // Artificial Intelligence 203: 1–18. – 2013. doi:10.1016/j.artint.2013.07.002.

4. Laplante, Ph. Requirements Engineering for Software and Systems (вид. 1st). Redmond, WA: CRC Press. - 2009. – 264 p.

5. Berenbach B. Software & Systems Requirements Engineering: In Practice. / Brian Berenbach, Daniel Paulish, Juergen Katzmeier, Arnold Rudorfer // New York: McGraw-Hill Professional. – 2009. – 546 p.