УДК 004.732.056 **doi: 10.26906/SUNZ.2025.3.64-71**

О. А. Горносталь, В. В. Челак

Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

КЛАСИФІКАЦІЯ МЕРЕЖЕВИХ АТАК МЕТОДАМИ МАШИННОГО НАВЧАННЯ В УМОВАХ ДИСБАЛАНСУ ТРЕНУВАЛЬНИХ ДАНИХ

Анотація. Об'єктом дослідження є процес виявлення мережевих вторгнень. **Предметом дослідження** є методи класифікації мережевих вторгнень. **Метою дослідження** є підвищення якості та швидкості ансамблевих класифікаторів в задачах класифікації мережевих вторгнень в умовах дисбалансу тренувальних даних. **Методи, що використовуються**: методи машинного навчання, методи попередньої обробки даних, ансамблеві класифікатори, метод перебалансування шляхом синтетичного збільшення меншості (SMOTE). **Отримані результати:** досліджено ефективність використання різних підходів для класифікації мережевих вторгнень в умовах дисбалансу класів у тренувальній вибірці. Запропоновано комплексний підхід, що передбачає попередню обробку даних за допомогою методу SMOTE для синтетичного балансування навчальної вибірки, а також його подальший аналіз з використанням ансамблевих моделей машинного навчання, що дало змогу покращити показники класифікації міноритарних класів. Найкращі результати отримано при поєднанні SMOTE з ансамблевими моделями, зокрема Bagging, Gradient Boosting та AdaBoost. **Висновки.** За результатами дослідження запропоновано удосконалений підхід до класифікації мережевого трафіку, який поєднує попереднє балансування вибірки методом SMOTE з ансамблевими алгоритмами беггінг та бустинг. Комплексне використання цих методів дозволило покращити значення метрики Recall для міноритарних класів. Загалом запропонований підхід забезпечив покращення якості класифікації: 18% для атак типу Infiltration, 33% для атак з використанням SQL-ін’єкцій та до 53% для атак XSS у порівнянні з базовими моделями машинного навчання без додаткового перебалансування вхідних даних.

Ключові слова: мережеві атаки, класифікація, машинне навчання, ансамблі, беггінг, бустинг, показники якості, SMOTE.

Вступ

Інформаційні системи давно стали невід’ємним атрибутом майже усіх сфер людського життя: від вирішення інженерних задач до обробки банківських транзакцій та зберігання медичних відомостей. При цьому будь-які потенційні або реальні загрози їх цілісності можуть спричинити значні підвищення витрат на їх обслуговування. Незважаючи на це, питання безпеки інформаційних систем залишається досить актуальним з урахуванням великої кількості кібератак за останні роки. Це підтверджується статистичними даними [1] та пояснюється тим фактом, що більшість з них працює в мережі та постійно обмінюється великими об’ємами інформації, що дозволяє зловмисникам легше отримувати до них доступ та швидше виконувати протиправні дії по відношенню до інфраструктури та даних. Крім цього, при роботі з комп’ютерними мережами необхідно враховувати апаратні збої, перевантаження та ризики, незалежні від зовнішніх чинників.

Задача також ускладняється наявністю великої кількості складників та параметрів, якими характеризується функціонування комп’ютерної мережі. Через це традиційні аналітичні методи або не дозволяють отримати бажаної точності виявлення вторгнень, або не можуть впоратися з їх якісним розрізненням, або працюють надто довго, що робить неможливим їх використання у динамічних системах, де необхідно приймати рішення в режимі реального часу. Завдяки цьому постає нагальна потреба в удосконаленні існуючих підходів та у створенні нових, які б могли задовольнити сучасним потребам.

Об'єктом дослідження є процес виявлення мережевих вторгнень. **Предметом дослідження є** методи класифікації мережевих вторгнень.

Огляд пов'язаних наукових публікацій. Виявлення мережевих вторгнень представляє собою комплексний процес збору та обробки інформації про поточні події в мережі. Одним з класичних підходів є сигнатурна група методів [2-3], які полягають у дослідженні відомих закономірностей та в пошуку схожих елементів з відомими загрозами. До них відносять [4] алгоритми порівняння шаблонів, хешей та використання Yara-правил [5]. Основним недоліком цієї групи методів є фактична неможливість виявлення раніше невідомих загроз, адже спостереження та аналіз ведуться з урахуванням вже досліджених вторгнень на основі виявлення точних або часткових співпадінь. Навіть якщо шаблони певної атаки існують, але зловмисники внесли деякі зміни в алгоритм своїх дій, то їх буде дуже складно виявляти подібними методами. Це обмеження також стосується обфускованих даних.

Альтернативним підходом, який може вирішити ці проблеми, є використання евристичних методів які полягають в аналізі поведінки процесів. Фактично, фокус зміщується з зовнішнього вигляду подій та їх структури на аналіз того, що саме відбувається та які закономірності з цим пов’язані. Найпростішими представниками цієї групи є статистичні методи, що використовують різні математичні критерії для виявлення нестандартної поведінки мережі чи окремих процесів в ній для виявлення аномалій та вторгнень. До евристичних підходів також можна віднести методи аналізу часових рядів та контрольні карти, які дозволяють побудувати шаблони нормальної поведінки середовища на основі наявних даних про його роботу в різних режимах для подальшого виявлення подій виходу за ці встановлені обмеження. Зазначені підходи дозволяють працювати з загрозами незалежно від того, який алгоритм чи програмний код

використовувався в процесі їх виконання, адже увага зосереджується саме на подіях.

Незважаючи на очевидні переваги класичні евристичні методи досить важко працюють зі складними закономірностями, які передбачають аналіз великого набору даних, наприклад, мережевих пакетів з десятками атрибутів. Розглянуті підходи також потребують постійного оновлення статистичних показників для адаптації до змінюваних умов та режимів роботи мережі. Крім того, зазвичай навчальні дані для аналізу різного роду реальних мереживах атак обмежені та мають значний дисбаланс класів, а розглянуті методи при недостатній кількості прикладів зразків атак можуть давати велику кількість хибних спрацьовувань або пропускати складні аномалії.

Саме тому для вирішення цих проблем необхідно застосовувати більш складні підходи, зокрема методи машинного навчання, які у задачах виявленні мережевих вторгнень [6] мають низку ключових переваг: вони здатні автоматично виявляти складні закономірності у великих обсягах даних, адаптуватися до нових типів атак, забезпечують високу точність класифікації та масштабованість для роботи з різноманітними мережевими середовищами.

Для досягнення більш стабільних і точних результатів використання машинного навчання за умови дисбалансу класів необхідно застосовувати додаткові підходи, зокрема методи перебалансування даних, до яких належать SMOTE [7], а також ансамблеві алгоритми [8], які дозволяють покращити узагальнюючу здатність моделей і підвищити якість класифікації міноритарних класів.

Постановка проблеми. Основна мета поточного дослідження полягає в аналізі ефективності використання різних методів підвищення якості класифікації мережевих вторгнень за умови обмежених та незбалансованих вхідних даних, коли кількість об’єктів одного класу значно переважає інші. Завдяки цьому стає нагальною потреба в розробці та вдосконалені такого методу, який би дозволив розрізняти як об’єкти мажоритарного класу, так і об’єкти міноритарного класу. Для виконання поставленої мети необхідно дослідити наступні напрямки підвищення ефективності роботи методів машинного навчання в умовах дисбалансу в навчальних даних:

1) використання підходів попередньої обробки даних, що полягають у перебалансуванні наявних екземплярів;

2) застосування ансамблевих методів з метою підвищення стабільності моделі та покращення її здатності розрізняти об’єкти міноритарних класів завдяки окремим спеціалізованим класифікаторам, які об’єднуються у групи;

3) комплексне використання зазначених підходів.

Для вирішення сформульованих задач необхідно діяти таким чином:

1) дослідити обраний набір даних, розподіл класів, розкид значень та інші статистичні характеристики;

2) розглянути ефективність роботи класичних методів машинного навчання в здачі бінарної (нормальний і аномальний стан мережі) та мультикласової класифікації (нормальний стан та декілька видів мережевих атак);

3) перевірити здатність методу перебалансуваня даних та підходу з використанням ансамблів впливати на якість виявлення окремих класів в обох підзадачах;

4) використати обидві стратегії в комплексі для оцінки їх ефективності підвищення якість класифікації;

5) сформувати висновки стосовно виявлення мережевих вторгнень в умовах незбалансованих тренувальних наборів.

Виконання поставлених кроків дозволить отримати набір рекомендацій для подальших досліджень з комплексним використанням спеціалізованих методів попередньої обробки даних та ансамблевих класифікаторів.

Огляд підходів та методів

В узагальненому вигляді проблему класифікації мережевих вторгнень можна представити у вигляді задачі пошуку вирішального правила, яке дозволяє зв’язати набір ознак з певною результуючою міткою. В такому випадку з метою формалізації представимо f(x) як модель класифікації, яка дозволяє отримати вхідний вектор ознак x і повернути значення відповідного класу. В такому випадку саме вирішальне правило набуває наступного вигляду:

$$\hat{y} = \arg\max_{c \in C} P\,(c|\,x), \qquad (1)$$

де $\hat{y}$ – передбачений клас; C – множина можливих класів; x – об’єкт, представлений множиною вхідних ознак; $P\,(C \mid x)$ – ймовірність належності об'єкта x до класу c; arg max – функція вибору мітки класу з найбільшою ймовірністю.

Якщо множина C містить більше двох елементів (наприклад, види атак: "SQL Injection", "Bruteforce", "XSS" тощо), то в такому випадку низька вірогідність приналежності до одного з класів не дає відповіді на запитання, до якої категорії віднести той чи інших об’єкт, бо є ще мінімум два варіанти. Задачі такого роду особливо ускладняються при наявності дисбалансу класів у використовуваних наборах даних.

Першим кроком вирішення цієї проблеми може бути використання спеціалізованих методів попередньої обробки незбалансованих даних. Одним з найпопулярнішою у цій категорії є SMOTE, який передбачає генерацію. нових екземпляри шляхом інтерполяції між випадковими точками меншості та їхніми найближчими сусідами того ж класу. Цю процедуру можна формалізувати наступним чином:

$$\tilde{x} = x_i + \delta \cdot (x_{NN} - x_i), \qquad (2)$$

де $\tilde{x}$ – новий синтетичний екземпляр класу; x_i – існуючий зразок міноритарного класу; один із x_{NN} k-найближчих сусідів x_i того ж класу; δ – коефіцієнт інтерполяції (зазвичай випадковий) на інтервалі від 0 до 1.

Наступним важливим кроком підвищення ефективності роботи з незбалансованими даними є використання ансамблевих методів. Вони дозволяють об'єднувати прогнози кількох базових моделей, зменшуючи ймовірність помилки, покращуючи стійкість до шуму та підвищуючи загальну точність класифікації. Завдяки різноманітності підходів до формування

ансамблю (наприклад, беггінг, бустинг), ці методи здатні краще виявляти складні шаблони навіть за умов дисбалансу.

Узагальнений процес об'єднання декількох моделей машинного навчання в ансамбль формалізується наступним виразом:

$$f(x) = \sum_{m=1}^{M} \alpha_m f_m(x), \quad (3)$$

де $f_m(x)$ – прогноз m-ї базової моделі; α_m – невід'ємна вага моделі m, що відображає її вплив у ансамблі; M – кількість моделей в ансамблі.

Ідея беггінг-ансамблів [9] полягає у навчанні кожної моделі $f_m^*(x)$ на випадковій вибірці з повторами. Ця процедура отримала назву бутстреп. В подальшому прогнози цих моделей об'єднуються за допомогою рівноважного голосування, що формалізується наступним виразом:

$$f_{bagging}(x) = \frac{1}{M}\sum_{m=1}^{M} f_m^*(x), \quad (4)$$

де $f_m^*(x)$ – прогноз m-ї базової моделі; M – кількість моделей в ансамблі.

Однією з варіацій беггінг-ансамблів є метод випадкового лісу (Random Forest), який об'єднує дерева рішень та при навчанні кожного з них додатково випадково вибирає підмножину ознак θ_m. В такому випадку процес формування результуючого прогнозу формалізується наступним виразом:

$$f_{RF}(x) = \frac{1}{M}\sum_{m=1}^{M} f_m^*(x, \theta_m), \quad (5)$$

де $f_m^*(x; \theta_m)$ – прогноз m-ї базової моделі, що навчалася з випадковою підмножиною ознак θ_m; M – кількість моделей в ансамблі.

З іншого боку ідея бустинг ансамблів полягає у почерговому покращенні наступних моделей з врахуванням помилок попередніх. AdaBoost [10] створює ансамбль послідовно навчених слабких моделей f_m^*, кожна з яких приділяє більше уваги прикладам, важким для попередніх моделей. В такому випадку фінальне прогнозування представляється таким виразом:

$$f_{Ada}(x) = sign(\sum_{m=1}^{M} \alpha_m \cdot f_m^*(x)), \quad (6)$$

де $f_m^*(x)$ – прогноз m-ї базової моделі; α_m – вага базової моделі в ансамблі; M – кількість моделей в ансамблі.

Ваговий коефіцієнт відображає, наскільки ансамбль довіряє конкретній моделі. Він розраховується за наступним виразом на основі похибки конкретного базового класифікатора:

$$\alpha_m = \frac{1}{2}\ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right), \quad (7)$$

де ε_m– зважена помилка класифікатора f_m^* та відображає частку об'єктів, які були класифіковані неправильно з урахуванням поточних ваг w_i прикладів.

Зважена помилка розрахується таким чином:

$$\varepsilon_m = \sum_{i=1}^{N} w_i \cdot I(f_m^*(x_i) \neq y_i), \quad (8)$$

де N – кількість прикладів; w_i – ваговий коефіцієнт прикладу; I – індикаторна функція (1, якщо класифіковано неправильно, 0 – в зворотному випадку); $f_m^*(x)$ – прогноз m-ої базової моделі

Інший ансамблевий метод градієнтного бустингу [11] будує модель поступово, додаючи на кожному кроці наступну слабку модель f_m^*, яка апроксимує негативний градієнт функції втрат L. Для цього за наступним виразом обчислюється негативний градієнт функції втрат за прогнозом попередньої моделі (псевдо-резидуал):

$$r_{im} = -\left[\frac{\partial \ell\,(y_i, F(x_i))}{\partial F\,(x_i)}\right]_{F(x_i)=F_{m-1}(x_i)}, \quad (9)$$

де r_{im} – псевдо-резидуал для об'єкта i на ітерації m; $\ell\,(y_i, F(x_i))$ – функція втрат між справжнім значенням y_i та прогнозом $F(x_i)$; $F_{m-1}(x_i)$ – прогноз моделі на попередній ітерації.

На кожному кроці алгоритму побудована модель апроксимує псевдо-резидуали та додається до поточного ансамблю з відповідним коефіцієнтом масштабування. Цей процес в градієнтному бустингу можна формалізувати у вигляді наступного виразу:

$$F_m(x) = F_{m-1}(x) + \gamma_m \cdot h_m(x), \quad (10)$$

де $F_m(x)$ – оновлений прогноз після m ітерацій; γ_m – коефіцієнт, який визначає вагу нової моделі; $h_m(x)$ – слабка модель, побудована для апроксимації r_{im} ; $F_{m-1}(x)$ – прогноз попередньої моделі.

Після завершення всіх ітерацій, фінальний прогноз моделі є сумою початкового припущення та всіх побудованих слабких моделей, що відображається у наступному виразі:

$$f_{Gradient}(x) = f_0(x) + \sum_{m=1}^{M} \gamma_m \cdot f_m^*(x)), \quad (11)$$

де $f_{Gradient}(x)$ – фінальний прогноз ансамблю; $f_0(x)$ – початкове значення (наприклад, середнє по y); M – кількість ітерацій (слабких моделей); γ_m – ваговий коефіцієнт окремої моделі; $f_m^*(x)$ – прогноз окремої моделі на ітерації m.

Формулювання теоретичних очікувань

Існує багато різних метрик для оцінювання якості класифікації, проте в умовах дисбалансу класів, коли цільовий (наприклад, тип мережевої атаки) є менш представленим, доцільно використовувати Recall як основну метрику, оскільки вона фокусує увагу на здатності моделі виявляти саме важливий міноритарний клас [12].

Використання методу синтетичного збільшення (передискретизації) вибірки меншості (SMOTE) дозволить зменшити вплив дисбалансу класів у тренувальній вибірці, підвищити Recall для міноритарних класів, зокрема таких атак, як XSS, SQL Injection або Infilteration, які у початковому датасеті представлені значно меншою кількістю прикладів, а також покращити стійкість класифікаторів до хибних негативних рішень.

Використання ансамблів, в свою чергу, має забезпечити підвищення точності та стабільності результатів класифікації завдяки поєднанню рішень багатьох слабких моделей, а також має забезпечити кращу здатність до узагальнення та зниження ймовірності перенавчання, яке часто виникає при використанні одиночних моделей на складних даних, особливо за умови значної переваги одного з класів.

Комплексне застосування SMOTE та ансамблів не лише має покращити окремі аспекти якості класифікації, а й створити надійнішу, збалансованішу та практично орієнтовану модель, здатну до ефективного виявлення як поширених, так і рідкісних атак в умовах

аналізу реального мережевого трафіку. Очікується зростання показника Recall для міноритарних класів, без істотного погіршення точності для мажоритарного класу. При цьому прогнозується отримання кращої здатності розрізняти близькі за характеристиками типи атак, що є слабким місцем при використанні окремих моделей без попереднього перебалансування.

Експериментальні дослідження

Перша етап експерименту полягає у аналізі набору даних та виявленні основних його характеристик. В рамках дослідження було використано публічний датасет мережевого трафіку CIC-IDS2017 [13], який був створений Канадським інститутом кібербезпеки (CIC) для дослідження виявлення вторгнень, що містить як нормальну активність, так і різноманітні типи мережевих атак у контрольованому середовищі. З нього було взято два набори даних, кожен з яких відповідає за певний день тижня та час, а також за конкретний вид загрози. Для підвищення об'єктивності дослідження було прийнято рішення обрати дослідження стану мережі у четвер, а саме набори Thursday-WorkingHours-Morning-WebAttacks та Thursday-WorkingHours-Afternoon-Infilteration. Перший набір даних є багатокласовим та зберігає дані як нормально стану мережі, так і опис характеристик мережі під час 3-х видів веб-атак, а саме: Bruteforce (метод атаки, при якому зловмисник випробовує всі можливі комбінації логінів і паролів для отримання несанкціонованого доступу до системи), XSS (міжсайтовий скриптинг, що дозволяє виконати шкідливий код на вебсторінці) та SQL Injection (використання шкідливих SQL-запитів). Цей датасет дозволяє дослідити особливості багатокласової класифікації в предметній галузі. Другий набір містить дані, що відповідають нормальному стану мережі, а також окремі приклади атаки Infilteration (проникнення, при якому дії зловмисника спрямовані на отримання несанкціонованого доступу до окремої комп'ютерної системи або мережі). Таким чином, цей датасет дозволяє дослідити особливості бінарної класифікації в предметній галузі.

Для дослідження характеристик наборів даних було виконано аналіз розподілу міток, а також розкиду значень атрибутів та оцінки кореляції між ними. Для останнього показника використовувалось значення сумарного показника кореляції.

На рис. 1–3 представлені результати дослідження характеристик набору даних, що відповідає за атаку типу Infiltration. На рис. 4–6 представлені результати дослідження характеристик набору даних, що відповідає за декілька видів вебатак.

Таким чином, як видно з наведених графічних залежностей, обидва набори даних мають певні вади: в першу чергу велику кореляцію окремих атрибутів, широкий розкид значень, а також дисбаланс у мітках. В такому разі за результатами першого етапу експериментального дослідження можна стверджувати, що для вдалого виконання як бінарної, так і багатокласової класифікації необхідно буде застосувати розглянуті раніше підходи для підвищення ефективності класифікації з використанням даних, що мають зазначені вади.

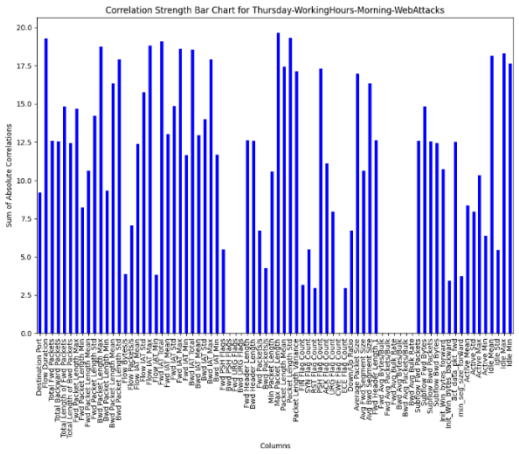

Рис. 1. Сумарний показник кореляції з-поміж атрибутів в наборі даних, що відповідає за атаку типу Infiltration

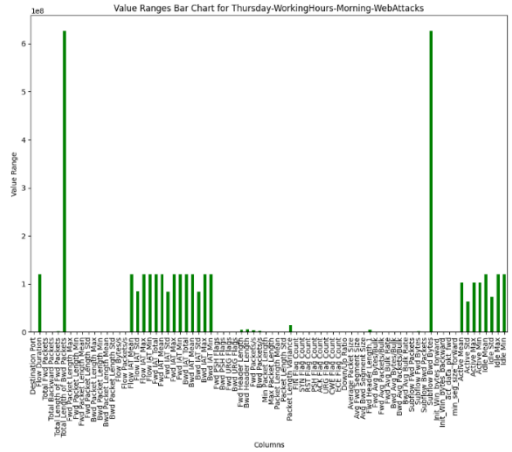

Рис. 2. Розкид значень атрибутів набору даних, що відповідає за атаку типу Infiltration

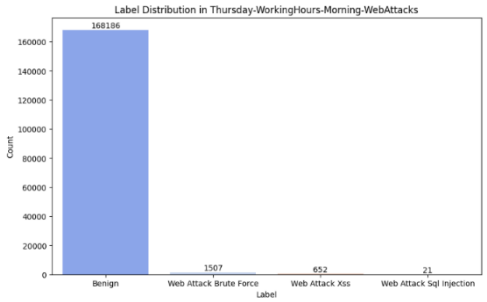

Рис. 3. Розподіл міток класів набору даних, що відповідає за атаку типу Infiltration

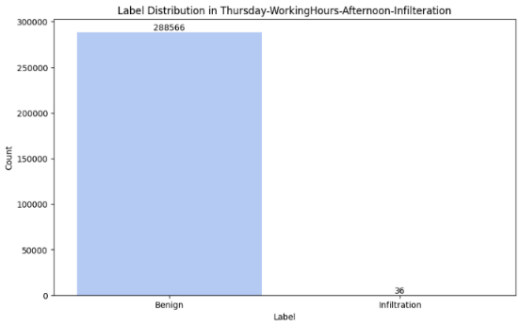

Рис. 4. Розподіл міток класів набору даних, що відповідає за декілька видів вебатак

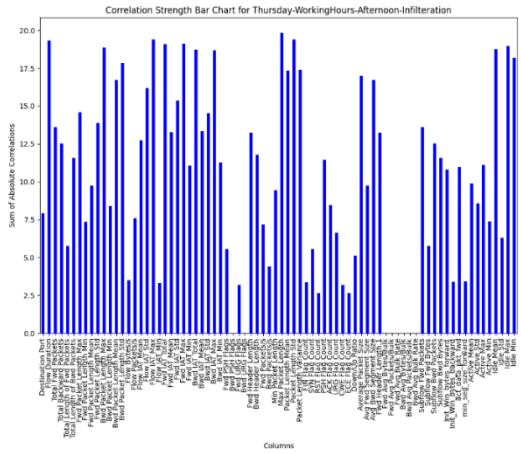

Рис. 5. Сумарний показник кореляції з-поміж атрибутів в наборі даних, що відповідає за декілька видів вебатак

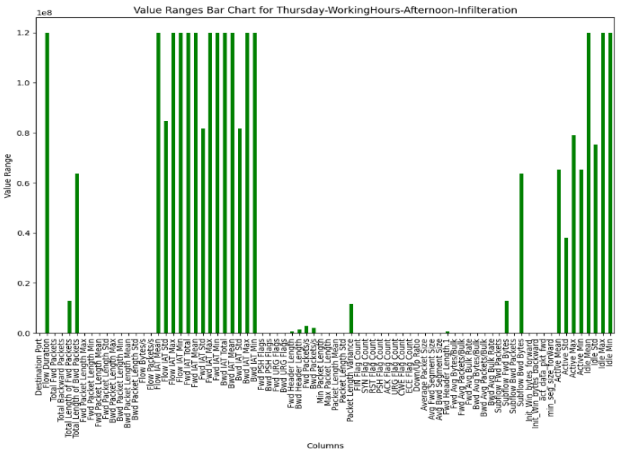

Рис. 6. Розкид значень атрибутів набору даних, що відповідає за декілька видів вебатак

Другий етап експериментального дослідження полягав у оцінці ефективності класифікації різними методами машинного навчання без використання додаткових технік зі стандартними налаштуваннями. Для цього були відібрані класифікатори, що працюють на принципах логістичної регресії, методу опорних векторів, методу К-найближчих сусідів та з використанням дерев рішення. Для кожного з них було використано два раніше розглянутих наборів даних, а також за результатами навчання та тестування побудовано графіки, що відображають основні показники якості, а саме Recall та F1-Score. На рис. 7 представлено результати оцінки якості класифікації з використанням методів машинного навчання для набора даних, що відповідає за атаку типу Infiltration або «проникнення» та за декілька видів вебатак.

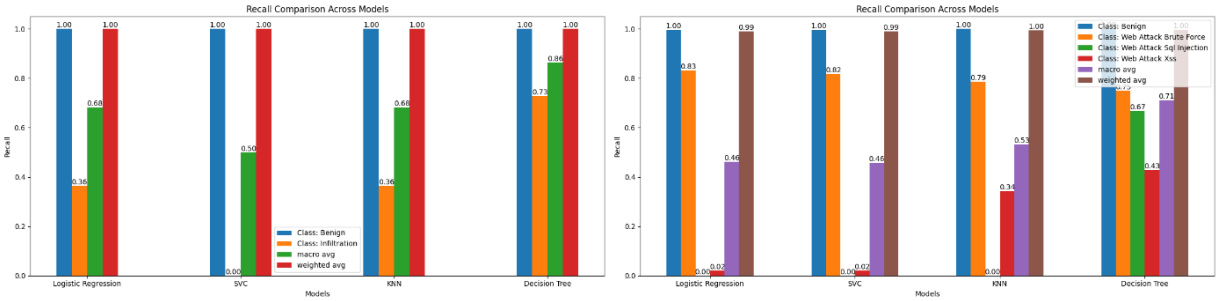

Рис. 7. Значення метрики Recall для результатів класифікації методами машинного навчання для набору даних, що відповідає за атаку типу Infiltration (зліва) та за декілька видів вебатак (справа)

За результатами дослідження отримано, що для задачі виявлення атаки типу Infiltration усі базові моделі, зокрема Logistic Regression, SVC, KNN та Decision Tree, досягають максимального значення метрики Recall для нормального трафіку. Проте значення цього показника якості для класу атаки суттєво нижчі: моделі Logistic Regression та KNN продемонстрували результат 0.36, Decision Tree – 0.73 (найкращий показник), а SVC взагалі не зміг побудувати модель. У випадку мультикласової задачі спостерігається ще помітніше падіння Recall для міноритарних класів. Наприклад, для атаки типу SQL Injection 3 з 4 моделей не змогли отримати значення, більше за 0, а модель Decision Tree продемонструвала результат на рівні 0.67 (найкращий результат). Для атаки типу XSS жодна з моделей не продемонструвала результат більший за 0.43. Трішки кращі результати були отримані при класифікації атаки Brute Forece – до 0.83, що пов'язано з більшою кількістю екземплярів навчальної вибірки. Лише для класу Benign (нормальний стан мережі) усі моделі досягають стабільно високих значень. Такі результати свідчать про те, що в умовах суттєвого дисбалансу в датасеті базові методи машинного навчання неспроможні ефективно виявляти малочисельні класи атак, а тому навіть незначна зміна в їхньому характері може спричинити істотне зниження якості виявлення. Таким чином, для досягнення прийнятної якості класифікації атак потрібно впроваджувати стратегії перебалансування вибірки та вдосконалення архітектури моделей.

Третій етап дослідження полягав у дослідження ефективності методів підвищення якості бінарної та багатокласової класифікації з використанням деяких підходів в умовах дисбалансу набору даних. Перший підхід полягав у використані попередньої обробки даних з перебалансуванням, що передбачало застосування методу SMOTE. Результати оцінки якості класифікації для обох наборів даних представлені на рис. 8.

Наступний підхід полягав у використання ансамблевих класифікаторів, що об'єднують в собі прості моделі машинного навчання. На рис. 9 представлені результати оцінки якості класифікації з використанням ансамблевих методів машинного навчання для обох наборів даних.

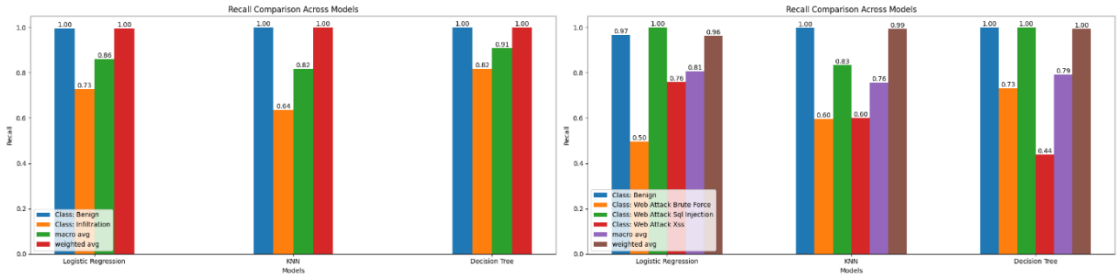

Рис. 8. Значення метрики Recall для результатів класифікації методами машинного навчання з застосуванням методу SMOTE для набору даних, що відповідає за атаку типу Infiltration (зліва) та за декілька видів атак (справа)

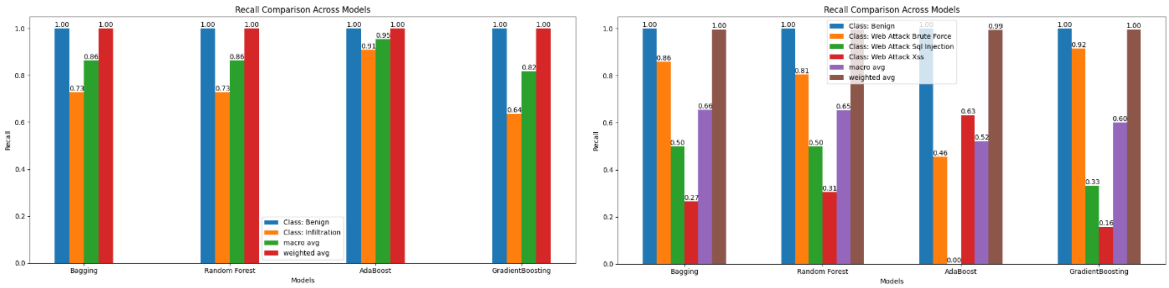

Рис. 9. Значення метрики Recall для результатів класифікації ансамблевими методами машинного навчання для набору даних, що відповідає за атаку типу Infiltration (зліва) та за декілька видів атак (справа)

Отримані результати продемонстрували, що балансування вибірки з використанням методу SMOTE для задач бінарної та мультикласової класифікації дозволяє збільшити значення метрики Recall для атаки типу Infiltration: Logistic Regression до 0.73, KNN до 0.64, Decision Tree до 0.82. Водночас Recall для класу Benign залишився стабільно високим для всіх моделей. У випадку мультикласової задачі також спостерігається покращення результатів для класів мережевих атак. Наприклад, для XSS метрика Recall зросла із майже нульового значення до 0.76, а для Brute Force – до 0.86 та 0.92 у випадку Bagging та Gradient Boosting відповідно. Проте не вдалося отримати кращих результатів для SQL Injection. Застосування ансамблевих методів класифікації, що поєднують кілька простих моделей машинного навчання при виявленні атаки типу Infiltration продемонструвало високі значення метрики Recall, наприклад 0.86 для Bagging та 0.91 для AdaBoost при еталонних показниках для нормального трафіку. У задачі мультикласової класифікації ансамблеві моделі також продемонстрували кращі результати порівняно з базовими: для атак типу XSS та Brute Force Recall зріс до 0.63 та 0.92 відповідно, проте залишився низьким для атаки SQL Injection.

Отже, обидва підходи забезпечили збільшення обраної метрики якості для більшості класів атак, проте не вдалося досягти таких результатів для SQL Injection класу. Саме тому було прийнято рішення перевірити, наскільки їх комбінація може вплинути на якість класифікації.

На рис. 10 представлені результати оцінки якості класифікації з комплексним використанням методу синтетичного перебалансування та ансамблевих класифікаторів.

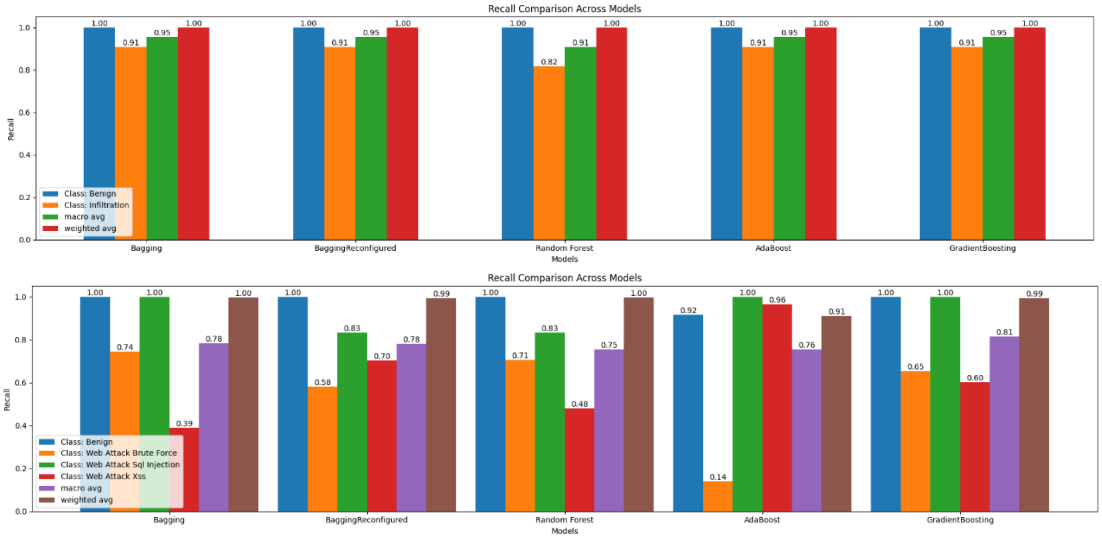

Рис. 10. Значення метрики Recall для результатів з комплексним використанням методу синтетичного перебалансування та ансамблевих класифікаторів для набору даних, що відповідає за атаку типу Infiltration (зверху) та за декілька видів атак (знизу)

Реалізація комплексного поєднання двох підходів (перебалансування даних методом SMOTE та подальша класифікація з використанням ансамблевих моделей) дозволила отримати кращі показники якості для обох задач. Результати бінарної класифікації продемонстрували найкращі значення Recall для всіх моделей для нормального трафіку – 1.00, а також від 0.82 до 0.91 для атаки Infiltration, що призвело до отримання найбільших значень Recall macro average на рівні 0.91-0.95. У випадку мультикласової задачі спостерігається вдосконалене загальне покриття для міноритарних класів з атаками: Recall для SQL Injection досяг 1.00 (Bagging, AdaBoost, Gradient Boosting), Brute Force – до 0.74 (Bagging), а XSS – до 0.96 у моделі AdaBoost. Крім того, вдалося досягти значення Recall macro avg на рівні 0.78 для моделі Bagging та 0.81 для моделі Gradient Boosting.

Ці результати підтверджують, що поєднання методів синтетичного перебалансування з ансамблевими класифікаторами є найбільш ефективною стратегією, яка дозволяє досягти високої чутливості як для основного класу, так і для рідкісних типів атак, представлених міноритарними екземплярами. Такий підхід забезпечує загальну стабільність і узгодженість результатів. Це критично важливо в умовах високої варіативності та дисбалансу мережевих атак, особливо коли система має бути чутливою до декількох рідкісних сценаріїв з обмеженим представленням у тренувальних даних.

Оцінка результатів та формування рекомендацій для подальших досліджень

Проведений аналіз показав, що при значному дисбалансі даних традиційні алгоритми машинного навчання втрачають здатність ефективно розпізнавати рідкісні типи мережевих атак. Це особливо помітно у випадку з класами, які представлені мінімальною кількістю прикладів у навчальній вибірці. Початкове значення метрики Recall для Infiltration для деяких методів становило лише 0.36 та 0.73 для дерев рішення, тоді як для XSS цей показник не перевищував 0.43, а для SQL Injection 0.67. Більш високі результати було отримано для Brute Force атаки, де максимально значення Recall склало 0.83.

Застосування методу SMOTE для штучного збільшення обсягу даних міноритарних класів дало відчутний ефект. Recall для Infiltration підвищився до 0.73 у Logistic Regression та до 0.82 у Decision Tree. Для XSS показник зріс до 0.76, а для Brute Force – до 0.92 у моделі Gradient Boosting. При цьому значення для класу SQL Injection також значно збільшилося, а для класу Benign залишилося стабільно високим, що свідчить про відсутність негативного впливу перебалансування на основний клас.

Покращення результатів також вдалося досягти завдяки поєднанню SMOTE з ансамблевими підходами. У задачі бінарної класифікації Recall для нормального трафіку досягав 1.00, а для Infiltration коливався в межах 0.82-0.91, що забезпечило Recall macro average на рівні 0.91-0.95. У мультикласовій класифікації також спостерігалося помітне зростання показників: Recall для SQL Injection піднявся до 1.00 (у моделях Bagging, AdaBoost та Gradient Boosting), для XSS – до 0.96 (AdaBoost), для Brute Force – до 0.74 (Bagging). Значення Recall macro average при цьому становили 0.78–0.81, що є істотним покращенням у порівнянні з базовими моделями.

Отримані результати свідчать, що поєднання синтетичного перебалансування та ансамблевих методів є дієвою стратегією для задач виявлення мережевих атак в умовах дисбалансу класів. Подальші дослідження доцільно спрямувати на глибше вивчення поведінки таких моделей у сценаріях зі змінними характеристиками мережевого трафіку та у випадках, коли нові типи атак з'являються в процесі або після навчання класифікаційної системи. Окрему увагу варто приділити оптимізації параметрів SMOTE для кожного конкретного класу. Також перспективним є застосування гібридних ансамблів, що поєднують сильні сторони різних алгоритмів, аби зменшити ризик падіння якості для окремих класів при загальному зростанні середніх показників.

Зрештою, метою подальших робіт має бути створення систем, здатних підтримувати високу точність і стабільність результатів навіть у складних умовах, коли розподіл даних є нерівномірним, а характер атак постійно змінюється.

Висновки

В рамках дослідження було перевірено декілька способів покращення роботи класифікаторів для виявлення різних мережевих атак. Для цього використовувалися набори даних з дисбалансом класів-атак, а також методи попередньої обробки з перебалансуванням та різні методи машинного навчання, у тому числі ансамблеві алгоритми. Найкращі результати дало поєднання синтетичного балансування (SMOTE) з ансамблевими моделями, такими як чистий Bagging, Random Forest і Gradient Boosting. Це дозволило значно підвищити чутливість до рідкісних атак, зокрема XSS та SQL Injection, а також утримати високу точність для нормального трафіку.

Порівняно з базовими моделями без перебалансування, запропонований комплексний підхід покращив значення метрики Recall на 18% для атаки Infiltration, на 33% для атаки SQL Injection та до 53% для атаки XSS у порівнянні з базовими моделями машинного навчання без додаткового перебалансування вхідних даних. Отримані результати підтверджують, що комбінування спеціалізованих методів обробки даних із сучасними ансамблевими алгоритмами є ефективним рішенням для виявлення мережевих атак у реальних умовах, де дані по різним станам часто є нерівномірно розподіленими.

Список літератури

1. Akamai Technologies, Inc., "Fighting the Heat: EMEA's Rising DDoS Threats," State of the Internet / Security, vol. 10, no. 02, 2024. [Online]. Available: https://www.akamai.com/content/dam/site/en/documents/state-of-the-internet/2023/2024/akamai-soti-2024-emea-ddos-report.pdf. Accessed: Jul. 27, 2025.

2. С. М. Лисенко & Р. В. Щука, "Аналіз методів виявлення шкідливого програмного забезпечення в комп'ютерних системах," Вісник Хмельницького національного університету. Технічні науки, no. 2, pp. 101–107, 2020, doi: 10.31891/2307-5732-2020-283-2-101-107.
3. L. Diana, P. Dini, and D. Paolini, "Overview on Intrusion Detection Systems for Computers Networking Security," Computers, vol. 14, no. 3, p. 87, 2025, doi: 10.3390/computers14030087.
4. Y. Otoum and A. Nayak, "AS-IDS: Anomaly and Signature Based IDS for the Internet of Things," Journal of Network and Systems Management, vol. 29, p. 23, 2021, doi: 10.1007/s10922-021-09589-6.
5. N. Naik, P. Jenkins, N. Savage, L. Yang, K. Naik and J. Song, "Embedding Fuzzy Rules with YARA Rules for Performance Optimisation of Malware Analysis," 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, UK, 2020, pp. 1-7, doi: 10.1109/FUZZ48607.2020.9177856..
6. О. Заковоротний та А. Хулап, "АНАЛІЗ МОДЕЛЕЙ ВИЯВЛЕННЯ ВТОРГНЕНЬ НА ОСНОВІ НЕЙРОМЕРЕЖ У СИСТЕМАХ ІНТЕРНЕТУ РЕЧЕЙ," Системи управління, навігації та зв'язку. Збірник наукових праць, 2(80), сс. 125-131, doi: 10.26906/SUNZ.2025.2.125.
7. С. Гавриленко, В. Зозуля, В. Омельченко, "ДОСЛІДЖЕННЯ МЕТОДІВ ПІДВИЩЕННЯ ЯКОСТІ КЛАСИФІКАЦІЇ НА НЕЗБАЛАНСОВАНИХ ДАНИХ", Системи управління навігації та зв'язку. Збірник наукових праць, 2(72), сс. 87-91, 2023, doi: 10.26906/SUNZ.2023.2.087.
8. O. Hornostal and S. Gavrylenko, "Application of heterogeneous ensembles in problems of computer system state identification," Advanced Information Systems, vol. 7, no. 4, pp. 5–12, Dec. 2023, doi: 10.20998/2522-9052.2023.4.01.
9. O. Hornostal and S. Gavrylenko, "Development of a method for identification of the state of computer systems based on bagging classifiers," Advanced Information Systems, vol. 5, no. 4, pp. 5–9, Dec. 2021, doi: 10.20998/2522-9052.2021.4.01.
10. H. Chandrasekaran, K. Murugesan, S. C. Mana, B. K. U. A. Barathi, and S. Ramaswamy, "Handling imbalanced data in intrusion detection using time weighted Adaboost support vector machine classifier and crossover boosted Dwarf Mongoose Optimization algorithm," Applied Soft Computing, vol. 167, p. 112327, 2024.
11. R. Vaishali, "A Hybrid Gradient Boost Model for Intrusion Detection," in Proc. 2023 7th Int. Conf. on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 1106–1111, doi: 10.1109/ICCMC56507.2023.10084018.
12. V. Chelak, O. Hornostal, Y. Chelak, and S. Gavrylenko, "ADVANCED METHODS FOR CLASSIFICATION QUALITY ASSESSMENT LEVERAGING ROC ANALYSIS AND MULTIDIMENSIONAL CONFUSION MATRIX," Advanced Information Systems, vol. 9, no. 1, pp. 24–34, 2025, doi: 10.20998/2522-9052.2025.1.03.
13. R. Panigrahi and S. Borah, "A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems," International Journal of Engineering & Technology, vol. 7, no. 3.24, pp. 479–482, 2018, doi: 10.14419/ijet.v7i3.24.22797.

Received (Надійшла) 23.06.2025
Accepted for publication (Прийнята до друку) 13.08.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Горносталь Олексій Андрійович – PhD, асистент кафедри "Комп'ютерна інженерія та програмування", Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Oleksii Hornostal – PhD, Assistant Professor of Department of "Computer Engineering and Programming", National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;
e-mail: gornostalaa@gmail.com; ORCID ID: https://orcid.org/0000-0001-5820-9999;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57189040595.

Челак Віктор Володимирович – PhD, доцент кафедри "Комп'ютерна інженерія та програмування", Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Viktor Chelak – PhD, Associate Professor of Department of "Computer Engineering and Programming", National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine;
e-mail: victor.chelak@gmail.com; ORCID ID: https://orcid.org/0000-0001-8810-3394;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57216331944&origin=resultslist.

Classification of network attacks using machine learning methods in conditions of training data imbalance

Oleksii Hornostal, Viktor Chelak

Abstract. **The object of the research** is the process of detecting network intrusions. **The subject of the research** is methods for classifying network intrusions**. The aim of the research** is to improve the quality and speed of ensemble classifiers in network intrusion classification problems under conditions of imbalance in training data. **Methods used**: machine learning methods, data preprocessing methods, ensemble classifiers, rebalancing method by synthetic minority augmentation. **Results obtained**: the effectiveness of using various approaches for classifying network intrusions under conditions of class imbalance in the training sample was investigated. A comprehensive approach was proposed, which involves preprocessing data using the SMOTE method for synthetic balancing of the training sample, as well as its subsequent analysis using ensemble machine learning models, which made it possible to improve the classification performance of minority classes. The best results were obtained when combining SMOTE with ensemble models, in particular Bagging, Gradient Boosting and AdaBoost. **Conclusions**. Based on the results of the study, an improved approach to network traffic classification was proposed, which combines pre-sampling by the SMOTE method with ensemble algorithms bagging and boosting. The combined use of these methods allowed to improve the value of the Recall metric for minority classes. Overall, the proposed approach provided an improvement in the classification quality: 18% for the Infiltration attack, 33% for the SQL Injection attack and up to 53% for the XSS attack compared to basic machine learning models without additional rebalancing of input data.

Keywords: network attacks, classification, machine learning, ensembles, bagging, boosting, quality metrics, SMOTE.