УДК 004.056.5:004.8 DOI 10.18372/2073-4751.87.21689

Пекуровський Г. В.,
orcid.org/0009-0006-1993-8615,
e-mail: hlib.pekurovskyi@npp.kai.edu.ua,
Щербак Т. Л.,
orcid.org/0000-0003-1170-8154,
e-mail: tetiana.shcherbak@npp.kai.edu.ua,
Климова А. С.,
orcid.org/0000-0002-4721-2241,
e-mail: asiia.klymova@npp.nau.edu.ua
Калашник М. О.,
orcid.org/0009-0003-8086-0325,
e-mail: 294776@stud.kai.edu.ua

ІНДЕКС СТАБІЛЬНОСТІ ТА ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СТІЙКОСТІ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ СИСТЕМ ВИЯВЛЕННЯ ВТОРГНЕНЬ В УМОВАХ ЗСУВУ КОНЦЕПЦІЇ

Національний університет «Київський авіаційний інститут»

Вступ

Системи виявлення вторгнень є критично важливим компонентом сучасної кібербезпеки. Вони дозволяють автоматично ідентифікувати аномалії в мережевому трафіку та потенційні атаки. У сучасних підходах все ширше застосовуються методи машинного навчання, які демонструють високу точність на історичних даних [1, 2].

Однак ключовою проблемою таких систем є їх залежність від припущення стаціонарності даних. У реальних мережах характер трафіку постійно змінюється: з'являються нові типи атак, змінюється поведінка користувачів, оновлюються протоколи передачі даних. Це призводить до явища зсуву концепції, тобто зміни розподілу даних у часі [3, 4].

Класичні метрики, такі як точність або F1-міра, не враховують часову деградацію моделей, що призводить до переоцінки їхньої реальної ефективності у динамічних середовищах.

У результаті моделі, які демонстрували високу точність під час навчання, поступово втрачають ефективність у реальних умовах. Метою роботи є аналіз стійкості моделей машинного навчання до зсуву концепції та введення узагальненої метрики оцінювання стабільності.

Наявні підходи до виявлення мережевих атак базуються на статистичних методах аналізу трафіку, глибоких нейронних мережах та ансамблевих моделях.

Однак більшість досліджень фокусується на статичних сценаріях, де навчальна та тестова вибірки мають однаковий розподіл. Проблема зсуву розглядається частково, проте відсутня уніфікована метрика, яка б дозволяла порівнювати різні моделі за стійкістю.

Методологія дослідження

Експериментальна система складається з наступних компонентів (рис. 1): модуль завантаження даних, модуль обробки даних, модуль навчання моделей, модуль симуляції зсуву концепції, модуль оцінювання та модуль візуалізації результатів.

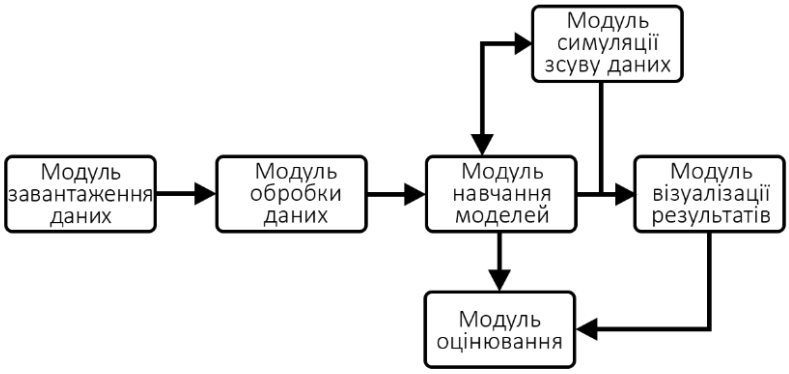


Рис. 1. Архітектура експериментальної системи дослідження

У дослідженні розглядаються наступні моделі машинного навчання: логістична регресія (Logistic Regression), випадковий ліс (Random Forest) та нейронна мережа у формі багатошарового перцептрону (Multi-layered Perceptron).

Дослідження проводяться на двох наборах даних, що дозволяє порівняти стійкість моделей до контрольованих змін характеристик даних у різних умовах:

1.NSL-KDD — покращена версія KDD'99 [5].

2.UNSW-NB15 — сучасний набір даних мережевого трафіку [6].

Попередня обробка датасетів передбачає кодування категоріальних ознак, нормалізацію числових ознак, поділ на навчальний та тестовий набори даних, а також уніфікацію форматів їхнього подання.

Досліджувані сценарії зміни характеристик даних включають дисбаланс класів, шум ознак та контрольне перемішування даних. Дисбаланс класів і шум ознак використовуються як сценарії контрольованої зміни даних, тоді як перемішування є контрольним сценарієм, який не повинен змінювати статистичні характеристики вибірки.

Використовуються наступні метрики оцінювання моделей машинного навчання: точність (accuracy), F1-міра (F1-score), характеристична крива оператора прийняття рішень (ROC-крива) та пов'язана із нею метрика площі під ROC-кривою (AUC), а також пропонується метрика індексу стабільності (stability index, SI).

Запропоновано метрику індексу стабільності, визначеного як відношення *F*1-міри на даних зі зсувом до F1-міри еталонної моделі:

$$SI = \frac{F1_{\text{зсув}}}{F1_{\text{еталон}}},$$

де $F1_{\text{зсув}}$ — F1-міра після зсуву концепції, а $F1_{\text{еталон}}$ — F1-міра до неї.

Для узагальненого порівняння стійкості моделі за сукупністю досліджуваних сценаріїв використовується середній індекс стабільності:

$$SI_{avg} = \frac{SI_{\text{дисбаланс}} + SI_{\text{шум}} + SI_{\text{перемішування}}}{3},$$

де SI_{avg} — середній індекс стабільності моделі за сукупністю досліджуваних сценаріїв; $SI_{\text{дисбаланс}}$ — індекс стабільності моделі після внесення дисбалансу класів; $SI_{\text{шум}}$ — індекс стабільності моделі після внесення шуму в ознаки; $SI_{\text{перемішування}}$ — індекс стабільності моделі після перемішування даних; 3 — кількість досліджуваних сценаріїв.

У роботі передбачено порівняння F1-міри до та після зміни характеристик даних, індексу стабільності для всіх моделей, деградаційних характеристик та порівняння результатів між двома датасетами.

Експериментальні результати дослідження датасету NSL-KDD

Аналіз результатів для датасету NSL-KDD (табл. 1–3, рис. 2–5) свідчить про високу базову якість класифікації всіх моделей (AUC ≈ 0.99–1.00, рис. 2). Перемішування даних не впливає на продуктивність (SI = 1.0), а дисбаланс класів чинить помірний вплив. Найбільшу деградацію викликає шум ознак: при цьому логістична регресія втрачає загальну якість класифікації ($F1_{\text{зсув}}$ = 0.876), а випадковий ліс демонструє найгостріше падіння індексу стабільності (SI = 0.762, рис. 4), що свідчить про високу чутливість цієї моделі до зашумлення ознак у досліджуваному сценарії. Нейронна мережа виявляє найвищу комплексну стійкість до всіх типів зсуву (SI ≥ 0.949).

Таблиця 1. Досліджувані метрики логістичної регресії для датасету NSL-KDD

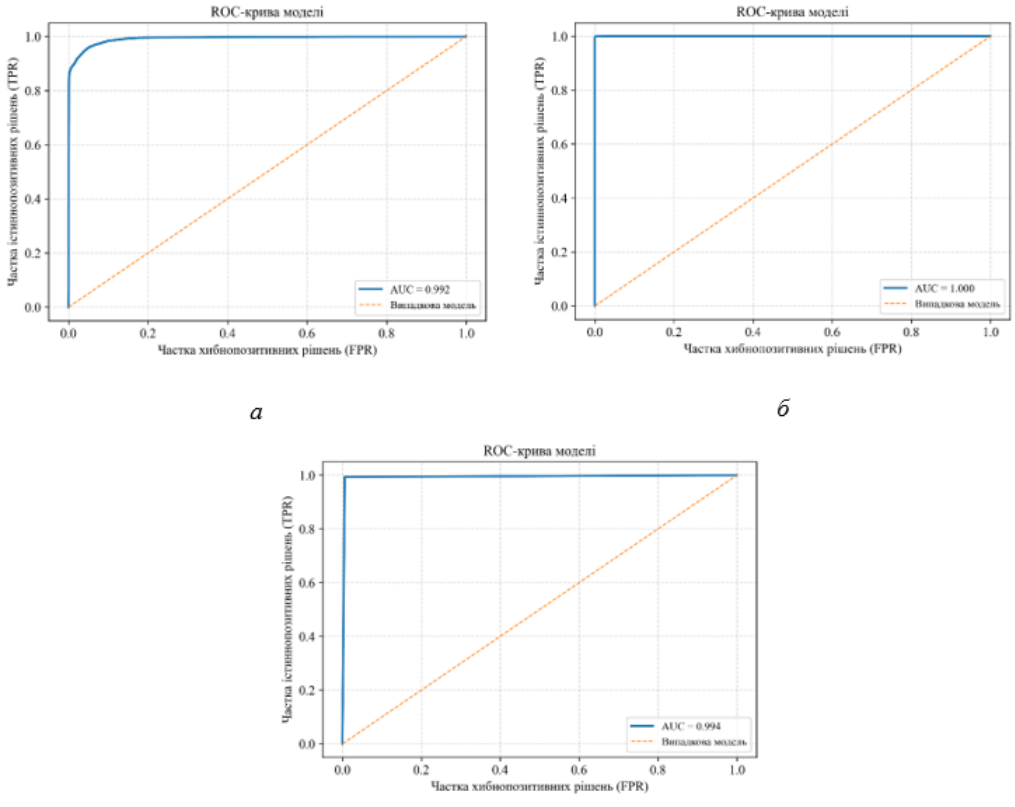
Типи зсуву		**Дисбаланс класів**	**Шум ознак**	**Перемішування даних**
Метрики	$F1_{еталон}$	0.949326	0.949326	0.949326
	$F1_{зсув}$	0.903094	0.876254	0.949326
	Індекс стабільності	0.951300	0.923028	1.000000

Таблиця 2. Досліджувані метрики випадкового лісу для датасету NSL-KDD

Типи зсуву		**Дисбаланс класів**	**Шум ознак**	**Перемішування даних**
Метрики	$F1_{еталон}$	0.998933	0.998933	0.998933
	$F1_{зсув}$	0.998294	0.760839	0.998933
	Індекс стабільності	0.999360	0.761651	1.000000

Таблиця 3. Досліджувані метрики нейронної мережі для датасету NSL-KDD

Типи зсуву		Дисбаланс класів	Шум ознак	Перемішування даних
Метрики	$F1_{еталон}$	0.993562	0.993562	0.993562
	$F1_{зсув}$	0.985749	0.943231	0.993562
	Індекс стабільності	0.992136	0.949342	1.000000



а *б* *в*

Рис. 2. ROC-крива: а — логістичної регресії; б — випадкового лісу; в —нейронної мережі

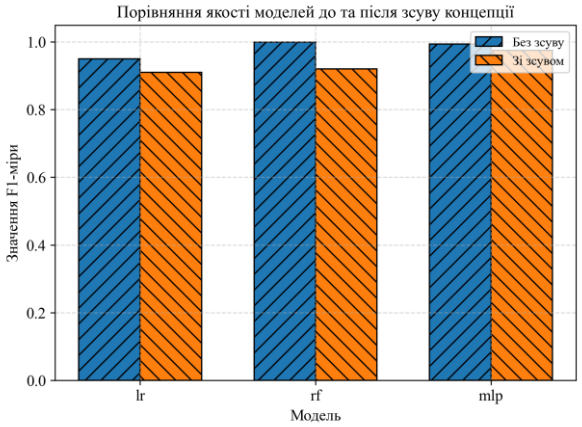


Рис. 3. Порівняльна характеристика F1-мір моделей (датасет NSL-KDD) до та після зміни характеристик даних: lr — логістична регресія, rf — випадковий ліс, mlp — нейронна мережа

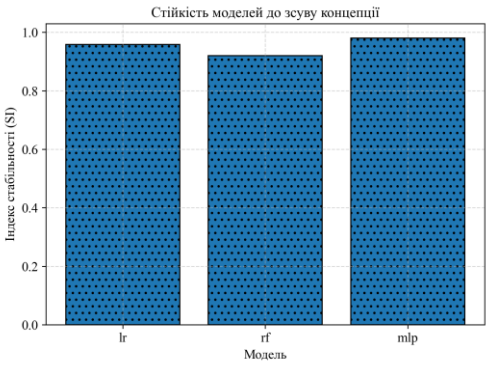


Рис. 4. Порівняльна характеристика індексів стабільності моделей (датасет NSL-KDD): lr — логістична регресія, rf — випадковий ліс, mlp — нейронна мережа

Експериментальні результати дослідження датасету UNSW-NB15

Отримані показники досліджуваних метрик стабільності для датасету UNSW-NB15, логістичної регресії, випадкового лісу та нейронної мережі показані у табл. 4–6 відповідно.

Таблиця 4. Досліджувані метрики логістичної регресії для датасету UNSW-NB15

Типи зсуву		Дисбаланс класів	Шум ознак	Перемішування даних
Метрики	$F1_{еталон}$	0.910696	0.910696	0.910696
	$F1_{зсув}$	0.837352	0.738863	0.910696
	Індекс стабільності	0.919464	0.811317	1.000000

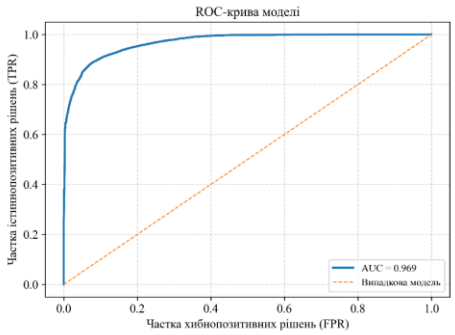
Таблиця 5. Досліджувані метрики випадкового лісу для датасету UNSW-NB15

Типи зсуву		Дисбаланс класів	Шум ознак	Перемішування даних
Метрики	$F1_{еталон}$	0.979153	0.979153	0.979153
	$F1_{зсув}$	0.771171	0.960930	0.979153
	Індекс стабільності	0.787590	0.981390	1.000000

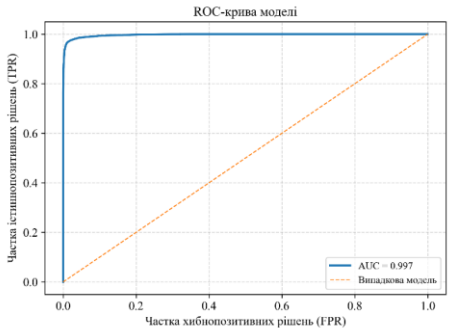
Таблиця 6. Досліджувані метрики нейронної мережі для датасету UNSW-NB15

Типи зсуву		Дисбаланс класів	Шум ознак	Перемішування даних
Метрики	$F1_{еталон}$	0.963879	0.963879	0.963879
	$F1_{зсув}$	0.776437	0.930041	0.963879
	Індекс стабільності	0.805533	0.964894	1.000000

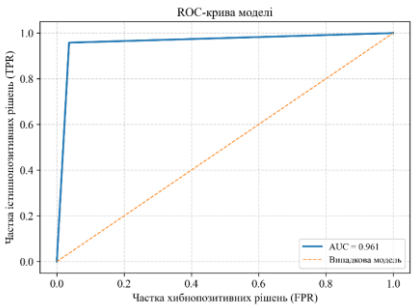
На датасеті UNSW-NB15 спостерігається загалом нижча стійкість моделей до зміни характеристик даних порівняно з NSL-KDD (табл. 4–6, рис. 5–7). Логістична регресія демонструє найнижчу еталонну F1-міру серед досліджуваних моделей (0.911) та найбільшу деградацію за умов шуму ознак (SI = 0.811). Водночас випадковий ліс та нейронна мережа зберігають високу стійкість до шуму ознак: їхні індекси стабільності становлять відповідно 0.981 та 0.965. Найбільш суттєва деградація випадкового лісу та нейронної мережі спостерігається за умов дисбалансу класів, де SI становить відповідно 0.788 та 0.806 (рис. 7). Підсумкове порівняння отриманих результатів для датасетів NSL-KDD та UNSW-NB15 подане у таблиці 7.



а



б



в

Рис. 5. ROC-крива: а — логістичної регресії; б — випадкового лісу; в — нейронної мережі (датасет UNSW-NB15)

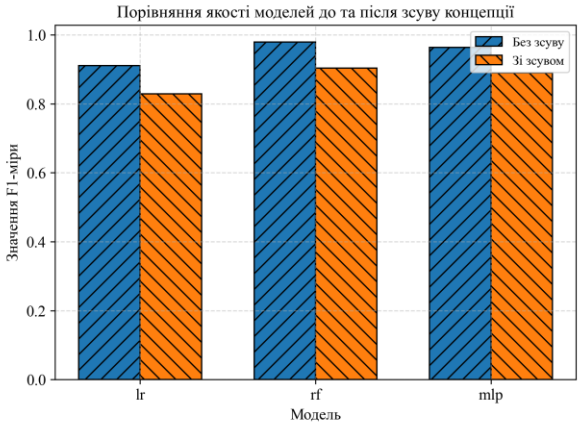


Рис. 6. Порівняльна характеристика F1-мір моделей (UNSW-NB15) до та після зміни характеристик даних: lr — логістична регресія, rf — випадковий ліс, mlp — нейронна мережа

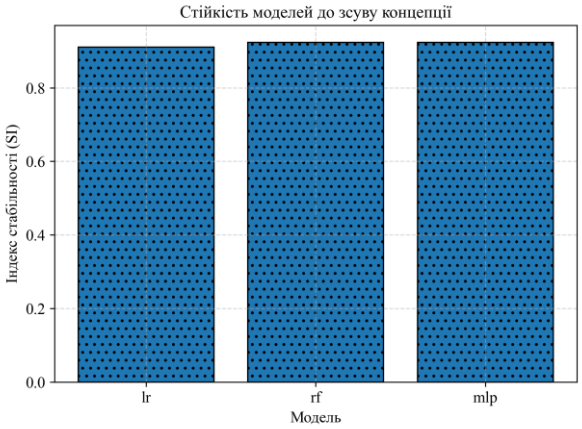


Рис. 7. Порівняльна характеристика індексів стабільності моделей (датасет UNSW-NB15): lr — логістична регресія, rf — випадковий ліс, mlp — нейронна мережа

Таблиця 7. Порівняльна характеристика результатів для датасетів NSL-KDD та UNSW-NB15

Характеристика	**NSL-KDD**	**UNSW-NB15**
Діапазон базової F1-міри	0.949–0.999	0.911–0.979
Середній ROC-AUC	0.995	0.976
Середній SI_{avg} по датасету	0.953	0.919
Найбільш виражена деградація	Шум ознак	Дисбаланс класів
Найвища базова F1-міра	Випадковий ліс (0.999)	Випадковий ліс (0.979)
Найвища середня стійкість	Нейронна мережа ($SI_{avg} \approx 0.980$)	Нейронна мережа ($SI_{avg} \approx 0.923$)
Контрольне перемішування	SI = 1.000 для всіх моделей	SI = 1.000 для всіх моделей
Особливість деградації	Найбільший вплив шуму на випадковий ліс	Найбільший вплив дисбалансу на випадковий ліс та нейронну мережу

Порівняльний аналіз показує, що NSL-KDD характеризується вищою загальною стійкістю моделей до досліджуваних змін характеристик даних. Для цього датасету найбільш виражена деградація спостерігається за умов шуму ознак. UNSW-NB15 демонструє нижчу загальну стійкість моделей, причому найбільш суттєве зниження якості випадкового лісу та нейронної мережі спостерігається за умов дисбалансу класів. Водночас за умов шуму ознак випадковий ліс та нейронна мережа зберігають високі значення індексу стабільності. Це свідчить про залежність характеру деградації не лише від типу моделі, а й від структури та характеристик конкретного датасету.

Обговорення результатів

Експерименти виявили залежність між типом зміни характеристик даних, структурою датасету та деградацією моделей. Контрольне перемішування даних не спричинило зміни F1-міри для жодної з досліджуваних моделей, що відповідає очікуваному результату для трансформації, яка не змінює склад та статистичні характеристики вибірки.

Для NSL-KDD найбільш виражена деградація спостерігається за умов шуму ознак. Найбільш чутливим до цього сценарію виявився випадковий ліс, для якого SI становить 0.762. Водночас багатошаровий перцептрон зберігає вищу стійкість до шуму ознак, демонструючи SI = 0.949.

Для UNSW-NB15 характер деградації відрізняється. Найбільш суттєве зниження якості випадкового лісу та нейронної мережі спостерігається за умов дисбалансу класів, де SI становить відповідно 0.788 та 0.806. Логістична регресія найбільше реагує на шум ознак, для якого SI дорівнює 0.811. Отже, один універсально найбільш руйнівний сценарій для всіх моделей і датасетів не визначається.

За сукупністю досліджуваних сценаріїв багатошаровий перцептрон демонструє найвище середнє значення індексу стабільності. Для NSL-KDD його SI_{avg} становить приблизно 0.980, а для UNSW-NB15 — приблизно 0.923. Водночас випадковий ліс забезпечує найвищу базову F1-міру на обох датасетах. Це демонструє відмінність між абсолютною якістю моделі та її стійкістю до зміни характеристик даних.

Таким чином, отримані результати показують, що оцінювання лише базової F1-міри не дає повної характеристики поведінки моделі за умов зміни характеристик даних. Використання SI як додаткової відносної метрики дозволяє кількісно оцінити ступінь деградації моделі відносно її базової продуктивності.

Висновки

У даній роботі досліджено стійкість моделей машинного навчання до контрольованих змін характеристик даних у задачі виявлення мережевих вторгнень. Розглянуто два сценарії зміни даних — дисбаланс класів та шум ознак — і контрольний сценарій перемішування даних. Для кількісної оцінки зміни якості використано індекс стабільності SI, який визначається як відношення F1-міри моделі в модифікованих умовах до її еталонного значення.

Експериментальні результати показали, що контрольне перемішування даних не впливає на якість моделей: для всіх досліджуваних моделей та обох датасетів SI становить 1.000.

Характер деградації моделей залежить від датасету та типу зміни даних. Для NSL-KDD найбільше зниження якості спостерігається за умов шуму ознак. Для UNSW-NB15 найбільш суттєва деградація випадкового лісу та нейронної мережі спостерігається за умов дисбалансу класів, тоді як логістична регресія найбільше реагує на шум ознак.

Випадковий ліс забезпечує найвищу базову F1-міру на обох датасетах. Водночас багатошаровий перцептрон демонструє найвище середнє значення індексу стабільності за сукупністю досліджених сценаріїв: $SI_{avg} \approx 0.980$ для NSL-KDD та $SI_{avg} \approx 0.923$ для UNSW-NB15.

Порівняння результатів показало, що середній індекс стабільності за всіма моделями та сценаріями становить приблизно 0.953 для NSL-KDD та 0.919 для

UNSW-NB15. Отже, в межах проведеного експерименту моделі демонструють вищу загальну стабільність на NSL-KDD.

Отримані результати показують, що статичної оцінки якості моделі недостатньо для повної характеристики її поведінки за умов зміни характеристик даних. Використання SI як додаткової відносної метрики дозволяє кількісно оцінити ступінь деградації моделі відносно її базової продуктивності та порівнювати стійкість різних алгоритмів у контрольованих сценаріях.

Перспективним напрямом подальших досліджень є розширення експерименту на часові послідовності даних, різні рівні та типи шуму, інші форми зміни розподілу, а також перевірка запропонованого індексу на додаткових датасетах і моделях.

Література

1. Бурков А. М. Машинне навчання без зайвих складнощів.. – Київ : Наш Формат, 2020. – 240 с.

2. Mitchell T. M. Machine Learning. – New York : McGraw-Hill, 1997. – 414 p.

3. Gama J. Knowledge Discovery from Data Streams. – Boca Raton : CRC Press, 2010. – 320 p.

4. Tsymbal A. The problem of concept drift: definitions and related work // Computer Science Department, Trinity College Dublin. – 2004. – 33 p.

5. NSL-KDD Dataset [Електронний ресурс]. – Режим доступу: https://www.unb.ca/cic/datasets/nsl.html

6. UNSW-NB15 Dataset [Електронний ресурс]. – Режим доступу: https://research.unsw.edu.au/projects/unsw-nb15-dataset.

Пекуровський Г. В., Щербак Т. Л., Климова А. С., Калашник М. О.

ІНДЕКС СТАБІЛЬНОСТІ ТА ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СТІЙКОСТІ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ СИСТЕМ ВИЯВЛЕННЯ ВТОРГНЕНЬ В УМОВАХ ЗСУВУ КОНЦЕПЦІЇ

У роботі досліджується поведінка моделей машинного навчання у задачі виявлення мережевих вторгнень в умовах зсуву концепції. Розглянуто три підходи: логістична регресія, випадковий ліс та нейронна мережа у вигляді багатошарового перцептрона. Запропоновано експериментальний підхід до моделювання зсуву концепції, що дозволяє оцінити стійкість моделей до зсуву концепції. Експерименти проводяться на двох відкритих датасетах: NSL-KDD та UNSW-NB15, що дозволяє порівняти стійкість моделей до контрольованих змін характеристик даних. Запропоновано метрику індексу стабільності, яка кількісно оцінює деградацію якості моделей між еталонним станом та умовами зсуву.

Ключові слова: *штучний інтелект, машинне навчання, зсув концепції, логістична регресія, випадковий ліс, нейронна мережа, система виявлення вторгнень.*

Pekurovskyi H. V., Shcherbak T. L., Klymova A. S., Kalashnyk M. O.

STABILITY INDEX AND EXPERIMENTAL STUDY OF MACHINE LEARNING MODEL ROBUSTNESS FOR INTRUSION DETECTION SYSTEMS UNDER CONCEPT DRIFT

This paper investigates the behavior of machine learning models in the task of network intrusion detection under concept drift conditions. Three approaches are examined: logistic regression, random forest, and a neural network in the form of a multilayer perceptron. An experimental approach to modeling concept drift is proposed, allowing the robustness of models to concept drift to be assessed. The experiments are conducted on two open datasets, NSL-KDD and UNSW-NB15, enabling a comparison of model robustness to controlled changes in data characteristics. A stability index metric is proposed, which quantitatively evaluates the degradation of model quality between the reference state and drift conditions.

Keywords: *artificial intelligence, machine learning, concept drift, logistic regression, random forest, neural network, intrusion detection system.*

Стаття подана до редакції: 07*/10/2026*
Стаття прийнята до опублікування:15/10/2026
Стаття опублікована: 10/11/2026

Стаття поширюється на умовах ліцензії CC BY 4.0