https://doi.org/10.31891/2307-5732-2026-365-21
УДК 004.02

ІЛЬЧИШИН ВЛАДИСЛАВ
Хмельницький національний університет
e-mail: ilchishin.vladislav@gmail.com
МАНЗЮК ЕДУАРД
Хмельницький національний університет
https://orcid.org/0000-0002-7310-2126
e-mail: eduard.em.km@gmail.com
ПЕТРОВСЬКИЙ СЕРГІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-0590-0484
e-mail: petrovskijs69@gmail.com
БАГРІЙ РУСЛАН
Хмельницький національний університет
https://orcid.org/0000-0001-5219-1185
e-mail: bahriiro@khmnu.edu.ua

МЕТОД ВИЯВЛЕННЯ ШАХРАЙСЬКИХ БАНКІВСЬКИХ ОПЕРАЦІЙ НА ОСНОВІ ТЕХНОЛОГІЙ ЗАСТОСУВАННЯ ЗАСОБІВ МАШИННОГО НАВЧАННЯ ТА ЗАХИСТУ ІНФОРМАЦІЇ В РОЗПОДІЛЕНО-ПАРАЛЕЛЬНОМУ ВИКОРИСТАННІ У ПРОЦЕСІ УПРАВЛІННЯ ІТ-ПРОЄКТАМИ

Швидке зростання обсягів цифрових фінансових транзакцій підвищує ризики здійснення шахрайських операцій та ускладнює їх своєчасне виявлення. Традиційні методи, що ґрунтуються на ручному аналізі або простих правилах, демонструють недостатню ефективність через високу кількість хибних спрацювань і низьку адаптивність до нових типів загроз. У сучасних умовах особливої актуальності набуває інтеграція технологій машинного навчання та засобів захисту інформації в розподілено-паралельних системах, що застосовуються у процесі управління ІТ-проєктами фінансового сектору.

У роботі запропоновано метод виявлення та класифікації шахрайських банківських операцій на основі алгоритму Random Forest із використанням техніки SMOTE для балансування класів. Розроблена модульна архітектура забезпечує ефективну обробку транзакційних даних, паралельне навчання моделей, валідацію та тестування у високонавантажених середовищах. Експериментальні дослідження на реальних банківських даних показали високу результативність запропонованого підходу: F1-міра – 0,83, точність – 0,86, повнота – 0,79, площа під ROC-кривою – 0,93. Метод дозволяє виявляти до 79% шахрайських операцій при рівні хибних спрацювань 4,5%.

Отримані результати мають наукове значення для подальшого розвитку алгоритмів машинного навчання у сфері фінансової безпеки, а також практичну цінність – підвищують рівень захисту банківських транзакцій, зменшують потенційні фінансові збитки та підвищують ефективність роботи систем моніторингу та служб безпеки фінансових установ.

Ключові слова: *фінансове шахрайство, класифікація транзакцій, машинне навчання, випадковий ліс, незбалансовані дані, автоматизація банківської безпеки, захист інформації, управління ІТ-проєктами.*

ILCHYSHYN VLADYSLAV, MANZIUK EDUARD, PETROVSKYI SERGII, BAHRII RUSLAN
Khmelnytskyi National University

METHOD FOR DETECTION OF FRAUDULENT BANKING TRANSACTIONS BASED ON TECHNOLOGIES OF APPLICATION OF MACHINE LEARNING TOOLS AND INFORMATION PROTECTION IN DISTRIBUTED-PARALLEL USE IN THE PROCESS OF IT PROJECT MANAGEMENT

The rapid growth of digital financial transactions has created a complex environment for ensuring banking security. Detecting fraudulent transactions is increasingly challenging due to the diversity and evolving nature of fraud schemes. Traditional monitoring approaches based on manual analysis and simple rule-based systems are inefficient because of high labor costs, numerous false positives, and limited adaptability to emerging types of fraud.

Advances in machine learning enable the development of adaptive fraud detection systems capable of automatically identifying complex patterns of suspicious behavior. Ensemble learning methods and techniques for handling class imbalance significantly improve transaction classification accuracy while reducing the workload on security services. The scarcity of labeled fraudulent examples limits conventional supervised models, and synthetic data generation methods such as SMOTE can create additional training samples, enhancing model performance.

This study presents a method for detecting and classifying fraudulent banking transactions using a Random Forest classifier enhanced with SMOTE and integrated into a distributed-parallel processing framework suitable for IT project management. The approach includes modular data processing, model training, validation, and testing, providing scalability and adaptability. Experimental evaluation on a real-world bank transaction dataset shows high effectiveness: F1-score 0.83, precision 0.86, recall 0.79, and area under the ROC curve 0.93, outperforming logistic regression, decision trees, and standard Random Forest. The method detects approximately 79% of fraudulent transactions while maintaining a low false positive rate of 4.5%, demonstrating practical applicability in operational banking environments.

The research contributes scientifically by advancing machine learning algorithms for anomaly detection in highly imbalanced financial data and developing adaptive detection systems. Practically, it reduces financial losses, improves transaction security, and increases the efficiency of bank security operations, enhancing trust in digital financial services. The proposed method provides a robust foundation for the development of intelligent and adaptive fraud detection systems in the financial sector.

Keywords: financial fraud, transaction classification, machine learning, random forest, unbalanced data, banking security automation, information protection, IT project management.

Стаття надійшла до редакції / Received 19.02.2026
Прийнята до друку / Accepted 11.03.2026
Опубліковано / Published 28.05.2026

Постановка проблеми у загальному вигляді
та її зв'язок із важливими науковими чи практичними завданнями

Сучасні фінансові системи функціонують у умовах значного зростання обсягів цифрових транзакцій, що створює складне середовище для забезпечення безпеки фінансових операцій. Однією з ключових проблем у цій сфері є виявлення шахрайських операцій, що стає все більш складним через різноманітність та динамічну еволюцію шахрайських схем. Традиційні підходи до моніторингу транзакцій, які базуються на ручному аналізі та простих правилах, демонструють низьку ефективність через високу трудомісткість, значну кількість помилкових спрацювань і обмежену здатність адаптуватися до нових типів шахрайства.

У контексті розвитку технологій захисту інформації, кібербезпеки та управління IT-проєктами виникає потреба у створенні інтелектуальних систем, які інтегрують методи машинного навчання з механізмами захисту даних і розподілено-паралельної обробки. Такі системи мають забезпечувати оперативне виявлення аномалій у великих потоках транзакційних даних, відповідати вимогам масштабованості та бути сумісними з існуючою банківською інфраструктурою, що становить важливе практичне завдання у сфері створення ПЗ.

Сучасні досягнення у галузі машинного навчання відкривають нові можливості для створення автоматизованих систем виявлення шахрайства, здатних аналізувати великі обсяги даних та виявляти складні патерни підозрілої поведінки. Зокрема, застосування ансамблевих методів та алгоритмів обробки незбалансованих даних дозволяє значно підвищити точність класифікації транзакцій та зменшити навантаження на служби безпеки, що є важливим з практичної точки зору для фінансових установ.

Ще однією складністю є дефіцит розмічених прикладів шахрайських транзакцій, який обмежує ефективність традиційних наглядових алгоритмів. Використання методів синтетичної генерації даних, таких як SMOTE, дозволяє створювати додаткові навчальні приклади для моделей, що підвищує їхню здатність до точного виявлення шахрайства.

Розробка автоматизованих систем виявлення шахрайства є комплексною задачею у галузі технологій створення програмних продуктів. Така система повинна забезпечувати не лише високу точність класифікації, але й відповідати вимогам інформаційної безпеки, масштабованості та інтеграції з існуючою банківською інфраструктурою. Це потребує застосування сучасних методів розробки ПЗ, включаючи модульну архітектуру, автоматизоване тестування та механізми безперервної інтеграції.

Таким чином, проблема виявлення шахрайських фінансових операцій має як наукове, так і практичне значення. Науково вона стимулює розвиток алгоритмів машинного навчання, моделей обробки незбалансованих даних та адаптивних систем детектування аномалій. Практично – її рішення забезпечує зменшення фінансових втрат, підвищення безпеки транзакцій та ефективності роботи служб безпеки банківських і фінансових установ. Реалізація таких систем є критично важливою для підтримки стабільності фінансового середовища та підвищення довіри клієнтів до цифрових фінансових сервісів.

Аналіз досліджень та публікацій

Проблематика виявлення шахрайських банківських операцій займає провідне місце в сучасних фінансових дослідженнях, оскільки розвиток електронних платіжних систем супроводжується зростанням складних шахрайських схем. У науковій літературі відзначається, що класичні підходи до моніторингу транзакцій недостатньо ефективні на тлі великих обсягів цифрових даних та еволюції методів зловмисників [1]. Це особливо актуально для розподілено-паралельних обчислювальних середовищ, що використовуються у процесах управління IT-проєктами банківського сектору.

Задачі класифікації транзакцій традиційно формалізуються як бінарні, де операції відносять до легальних або шахрайських [2]. Однією з ключових проблем є суттєва незбалансованість класів: частка шахрайських транзакцій зазвичай менше 1%, що ускладнює навчання моделей [3, 4]. Традиційні моделі демонструють високу загальну точність, але низьку чутливість до рідкісних аномалій [5].

У літературі активно досліджуються методи попередньої обробки й балансування даних, включно з undersampling, oversampling та SMOTE [6–9]. Поєднання цих методів з оптимізацією порогів класифікації підвищує здатність моделей виявляти аномальні транзакції без значного збільшення хибних спрацювань.

Серед класичних методів машинного навчання увагу привертають ансамблеві алгоритми, зокрема випадковий ліс, який стабільно працює з великою кількістю ознак і дозволяє оцінювати важливість ознак [10, 11]. Метод градієнтного бустингу, зокрема XGBoost та LightGBM, ефективно підвищує якість класифікації шляхом послідовного навчання на помилках попередніх моделей [12–14].

Значний інтерес викликають рекурентні нейронні мережі, такі як LSTM та GRU, що дозволяють моделювати часові залежності транзакцій одного клієнта [15]. Аналіз транзакційних послідовностей допомагає підвищити точність виявлення аномальної поведінки [16]. Окрему групу складають графові нейронні мережі (GNNs), ефективні у моделюванні взаємозв'язків між транзакціями та рахунками [17].

Огляд наукових праць свідчить про активний розвиток багатокомпонентних систем виявлення шахрайства, у яких поєднуються машинне навчання, захист інформації та масштабовані розподілені архітектури. Це створює підґрунтя для розробки ефективних інтелектуальних методів детектування, релевантних завданням сучасного управління IT-проєктами у фінансовій сфері.

Формулювання цілей статті

Мета роботи полягає в підвищенні точності виявлення та класифікації шахрайських банківських операцій шляхом розробки методу на основі ансамблевого навчання з балансуванням класів.

Задачі дослідження:

– провести аналіз існуючих методів та підходів до виявлення фінансового шахрайства з використанням методів машинного навчання;

– розробити метод виявлення та класифікації шахрайських транзакцій на основі алгоритму випадкового лісу з інтегрованою технікою SMOTE для вирішення проблеми незбалансованості класів;

– створити програмну реалізацію методу класифікації банківських транзакцій з модульною архітектурою, що забезпечує масштабованість, адаптивність та інтеграцію у розподілено-паралельне середовище ІТ-проєктів;

– провести експериментальне дослідження ефективності спроектованого методу шляхом порівняння з альтернативними алгоритмами класифікації та оцінки його точності на реальних транзакційних даних.

Суміжні дослідження демонструють успішне застосування методів машинного навчання та глибокого навчання в різних предметних областях, що підтверджує універсальність підходів класифікації та виявлення аномалій. Зокрема, методи кластеризації та аналізу даних показують високу ефективність при обробці транспортних потоків [18, 19], техніки пояснюваного глибокого навчання демонструють надійність у медичній діагностиці [20, 21, 23, 23], а підходи структурного вирівнювання онтологій забезпечують якісну обробку концептуальних категорій [22]. Ці роботи підкреслюють важливість застосування ансамблевих методів, балансування даних та інтерпретованості моделей, що є ключовими принципами і для систем виявлення фінансового шахрайства.

Виклад основного матеріалу

Розроблений метод виявлення шахрайських банківських операцій базується на застосуванні алгоритмів машинного навчання для автоматичної класифікації транзакцій у процесі управління ІТ-проєктами та забезпечення безпеки інформації. Модель навчається на історичних даних, що містять приклади як легальних, так і шахрайських операцій, після чого використовується для аналізу нових транзакцій у реальному часі. Метод ґрунтується на припущенні, що шахрайські операції мають специфічні ознаки та поведінкові патерни –зокрема, відмінності у сумі, часі, геолокації чи типі рахунку. Загальна схема методу (рис. 1) демонструє послідовність етапів та можливі зворотні зв'язки, що виникають під час оптимізації моделі. Важливою властивістю підходу є **модульність архітектури**, що дозволяє замінювати окремі компоненти – алгоритми навчання, методи попередньої обробки чи балансування класів – без потреби перебудови всієї системи. Це забезпечує високу гнучкість, масштабованість та адаптивність системи, що є критично важливим для інтеграції в існуючу ІТ-інфраструктуру банківських установ та підвищення ефективності служб безпеки при обробці великих обсягів транзакційних даних.

Розроблений метод виявлення шахрайських банківських операцій інтегрує технології машинного навчання та засоби захисту інформації у розподілено-паралельному середовищі, що забезпечує високу швидкодію та масштабованість у процесі управління ІТ-проєктами фінансового сектору. Метод включає структуровану послідовність етапів обробки та аналізу транзакційних даних. На першому етапі проводиться завантаження та первинний аналіз транзакцій для оцінки структури датасета та визначення потенційних проблем. Другий етап передбачає підготовку даних: обробку пропусків, усунення викидів, нормалізацію ознак та кодування категоріальних змінних, що забезпечує коректну роботу алгоритмів.

Третім етапом є подолання незбалансованості класів, оскільки частка шахрайських операцій є мінімальною. Для цього застосовуються методи балансування, які підвищують здатність моделі розпізнавати рідкісні аномалії. На завершальному четвертому етапі здійснюється навчання моделі на підготовлених даних, включно з налаштуванням параметрів та валідацією. Отримана модель формує стійкий і точний механізм класифікації транзакцій, здатний ефективно розмежовувати легальні та шахрайські операції в реальному часі та інтегруватися у системи підтримки прийняття рішень під час управління ІТ-проєктами у банківській сфері.

П'ятий етап методу передбачає тестування навченої моделі на незалежному тестовому наборі, що дає змогу оцінити її здатність узагальнювати знання на нових транзакціях. На цьому етапі обчислюються основні метрики якості класифікації. Завершальний етап включає застосування моделі для класифікації нових операцій: після проходження стандартної попередньої обробки модель визначає ймовірність шахрайства, на основі якої приймається рішення щодо транзакції. Такий підхід забезпечує оперативність та надійність, що є критично важливими для процесів управління ІТ-проєктами у фінансовій сфері.

Метод також підтримує адаптивність: періодичне перенавчання на оновлених даних забезпечує актуальність моделі та її здатність реагувати на зміну характеру шахрайських операцій у банківській сфері.

Для класифікації банківських транзакцій обрано алгоритм випадкового лісу – ансамблевий метод машинного навчання, що характеризується високою точністю, стійкістю до перенавчання та здатністю працювати з великою кількістю як числових, так і категоріальних ознак. Ці властивості, разом із можливістю інтеграції у розподілені обчислювальні середовища та системи захисту інформації, роблять його доцільним інструментом для виявлення шахрайських операцій та реалізації надійних ІТ-проєктів у банківській галузі.

Випадковий ліс складається з множини дерев рішень, кожне з яких навчається на бутстреп-підвибірці навчальних даних. Завдяки незалежності дерев їх помилки взаємно компенсуються, що підвищує якість класифікації порівняно з окремою моделлю.

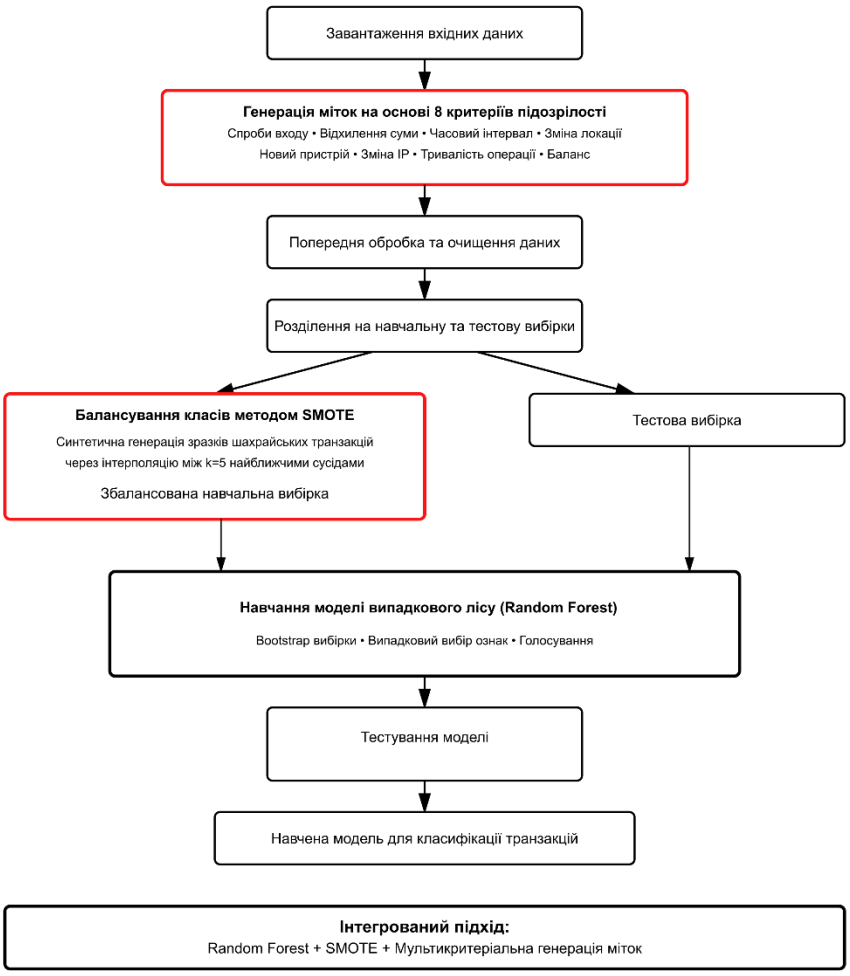

Рис. 1. Загальна схема методу виявлення шахрайських операцій

У процесі навчання для кожної підвибірки будується дерево, причому на кожному кроці розглядається випадкова підмножина ознак – зазвичай розміром, що дорівнює квадратному кореню з їх загальної кількості. Це забезпечує різноманітність дерев та зменшує кореляцію між ними, що є критично важливим для стабільності системи у розподілено-паралельних обчисленнях. Дерева зазвичай вирощуються до повної глибини, а перенавчання запобігається завдяки подальшому усередненню їхніх передбачень.

Під час класифікації нової транзакції вона проходить через усі дерева, кожне з яких формує власне рішення. Остаточний клас визначається більшістю голосів, що забезпечує надійність та стабільність результатів моделі навіть за умов динамічних змін у патернах шахрайської активності. Такий механізм дозволяє реалізувати ефективний, масштабований і захищений інструмент виявлення шахрайства, який може бути інтегрований в IT-інфраструктуру банківського проєкту та підтриманий методами управління IT-проєктами.

У задачі виявлення шахрайства випадковий ліс може формувати ймовірнісну оцінку належності транзакції до класу шахрайських — як частку дерев, що віднесли її до цього класу. Це дозволяє встановлювати порогові значення та регулювати баланс між виявленням фроду та кількістю хибних спрацювань.

Алгоритм також забезпечує оцінку важливості ознак, визначаючи внесок кожної з них у зменшення помилки класифікації. Такі оцінки дають змогу визначити, які характеристики транзакцій найбільше впливають на виявлення шахрайства.

Архітектура запропонованого алгоритму виявлення шахрайських банківських операцій визначається набором ключових параметрів, які забезпечують баланс між точністю класифікації, обчислювальною ефективністю та можливістю інтеграції в розподілено-паралельні системи обробки транзакцій. До таких параметрів належать кількість дерев ансамблю, їх максимальна глибина, мінімальна кількість зразків для побудови розбиття та формування листових вузлів. Збільшення числа дерев традиційно сприяє підвищенню точності моделі, однак водночас потребує більше обчислювальних ресурсів, що є критичним фактором при впровадженні системи у рамках масштабованих IT-проєктів. У цьому дослідженні використано 100 дерев, що забезпечує стабільність без надмірних витрат. Глибину дерев не обмежено, мінімальну кількість зразків для розбиття встановлено на рівні 2, а для листа – 1. Кількість ознак для кожного розбиття обчислюється як корінь квадратний із загальної кількості ознак. Для оцінки якості розбиттів застосовано критерій Джині.

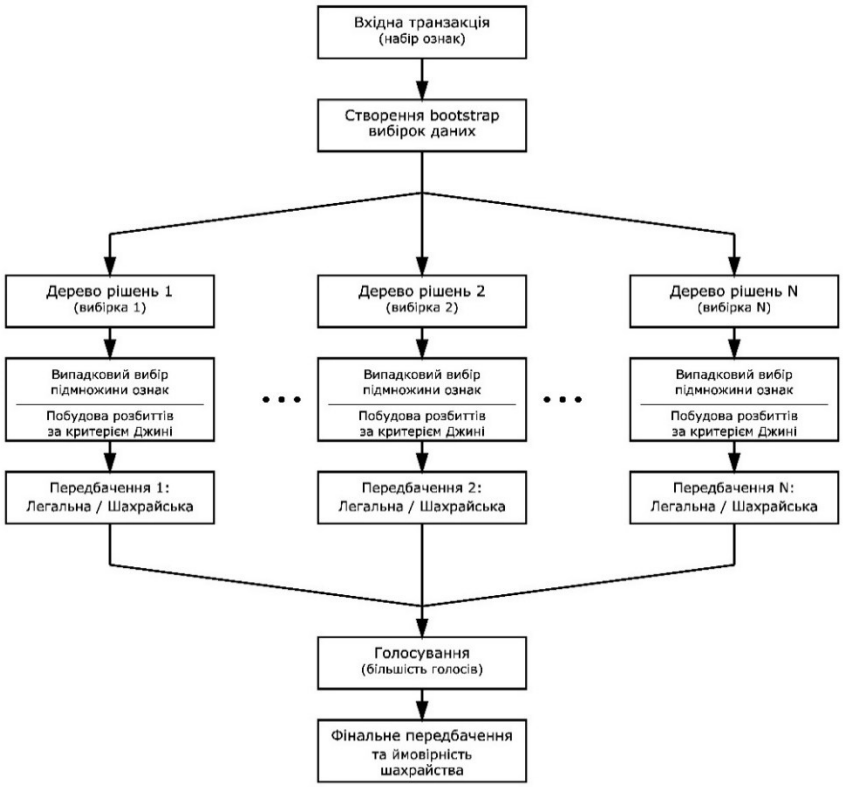

Рис. 2. Архітектура моделі класифікації транзакцій

Експериментальні дослідження

Для експериментальної перевірки методу використано Bank Transaction Dataset, що містить інформацію про банківські операції. Кожна транзакція описується набором ознак: ідентифікатори транзакції та рахунку, сума, тип операції, дата й час, геолокація, ідентифікатор пристрою, IP-адреса, тривалість, кількість спроб входу, баланс рахунку.

Оскільки датасет не містив міток класів, їх було автоматично сформовано на основі восьми критеріїв підозрілості (надмірні спроби входу, аномальна сума, надто малий інтервал між операціями, різке географічне переміщення, новий пристрій, зміна IP, нетипова тривалість, низький баланс). Транзакції з ≥3 балами позначались як шахрайські. У результаті отримано 126 шахрайських та 2386 легальних операцій.

Підготовка даних включала аналіз і заповнення пропусків: 12 відсутніх значень для геолокації, 80 – для ідентифікатора пристрою, 5 – для IP-адреси. Для числових ознак застосовувалась медіана, для категоріальних – найчастіше значення (SimpleImputer). Результати застосування SMOTE наведено на рисунку 3, де показано порівняння метрик моделі до та після балансування. Балансування суттєво підвищило здатність моделі виявляти шахрайські операції: повнота зросла з 0.38 до 0.79.

Числові ознаки (сума, тривалість, кількість спроб входу, баланс) нормалізовано методом стандартизації (StandardScaler). Категоріальні ознаки – тип транзакції, геолокація, ідентифікатор пристрою – закодовано через OneHotEncoder.

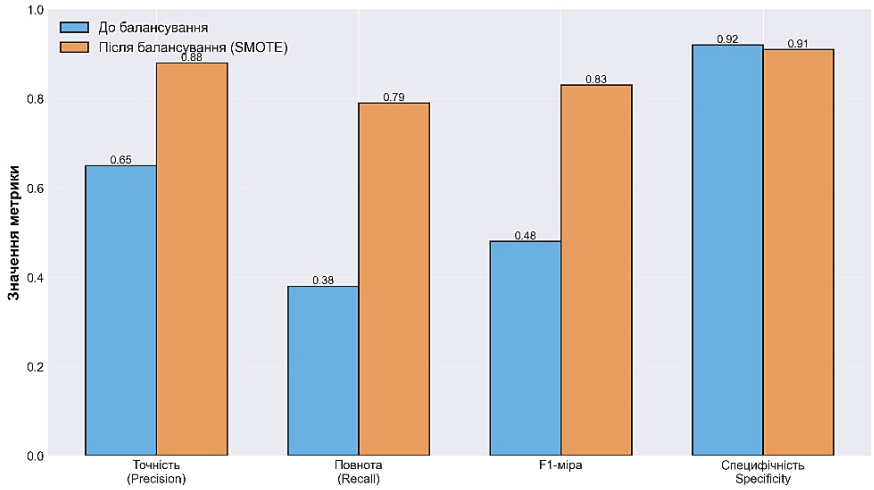

Рис.3. Порівняння метрик класифікації до та після балансування класів

Дані поділено на вибірки зі збереженням пропорції класів: 1758 транзакцій у навчальній вибірці, 377 у валідаційній та 377 у тестовій. Сильний дисбаланс класів у навчальній вибірці (≈95% легальних проти 5% шахрайських транзакцій) призводив до того, що модель могла досягати високої точності, фактично не виявляючи шахрайства. Для усунення цієї проблеми застосовано SMOTE, який генерує синтетичні зразки: для кожної шахрайської транзакції визначається п'ятеро найближчих сусідів, випадково обирається один із них, і новий зразок створюється шляхом інтерполяції між двома точками. Для навчання класифікатора використано алгоритм Random Forest із бібліотеки scikit-learn. Він формує ансамбль дерев рішень, кожне з яких тренується на випадковій підмножині даних та ознак. Основним параметром є кількість дерев, яку варіювали від 10 до 200, оцінюючи F1-міру на валідаційній вибірці. Графік (рис. 4) показує зростання F1-міри зі збільшенням числа дерев, але після 100 дерев приріст майже зникає: при 100 деревах F1 = 0.82, при 200 — 0.83.

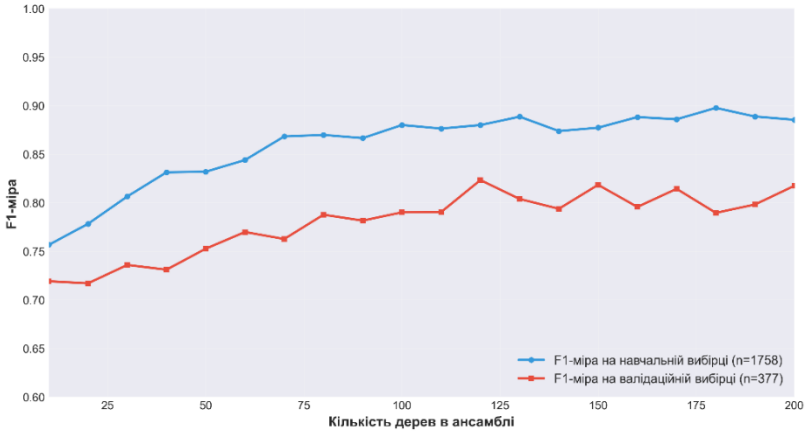

Рис. 4. Залежність якості класифікації від кількості дерев

Процес генерації тривав до вирівнювання класів: у вихідній навчальній вибірці було близько 1670 легальних та 88 шахрайських транзакцій, після застосування SMOTE кількість шахрайських зразків збільшено до 1670. Балансування виконувалось лише для навчальної вибірки, тоді як валідаційна та тестова частини залишалися незмінними, що забезпечило реалістичну та об'єктивну оцінку моделі.

Для балансу між якістю класифікації та складністю моделі вибрали 100 дерев. Як критерій розбиття встановлено індекс Джині, а параметр *max_features* – корінь квадратний із загальної кількості ознак, що забезпечує різноманітність дерев. Глибина дерев не обмежувалась; мінімальна кількість зразків у вузлі – 2, у листі – 1. Після навчання обчислено важливість ознак як середнє зменшення індексу Джині; результати наведені на рисунку 5.

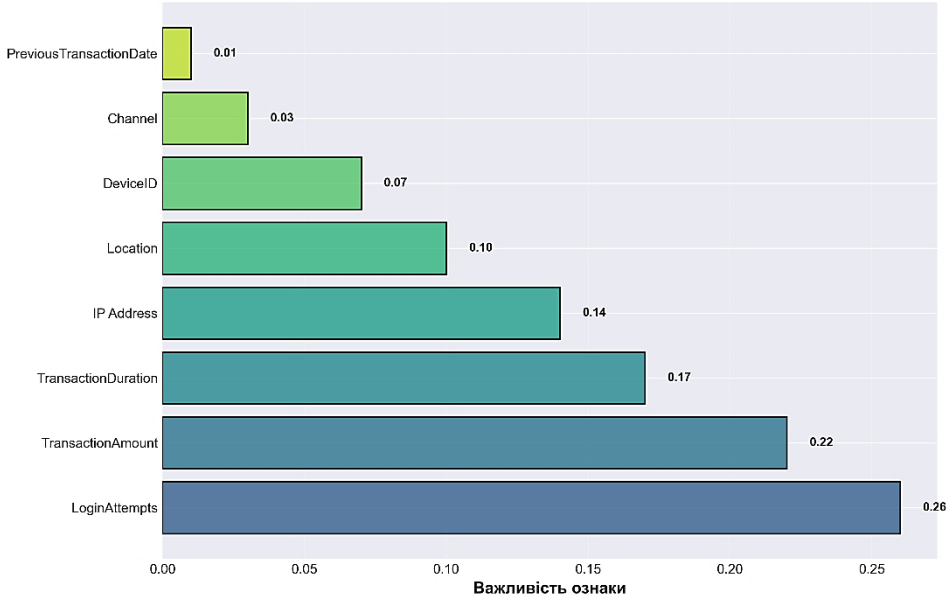

Рис.5. Важливість ознак для виявлення шахрайських операцій

Найважливішою ознакою стала кількість спроб входу (0.26), що найкраще відрізняє шахрайські транзакції від легальних. Далі за значущістю йдуть сума операції (0.22) та її тривалість (0.17). Геолокація та IP-адреса мають нижчу важливість – 0.10 і 0.14. Найменшу важливість показали канал, ідентифікатор пристрою та попередня транзакція – 0.03, 0.07 і 0.01.

Для оцінки якості моделі побудовано криву точності-повноти та ROC-криву, що демонструють роботу класифікатора при різних порогах. Площа під кривою точності-повноти становить 0.84, що свідчить про високу

якість моделі. Робоча точка при порозі 0.5 дає точність 0.86 і повноту 0.79. Підвищення порогу збільшує точність, але знижує повноту (наприклад, при 0.7: точність 0.93, повнота 0.63). Зниження порогу до 0.3 підвищує повноту до 0.89, але зменшує точність до 0.71. Вибір порогу залежить від пріоритетів – максимальне виявлення шахрайства чи мінімізація хибних спрацювань.

ROC-крива відображає співвідношення істинно та хибно позитивних спрацювань. Графік ROC-кривої представлений на рисунку 6. Діагональ відповідає випадковому класифікатору.

Площа під ROC-кривою становить 0.91, що вказує на високу здатність моделі розрізняти класи. Робоча точка має координати: хибно позитивні 0.09, істинно позитивні 0.79. ROC-крива значно перевищує діагональ і має вигин у верхній лівий кут, що демонструє високу чутливість при низькій частці хибних спрацювань. Обидві криві підтверджують високу якість моделі: площі 0.84 і 0.93 значно перевищують рівень базового класифікатора та свідчать про надійне розрізнення шахрайських і легальних транзакцій.

Для оцінки ефективності методу проведено порівняння з популярними алгоритмами, навченими на однакових даних. Логістична регресія як базова модель дала F1 = 0.64 (точність 0.71, повнота 0.58), що свідчить про пропуск багатьох шахрайських транзакцій через лінійність моделі. Дерево рішень (критерій Джині, глибина 10) показало кращі результати: F1 = 0.72, точність 0.78, повнота 0.67, проте лишається ризик перенавчання. Третя модель базовий Random Forest із 100 дерев без балансування класів, навчений на незбалансованій вибірці. Порівняльні результати наведені на рисунку 7.

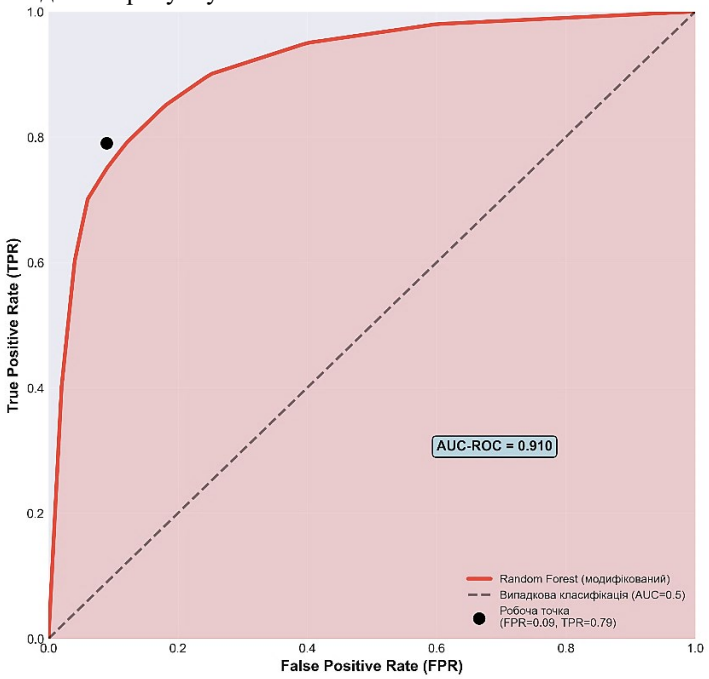

Рис.6. ROC-крива моделі класифікації

Базовий Random Forest дав F1 = 0.76 (точність 0.83, повнота 0.71), що краще за дерево рішень, але гірше за розроблений метод, оскільки незбалансовані дані погіршили виявлення шахрайства. Розроблений Random Forest із SMOTE досяг F1 = 0.82 (точність 0.86, повнота 0.79), значно покращивши розпізнавання шахрайських транзакцій. XGBoost (100 дерев, глибина 6, learning rate 0.1) показав F1 = 0.78, що трохи нижче за результат SMOTE-Random Forest, ймовірно через особливості даних або потребу в донастроюванні. Загалом, метод Random Forest із SMOTE продемонстрував найкращу F1-міру серед усіх алгоритмів.

Порівняння з базовим Random Forest показує, що балансування класів суттєво підвищує якість: SMOTE збільшив F1 з 0.76 до 0.82 та повноту з 0.71 до 0.79, забезпечивши додаткове виявлення 8% шахрайських транзакцій. Перевага над XGBoost свідчить, що саме поєднання Random Forest і SMOTE виявилось найефективнішим для цієї задачі.

Аналіз помилок системи класифікації дає змогу зрозуміти слабкі місця моделі та напрямки для подальшого покращення. Матриця помилок показала 4 хибно негативні та 2 хибно позитивні помилки. Пропущені шахрайські транзакції були малопомітними: звичні суми, відомі пристрої, стандартні локації, лише поодинокі слабкі індикатори (зміна IP, швидке виконання, невдалий вхід). Модель оцінила їх як недостатньо підозрілі, тому 4 із 19 випадків були класифіковані помилково. Хибно позитивні помилки виникали через незвичні для клієнтів операції нові пристрої після зміни телефону, великі нетипові суми або транзакції під час подорожей. Аналіз показує можливість покращення моделі за рахунок нових та оптимізації порогу класифікації.

Експеримент з різним обсягом навчальних даних показав, що якість моделі швидко зростає від 10% до 40% даних (F1 від 0.58 до 0.75), після чого покращення сповільнюється. На 70% даних модель досягає F1 = 0.82 значення, використане в основних експериментах. Подальше збільшення вибірки дає мінімальний приріст (до F1 = 0.84 на 100%), що свідчить про вихід на плато. Для досягнення F1 > 0.75 потрібно щонайменше 40% даних (~700 транзакцій).

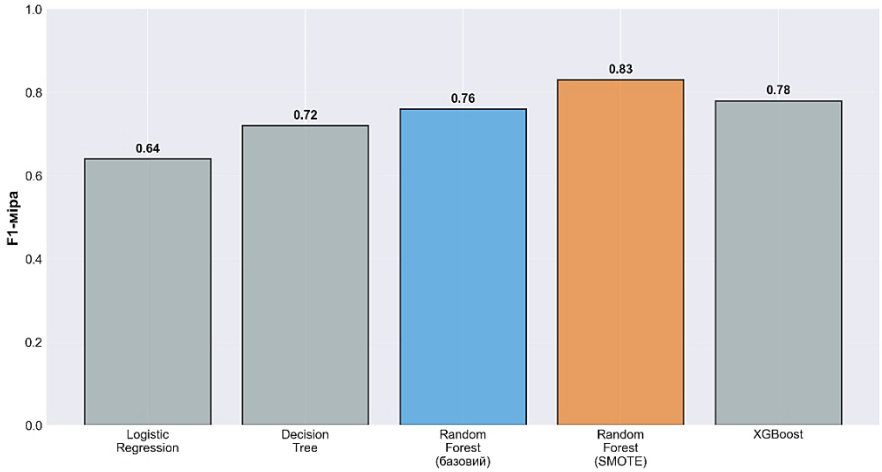

Рис.7. Порівняння якості моделей

Практичний аналіз показує, що при повноті 0.79 модель виявляє 79 зі 100 шахрайських операцій (близько 395 з 500 на день), пропускаючи приблизно 105. Точність 0.88 означає, що з усіх позначених системою підозрілих 88% є шахрайськими, що створює близько 54 хибних спрацювань на день прийнятне навантаження для служби безпеки. Специфічність 0.99 гарантує мінімальні незручності для клієнтів.

Порівняно з альтернативами, модель істотно краща: логістична регресія (F1 = 0.64) пропустила б додатково ~105 випадків шахрайства, базовий Random Forest (F1 = 0.76. Найважливіші ознаки кількість спроб входу та сума транзакції; вони вказують на ключові напрямки моніторингу.

Аналіз помилок визначив ситуації, що потребують особливої уваги: транзакції з нових пристроїв, під час подорожей, або нетипово великі суми вони часто спричиняють хибні спрацювання. Модель можна покращити додаванням поведінкових ознак та адаптивними порогами: підвищеним для великих сум і зниженим для транзакцій з нових локацій.

Практична система може працювати в режимах автоматичного блокування (ймовірність >0.9), попередження (0.5–0.9) та моніторингу (0.3–0.5). Такий підхід зменшує втрати від шахрайства, покращує роботу служби безпеки та підвищує комфорт клієнтів.

Висновки

У роботі запропоновано метод виявлення шахрайських банківських операцій на основі випадкового лісу та техніки SMOTE, що дозволяє ефективно балансувати класи транзакцій та підвищувати точність класифікації. Розроблена модульна програмна система забезпечує повний цикл обробки транзакцій та гнучкість у тестуванні й модифікації компонентів. Застосування сучасних технологій створення програмних продуктів, підходів до захисту інформації та паралельних обчислень робить систему придатною для інтеграції в банківські IT-проєкти та підтримки стабільної роботи фінансових сервісів з дотриманням вимог кібербезпеки та високою швидкодією при обробці транзакційних потоків.

Запропонована система реалізована з використанням сучасних технологій створення програмних продуктів. Модульна архітектура забезпечує гнучкість та масштабованість, дозволяючи незалежно розробляти та тестувати окремі компоненти. Система реалізована з використанням сучасних технологій створення програмних продуктів та орієнтована на інтеграцію в інфраструктуру банківських IT-проєктів. Модульна архітектура забезпечує високу гнучкість і масштабованість, що дозволяє ефективно управляти життєвим циклом програмного забезпечення, а також незалежно розробляти, тестувати й оновлювати окремі компоненти. Програмне забезпечення розроблено з урахуванням принципів безпечного кодування та відповідає вимогам захисту інформації, що є критично важливим для фінансових застосунків. Система включає компоненти попередньої обробки даних, навчання моделі, валідації та моніторингу, які взаємодіють через чітко визначені інтерфейси.

Для забезпечення високої продуктивності при обробці великих обсягів транзакційних даних у системі застосовано технології паралельних обчислень. Алгоритм Random Forest природно підтягається до паралелізації, оскільки окремі дерева рішень у ансамблі можуть навчатися незалежно. Це дозволяє ефективно використовувати багатоядерні процесори та розподіляти обчислювальне навантаження, суттєво скорочуючи час навчання моделі та прискорюючи процес класифікації транзакцій у реальному часі.

Експериментальне дослідження показало високу ефективність методу: F1-міра 0.83, точність 0.86, повнота 0.79, площа під ROC-кривою 0.93. Розроблений підхід дозволяє виявляти близько 79% шахрайських операцій при хибних спрацюваннях лише 4,5%, що перевищує результати традиційних методів та окремих алгоритмів класифікації.

З точки зору технологій захисту інформації та кібербезпеки, запропонований метод забезпечує додатковий рівень захисту фінансових операцій клієнтів, зменшуючи ризики несанкціонованого доступу та шахрайських транзакцій. Модульна архітектура програмного забезпечення дає змогу гнучко адаптувати систему до нових типів кіберзагроз, а також розширювати її функціональність шляхом інтеграції додаткових механізмів захисту та оптимізації моделей машинного навчання. Такий підхід сприяє ефективному управлінню IT-проєктами, орієнтованими на забезпечення кіберстійкості та високої продуктивності.

Отримані результати підтверджують практичну придатність методу для фінансових установ і його потенціал для зниження фінансових втрат та підвищення якості обслуговування клієнтів.

Література

1. Siam A. M., Bhowmik P., Uddin M. P. Hybrid feature selection framework for enhanced credit card fraud detection using machine learning models. *PLOS One*. 2025. Vol. 20, No. 7. URL https//doi.org/10.1371/journal.pone.0326975.

2. Hafez I. Y., Hafez A. Y., Saleh A., Abd El-Mageed A. A., Abohany A. A. A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*. 2025. Vol. 12, No. 1. Pp. 6. URL https//doi.org/10.1186/s40537-024-01048-8.

3. Credit Card Fraud Identification Using Machine Learning on Graphs – RelationalAI. URL https//www.relational.ai/post/credit-card-fraud-detection-machine-learning-graphs.

4. Alatawi M. N. Detection of fraud in IoT based credit card collected dataset using machine learning. *Machine Learning with Applications*. 2025. Vol. 19. Pp. 100603. URL https//doi.org/10.1016/j.mlwa.2024.100603.

5. Baisholan N., Dietz J. E., Gnatyuk S., Turdalyuly M., Matson E. T., Baisholanova K. FraudX AI An Interpretable Machine Learning Framework for Credit Card Fraud Detection on Imbalanced Datasets. *Computers*. 2025. Vol. 14, No. 4. Pp. 120. URL https//doi.org/10.3390/computers14040120.

6. Hayat K., Magnier B. Data Leakage and Deceptive Performance A Critical Examination of Credit Card Fraud Detection Methodologies. *Mathematics*. 2025. Vol. 13, No. 16. Pp. 2563. URL https//doi.org/10.3390/math13162563.

7. Moradi F., Tarif Hokmabadi M., Homaei M. A Systematic Review of Machine Learning in Credit Card Fraud Detection. Computer Science and Mathematics, 2025. URL https//doi.org/10.20944/preprints202507.1085.v1.

8. Thennakoon A., Bhagyani C., Premadasa S., Mihiranga S., Kuruwitaarachchi N. Real-time Credit Card Fraud Detection Using Machine Learning / *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, January 2019. Pp. 488–493. URL https//doi.org/10.1109/CONFLUENCE.2019.8776942.

9. Tanouz D., Subramanian R. R., Eswar D., Reddy G. V. P., Kumar A. R., Praneeth C. V. N. M. Credit Card Fraud Detection Using Machine Learning / *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2021. Pp. 967–972. URL https//doi.org/10.1109/ICICCS51141.2021.9432308.

10. Chen Y., Zhao C., Xu Y., Nie C., Zhang Y. Year-over-Year Developments in Financial Fraud Detection via Deep Learning A Systematic Literature Review. arXiv, 2025. URL https//doi.org/10.48550/arXiv.2502.00201.

11. Chen Y., Zhao C., Xu Y., Nie C., Zhang Y. Deep Learning in Financial Fraud Detection Innovations, Challenges, and Applications. *Data Science and Management*. 2025. URL https//doi.org/10.1016/j.dsm.2025.08.002.

12. Jin J., Zhang Y. The analysis of fraud detection in financial market under machine learning. *Scientific Reports*. 2025. Vol. 15, No. 1. Pp. 29959. URL https//doi.org/10.1038/s41598-025-15783-2.

13. Afriyie J. K., Tawiah K., Pels W. A., Addai-Henne S., Dwamena H. A., Owiredu E. O., Ayeh S. A., Eshun J. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*. 2023. Vol. 6. Pp. 100163. URL https//doi.org/10.1016/j.dajour.2023.100163.

14. Salomon S. What is Fraud Detection for Machine Learning? *Feedzai*. URL https//www.feedzai.com/blog/what-is-fraud-detection-for-machine-learning/.

15. Credit Card Fraud Detection. URL https//www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

16. Fraud Detection Dataset. URL https//www.kaggle.com/datasets/goyaladi/fraud-detection-dataset.

17. Synthetic Financial Datasets For Fraud Detection. URL https//www.kaggle.com/datasets/ealaxi/paysim1.

18. Ryzhanskyi O., Pavlyshyn V., Radiuk P., Manziuk E., Barmak O., Krak I. AI-Driven Traffic Signal Control System to Reduce CO2 Emissions / CEUR Workshop Proc., CEUR-WS, 2025. Pp. 18–27. URL: https://ceur-ws.org/Vol-3974/paper02.pdf.

19. Pavlyshyn V., Ryzhanskyi O., Manziuk E., Radiuk P., Barmak O., Krak I. Establishing Patterns of the Urban Transport Flows on Clustering Analysis. CEUR Workshop Proceedings. 2025. Vol. 3974. Pp. 1–9. URL: https://ceur-ws.org/Vol-3974/paper01.pdf.

20. Manziuk E., Barmak O., Krak I., Petliak N., Jin Z., Radiuk P. Explainable Deep Learning for Interpretable Brain Tumor Diagnosis from MRI Images / Lecture Notes in Data Engineering, Computational Intelligence, and Decision-Making, Volume 1, Cham, Springer Nature Switzerland, 2024. Pp. 326–348. URL: https://doi.org/10.1007/978-3-031-70959-3_17.

21. Barmak O., Krak I., Yakovlev S., Manziuk E., Radiuk P., Kuznetsov V. Toward explainable deep learning in healthcare through transition matrix and user-friendly features. Frontiers in Artificial Intelligence. 2024. Vol. 7. Pp. 1482141. URL: https://doi.org/10.3389/frai.2024.1482141.

22. Manziuk E., Krak I., Barmak O., Mazurets O., Kuznetsov V., Pylypiak O. Structural alignment method of conceptual categories of ontology and formalized domain. 2021. Pp. 11–22.

23. An adaptive approach to detecting fake news based on generalized text features(Conference Paper) Shupta, A., Barmak, O., Wierzbicki, A., Skrypnyk, T. // 7th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Machine Learning Workshop, CoLInS 2023; Kharkiv; Ukraine; 20 April 2023 до 21 April 2023; Код 188444 // CEUR Workshop Proceedings Volume 3387, 2023, Pages 300-310

24. E. Manziuk, O. Barmak, I. Krak, O. Mazurets, and T. Skrypnyk, “Formal Model of Trustworthy Artificial Intelligence Based on Standardization.,” in CEUR Workshop Proceedings, Khmelnytskyi, Ukraine, Mar. 2021, vol. 2853, pp. 190–197. http://ceur-ws.org/Vol-2853/