УДК 004.056.53:004.421:004.85
DOI: 10.31673/2409-7292.2025.030315

Гайдур Г.І., Гамза Д.Є.

ГІБРИДНИЙ МЕТОД ВИЯВЛЕННЯ ШКІДЛИВОЇ АКТИВНОСТІ НА ОСНОВІ СТЕКІНГ-АНСАМБЛЮ КЛАСИФІКАТОРІВ

У статті представлено гібридний метод виявлення шкідливої активності в інформаційних системах організацій, розроблений на основі ансамблевого підходу з використанням стекінгу. Запропонована архітектура поєднує класичні алгоритми машинного навчання (SVM, Random Forest, kNN) та сучасні високоефективні бустингові моделі (XGBoost, LightGBM, CatBoost), тоді як роль мета-класифікаторів виконують логістична регресія, XGBoost, Gradient Boosting та Random Forest. Такий підхід забезпечує інтеграцію сильних сторін різних методів, що дозволяє суттєво підвищити точність класифікації, стійкість до шуму та здатність системи до узагальнення. Особливу увагу приділено попередній обробці даних, яка включає видалення нерелевантних ознак, нормалізацію числових характеристик, балансування диспропорції класів за допомогою алгоритму SMOTE, зниження розмірності із застосуванням PCA та інженерію часових ознак. Застосування цих методів дозволило зменшити ризик перенавчання, прискорити обчислення та зберегти інформативність ключових характеристик трафіку. Для вибору оптимальних моделей застосовано два методи: побудову Pareto-фронту та евристичну фільтрацію за середніми значеннями метрик, що дало змогу забезпечити збалансоване співвідношення між точністю, F1-мірою та швидкодією. Експериментальна перевірка запропонованого підходу проведена на одному з найбільш репрезентативних датасетів у сфері кібербезпеки – CSE-CIC-IDS2018. Отримані результати засвідчили досягнення точності на рівні 98,07%, F1-міри 96,57% та середнього часу прогнозу 7,16 мс, що відповідає сучасним вимогам до IDS, здатних функціонувати в умовах високого навантаження в реальному часі. Запропонована система продемонструвала кращу ефективність порівняно з поодинокими моделями, що підтверджує доцільність використання гібридних ансамблевих методів у завданнях кіберзахисту.

Ключові слова: загрози, виявлення вторгнень, гібридна класифікація, стекінг, кібербезпека, кіберзахист, машинне навчання, моделі, шкідлива активність.

Вступ та постановка проблеми

Стрімке зростання цифрової інфраструктури та ускладнення інформаційних систем організацій призвели до суттєвого розширення поверхні атак для кіберзагроз [1, 2, 3, 4]. Традиційні системи виявлення вторгнень (IDS), що базуються переважно на сигнатурному або аналізі на основі правил, дедалі частіше виявляються неефективними у виявленні нових, поліморфних або прихованих атак. Їхня залежність від заздалегідь визначених шаблонів обмежує здатність до узагальнення та адаптації до нових сценаріїв поведінки.

У відповідь на ці обмеження, методи машинного навчання (ML) набули широкого поширення як основа для побудови адаптивних та інтелектуальних IDS. Такі моделі здатні навчатися на історичних даних, виявляти приховані закономірності та ідентифікувати аномалії в режимі реального часу. Проте використання окремих класифікаторів має низку недоліків: метод опорних векторів (SVM) чутливий до масштабування ознак, метод k-ближчих сусідів (kNN) погано масштабується у високовимірних просторах, а глибокі нейронні мережі вимагають значних обчислювальних ресурсів і великих обсягів помічених даних [5, 6].

Ансамблеві методи, зокрема стекінг (stacking), демонструють високу ефективність у подоланні цих обмежень [7]. Стекінг дозволяє поєднувати прогнози кількох гетерогенних моделей за допомогою мета-класифікатора, що забезпечує покращену узагальнюючу здатність, стійкість до шуму та зниження ризику перенавчання.

Водночас, навіть у межах стекінг-підходу залишаються відкритими кілька важливих наукових і прикладних питань щодо балансу між точністю класифікації та швидкодією моделі для використання в реальному часі, адаптування моделі до змін у поведінці мережевого трафіку без повного перенавчання та зменшення впливу дисбалансу класів та надмірної розмірності ознак на якість класифікації [8].

З метою вирішення цих проблем у даному дослідженні запропоновано гібридну стекінг-архітектуру для виявлення шкідливої активності, яка поєднує класичні (SVM, Random Forest, kNN) та бустингові моделі (XGBoost, LightGBM, CatBoost), із логістичною регресією як мета-класифікатором. Система включає сучасні методи попередньої обробки даних, зокрема

SMOTE для балансування класів, PCA для зниження розмірності, інженерію часових ознак та адаптивне зважування [9, 10, 11].

Метою дослідження є розробка, реалізація та експериментальна перевірка масштабованої, високоточної та низьколатентної системи виявлення шкідливої активності, здатної функціонувати в умовах реального часу в корпоративних інформаційних середовищах. Запропонована модель валідується на датасеті CSE-CIC-IDS2018 – одному з найавторитетніших у сфері кібербезпеки – та порівнюється з сучасними підходами.

Мета статті

Метою цієї статті є розробка, реалізація та експериментальна перевірка гібридної стекінг-архітектури для виявлення шкідливої активності в інформаційних системах, яка поєднує точність, адаптивність і здатність до роботи в режимі реального часу. Запропонована модель базується на поєднанні гетерогенних класифікаторів машинного навчання (SVM, Random Forest, kNN, XGBoost, LightGBM, CatBoost) із логістичною регресією як мета-класифікатором. Особливу увагу приділено попередній обробці даних, балансуванню класів, зниженню розмірності, інженерії часових ознак та адаптивному зважуванню. Ефективність моделі оцінюється на основі репрезентативного датасету CSE-CIC-IDS2018 з урахуванням ключових метрик якості класифікації та продуктивності системи в умовах високого мережевого навантаження. Сучасні IDS досягають точності в межах 93-99%, F1 score 92-98 і час прогнозу 5-10 мс. Мета роботи буде вважатися досягнутою, якщо запропонований метод досягне або перевершить ці значення.

Виклад основного матеріалу

Основні положення створення методу на базі гібридної класифікації

Стекінг – це мета-ансамблевий метод, у якому комбінація прогнозів декількох базових класифікаторів подається як вхідні дані до мета-класифікатора (мета-моделі). Формально це можна представити наступним чином:

$$\text{Прогноз}=M(P1,P2,\ldots,PN), \quad (1)$$

де Pi – прогноз і-ї базової моделі, а M – мета-модель, що навчається на виходах цих базових моделей [7].

У запропонованій гібридній системі виявлення шкідливої активності використовуються наступні базові алгоритми, обрані з огляду на їхні доведені переваги у задачах класифікації та виявлення аномалій:

– метод опорних векторів (Support Vector Machine, SVM): відомий своєю здатністю ефективно розділяти класи навіть у високовимірному просторі за допомогою побудови оптимальної гіперплощини;

– випадкові ліси (Random Forest): ансамбль дерев рішень, що ефективно працює з табличними даними, стійкий до перенавчання та здатен обробляти велику кількість ознак;

– метод k-найближчих сусідів (k-Nearest Neighbors, KNN): простий та інтуїтивно зрозумілий алгоритм, що базується на вимірюванні відстані до найближчих елементів, добре зарекомендував себе в задачах класифікації;

– XGBoost (Extreme Gradient Boosting): високоефективна та оптимізована реалізація градієнтного бустингу, відома своєю швидкістю та точністю;

– LightGBM: ще одна швидка та пам'яттєво ефективна реалізація градієнтного бустингу, що добре масштабується на великих наборах даних;

– CatBoost: бустингова модель, що має вбудовану підтримку для обробки категоріальних ознак та менш чутлива до вибору гіперпараметрів.

Як мета-класифікатор обрано логістичну регресію, XGBoost, GradientBoost, Random Forest. Вибір мета-класифікаторів у стекінгу обумовлений поєднанням простоти, інтерпретованості та високої продуктивності. Логістична регресія ефективно агрегує

ймовірнісні прогнози та мінімізує перенавчання, тоді як XGBoost і Gradient Boosting забезпечують потужне моделювання складних залежностей і високу точність, а Random Forest додає стабільність і стійкість за рахунок агрегації багатьох дерев, що разом підвищує точність і надійність системи виявлення шкідливої активності [8]. Комбінування різних алгоритмів дозволяє компенсувати слабкі сторони кожного окремого методу, мінімізувати ризик перенавчання та, як наслідок, забезпечити суттєве покращення точності та стійкість системи до нових, раніше невідомих типів кібератак.

Методи оптимізації обробки даних гібридної моделі класифікації

Для підвищення ефективності гібридної моделі було реалізовано метод оптимізації оброби даних, який включає в себе низку процедур попередньої обробки даних (рис.1).

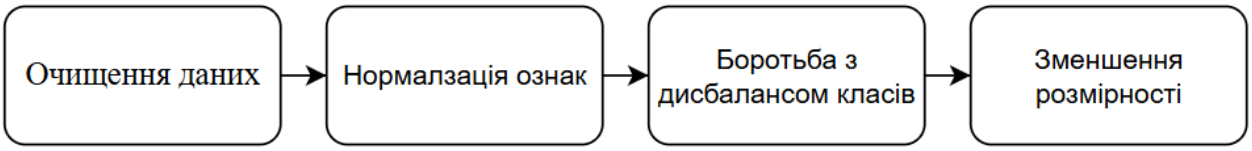

Рис. 1. Метод оптимізації обробки даних гібридної моделі класифікації

Основні кроки включають наступні.

Очищення даних. Видалено ознаки, які:

- мають нульове стандартне відхилення (тобто постійне значення);
- містять значну кількість пропущених значень (NaN);
- є унікальними для кожного запису (наприклад, ідентифікатор потоку).

Видалення ознак дозволяє зменшити шум у даних та знизити ризик перенавчання моделі.

Нормалізація ознак. Щоб усі ознаки мали однаковий масштаб, застосовано Min-Max нормалізацію. Так нормалізація дозволяє уникнути домінування ознак з великим числовим діапазоном при навчанні моделей, особливо важливо для SVM та KNN.

Боротьба з дисбалансом класів. У датасеті CSE-CIC-IDS2018 спостерігається значна диспропорція між кількістю прикладів «BENIGN» та атакуючих класів. Для балансування вибірки застосовано метод SMOTE (Synthetic Minority Oversampling Technique), який генерує синтетичні зразки для меншості.

Зменшення розмірності. Для уникнення перенавчання та прискорення класифікації застосовано метод головних компонент (PCA). Його мета – проектування даних на простір меншої розмірності з максимально збереженою дисперсією. Нехай – матриця ознак розмірності , тоді PCA виконує такі кроки:

1. Центрування ознак;
2. Обчислення коваріаційної матриці;
3. Власні вектори та власні значення;
4. Відбір головних компонент.

Цей підхід дозволяє зменшити кількість ознак без втрати суттєвої інформації, особливо важливо для моделей із високою обчислювальною складністю.

Виконання цих кроків дозволяє не тільки покращити швидкодію та точність моделі, але й мінімізувати ризик перенавчання (overfitting), що є критично важливим для узагальнюючої здатності системи у реальних умовах.

Результати дослідження

Результати класифікації окремих моделей

Оцінка проводилася за метриками точності (Accuracy), F1-міри та середнього часу прогнозу на один запис (мс). Кожна модель має власну специфіку та демонструє різну ефективність залежно від типу даних та обчислювальних ресурсів. Результати класифікації обраних моделей представлено в табл. 1.

© Г.І. Гайдур, Д.Є. Гамза. (2025) Гібридний метод виявлення шкідливої активності на основі стекінг-ансамблю класифікаторів. Сучасний захист інформації, 3(63), 20–26. https://doi.org/10.31673/2409-7292.2025.030315

Таблиця 1.

Результати класифікації окремих моделей

Модель	Accuracy	F1-score	Час прогнозу (мс)
SVM	0.91	0.89	11.3
Random Forest	0.94	0.93	8.2
KNN	0.9	0.88	14.7
XGBoost	0.95	0.94	10.1
LightGBM	0.94	0.93	6.8
CatBoost	0.95	0.94	7.5
Extra Trees	0.93	0.92	7.9
MLP	0.92	0.91	13.2
Logistic Regression	0.89	0.87	5.4
Naive Bayes	0.85	0.82	2.1

Сучасні IDS досягають точності в межах 93-99%, F1 score 92-98 і час прогнозу 5-10 мс [9]. Як видно з таб.1, деякі поодиноко взяті моделі вже досягають рівня діапазонів зазначених вище, такі як LightBGM, CatBoost, Random Forest, Extra Trees. Але модель зі стекінгом має підвищити значення поодиноких моделей і показати кращі можливості для виявлення шкідливої активності.

Порівняння гібридних моделей

Порівняння продуктивності різних комбінацій базових класифікаторів у рамках стекінг-підходу демонструє значне покращення порівняно з поодинокими моделями (табл. 2). Це підтверджує ефективність ансамблевого навчання для вирішення складних задач кібербезпеки для виявлення шкідливої активності в інформаційних системах.

Таблиця 2.

Порівняння гібридних моделей

Базові моделі	Мета-класифікатор	Accuracy	F1-score	Час прогнозу (мс)
XGBoost + CatBoost + LightGBM	XGBoost	0.9807	0.9657	7.16
XGBoost + CatBoost + Random Forest	XGBoost	0.9807	0.9657	7.57
XGBoost + CatBoost + Random Forest	Gradient Boosting	0.9797	0.9637	7.83
XGBoost + CatBoost + LightGBM	Gradient Boosting	0.9797	0.9637	7.4
XGBoost + CatBoost + LightGBM	Random Forest	0.9787	0.9647	7.48
XGBoost + CatBoost + Random Forest	Random Forest	0.9787	0.9647	7.91
XGBoost + CatBoost + LightGBM	Logistic Regression	0.9767	0.9617	6.91
XGBoost + CatBoost + Random Forest	Logistic Regression	0.9767	0.9617	7.31
CatBoost + LightGBM + Extra Trees	XGBoost	0.974	0.959	6.51
CatBoost + LightGBM + Extra Trees	Gradient Boosting	0.973	0.957	6.73

Критерії відбору оптимальної моделі

Для вибору найкращої комбінації моделей використано два підходи: порівняння із середніми значеннями метрик та побудову Pareto-фронту. Перший метод ґрунтується на принципі above-average rule, де комбінація вважається оптимальною, якщо перевищує середнє значення за всіма критеріями одночасно. Другий підхід – Pareto-фронт – дозволяє відібрати множину моделей, які не поступаються одна одній за всіма метриками і являють собою найкращі компроміси між точністю, F1-мірою та часом прогнозування. Це дає змогу обрати варіант, який найкраще відповідає вимогам конкретного застосування IDS у реальному часі.

Метод Pareto-фронту

Для аналізу ефективності комбінацій моделей було застосовано метод Pareto-фронту, який дозволяє визначити множину архітектур, що не поступаються одна одній одночасно за точністю, F1-мірою та часом прогнозування, і забезпечує оптимальний вибір залежно від вимог системи.

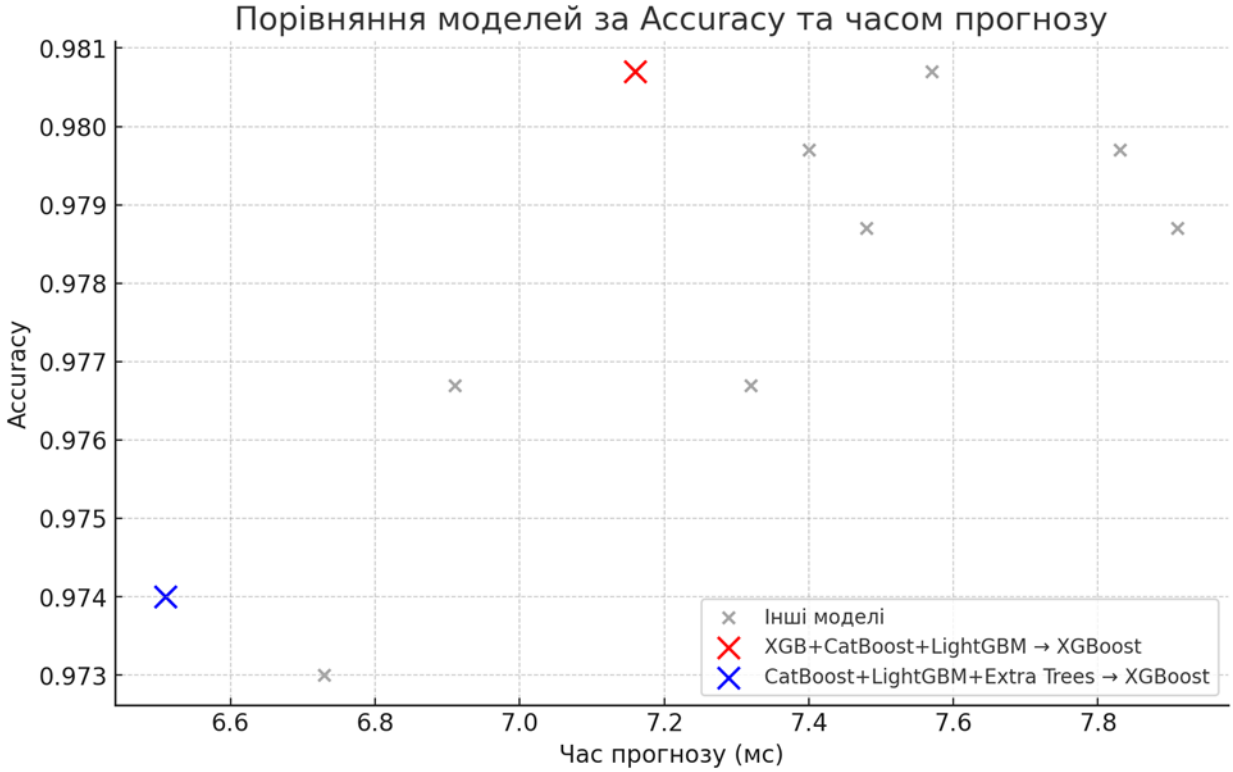

Рис.2. Порівняння моделей за точністю та часом прогнозу з кращими значеннями за методом Pareto-фронту

На основі даних було виявлено дві недоміновані комбінації (рис. 2):

1. Accuracy = 0.9807, F1-score = 0.9657, час = 7.16 мс (XGBoost, CatBoost, LightGBM з XGBoost як мета-класифікатором);

2. Accuracy = 0.974, F1-score = 0.959, час = 6.51 мс (CatBoost, LightGBM, Extra Trees з XGBoost як мета-класифікатором).

Перша комбінація забезпечує найвищу точність, друга — найменший час обробки. Обидві комбінації знаходяться на Pareto-фронті й можуть бути рекомендовані залежно від пріоритетів системи (точність або швидкодія).

Метод фільтрації за середніми значеннями

Цей евристичний метод ґрунтується на простому правилі: комбінація вважається кращою, якщо всі її показники перевищують середні значення по вибірці (рис. 3).

Рис.3. Порівняння моделей за точністю та часом прогнозу
з кращими значеннями за методом фільтрації за середніми значеннями

Для набору з 10 комбінацій було обчислено такі середні значення:

– середній Accuracy = 0.97842;
– середній F1-score = 0.96362;
– середній час прогнозу = 7.281 мс.

Серед усіх комбінацій лише одна відповідала цим умовам:

– Accuracy = 0.9807;
– F1-score = 0.9657;
– Час прогнозу = 7.16 мс.

Ця комбінація реалізована на базі моделей XGBoost, CatBoost, LightGBM з XGBoost як мета-класифікатор.

Висновки

Результати експериментів свідчать про досягнення точності до 98,07%, F1-міри 96,57% та середнього часу прогнозу 7,16 мс, що відповідає вимогам до IDS, здатних функціонувати в реальному часі. Для відбору оптимальних конфігурацій моделей використано два підходи: фільтрацію за середніми значеннями та побудову Pareto-фронту, що дозволило об'єктивно оцінити компроміс між точністю та швидкодією.

Отримані результати демонструють, що стекінгова модель здатна перевершити ефективність поодиноких класифікаторів та може бути практично реалізована у високонавантажених корпоративних інформаційних системах.

Перелік посилань:

1. Гайдур,Г. І., Гахов,С. О., Гамза,Д. Є.(2024). Модель виявлення шкідливої активності в інформаційній системі організації на основі гібридної класифікації. Сучасний захист інформації, 4(60), 30-38. DOI: 10.31673/2409-7292.2024.040003.

2. IDS 2018 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. (2023, December 21). Retrieved from https://www.unb.ca/cic/datasets/ids-2018.html.

3. Kaur, G., & Saini, H. S. (2023). Stacking ensemble learning for network intrusion detection systems. International Journal of Computer Applications, 184(12), 15–23. DOI: 10.29130/dubited.737211

4. Cisco. (2023). Annual Cybersecurity Report. Cisco Systems. https://www.cisco.com/c/m/en_us/products/security/cybersecurity-reports/cybersecurity-readiness-index.html.

5. Савченко В. А., Смолєв Є. С., Гамза Д. Є. Методика виявлення аномалій взаємодії користувачів з інформаційними ресурсами організації. Сучасний захист інформації. № 4 (2023). С. 6-12 DOI: 10.31673/2409-7292.2023.030101.

6. ENISA. (2023). ENISA Threat Landscape Report 2023. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023.

7. Murat U., Emine U., Mürsel O.(2021). A Stacking Ensemble Learning Approach for Intrusion Detection System. Düzce University Journal of Science & Technology, 184(12), 15–23. DOI:10.29130/dubited.737211.

8. Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: Improving accuracy through combining DOI: 10.2200/S00240ED1V01Y200912DMK002.

9. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. Wiley. DOI: 10.1002/9781118548387.

10. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31 DOI: 10.1016/j.jnca.2015.11.016

11. Pinto, A.; Herrera, L.C.; Donoso, Y.; Gutierrez, J.A. Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure. Sensors 2023, 23, 2415, DOI: 10.3390/s23052415.

Надійшла 06.06.2025