УДК 004.056.5 **doi: 10.26906/SUNZ.2023.1.078**

С. Ю. Гавриленко, В. В. Челак

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

РОЗРОБКА МЕТОДУ ІДЕНТИФІКАЦІЇ СТАНУ КОМП'ЮТЕРНОЇ СИСТЕМИ НА ОСНОВІ НЕЧІТКИХ ДЕРЕВ РІШЕНЬ

Анотація. Предметом дослідження є методи та засоби ідентифікації стану комп'ютерної системи**. Метою статті** є підвищення якості класифікації даних за рахунок розробки методу ідентифікації стану комп’ютерної системи. **Завдання**: дослідити методи ідентифікації стану комп’ютерної системи та розробити метод класифікації стану комп'ютерної системи з метою захисту даних. Використовуваними **методами** є: методи штучного інтелекту, машинного навчання. Отримано такі **результати**: досліджено методи ідентифікації стану комп’ютерної системи KNN (k-nearest neighbors), метод опорних векторів (SVM), нейронні мережі, дерева рішень. Запропоновано метод ідентифікації стану комп’ютерної системи на основі нечітких дерев рішень, який відрізняється від відомих методів побудови наявністю спеціальної процедури фазифікації атрибутів вихідних даних та побудови функції приналежності. Розроблено програмне забезпечення, в якому реалізовано та досліджено запропонований метод вирішення задачі ідентифікації стану комп'ютерної системи. **Висновки.** Наукова новизна отриманих результатів полягає у дослідженні методів ідентифікації стану комп'ютерної системи, розробці методу на основі нечітких дерев рішень, оцінці якості моделі на етапі навчання та тестування, виконання порівняльного аналізу.

Ключові слова: комп'ютерна система, ідентифікація стану, обробка даних, машинне навчання, нечітка логіка, нечіткі дерева прийняття рішень.

Вступ

Сьогодні комп’ютерні технології є однією із ключових складових функціонування будь-якої держави і впливають на її безпеку та економічний успіх.

З початком повномасштабного вторгнення Росії в Україну в 2022 році кількість атак зросла в десятки разів. Кібератаки стали радше інструментом ведення бойових дій, аніж засобом тиску на окремих людей та їхні приватні дані.

Більшість кібератак була спрямована на державні сайти, транспортні компанії, енергетичний і фінансовий сектор [1].

Згідно статистики державної служби спеціального зв'язку та захисту інформації України, якщо до війни на державний сектор здійснили близько 143 тисяч атак, то в наступні місяці ця цифра стрімко зросла і досягла 25,1 мільйонів у вересні 2022 року. [1]

Як прогнозують експерти, кількість атак також буде збільшуватися. Це вимагає вдосконалення інфраструктури, перегляду стратегії інформаційної безпеки, архітектури комп’ютерної системи і методів впровадження та засобів ідентифікації їх стану, особливо в умовах постійної трансформації глобальної системи безпеки.

Крім того, сучасні умови політичного та соціально-економічного розвитку країни потребують балансу у розширенні вільного обміну інформацією та необхідністю дії окремих регламентованих обмежень на її поширення.

Комп’ютерна система (КС) характеризується великим обсягом показників (атрибутів) її функціонування. Це призводить до наявності труднощів з адекватного відбору атрибутів для ідентифікації стану КС в умовах зовнішніх впливів і розробки критерію оцінки, що відповідає обраним показникам [3].

Об’єктом дослідження є процес ідентифікації стану комп’ютерної системи.

Предметом дослідження є методи ідентифікації стану комп’ютерної системи.

Метою дослідження є розробка методу ідентифікації стану комп’ютерної системи на основі використання нечітких дерев рішень.

Огляд літератури. Функціонування КС характеризується великою кількістю процесів [3], в тому числі породжених впливом шкідливого програмного забезпечення. Для аналізу цих даних використовуються складні математичні алгоритми, що базуються на машинних методах навчання [4]. Найбільш популярні алгоритми машинного навчання наведено в [5]. Основою зазначених методів є технології і процедури, які вирішують завдання ідентифікації стану КС та виявлення аномалій в даних [6, 7]. Наразі активно використовуються метод k найближчих сусідів (k-nearest neighbors)[8, 9]; SVM – метод опорних векторів [10, 11]; нейронні мережі, ансамблеві методи [12] та ін.

Але більшість із моделей, наприклад, нейромережеві моделі поводяться як «чорна скринька», яка не пояснює, чому з таких вхідних даних вийшов саме цей результат. Крім того відсутня достатня для коректного навчання ML-моделей кількості навчальних та тестових даних у всіх напрямках кіберзагроз, від комп'ютерних вірусів до прийомів соціальної інженерії. Зловмисники також використовують алгоритми машинного навчання для створення шкідливих програм, аналізу користувача поведінки, розробки ботів-збирачів персональних даних, пошуку вразливостей, підбору паролів, підміни особистості, обходу систем захисту та ін. Можливі специфічні атаки на ML-алгоритми та використовувані набори даних, що може призвести до невірних рішень, пропущених атак або помилкових спрацьовувань;

Проведений аналіз показав, що основними недоліками даних методів є зневага факторами нечітких даних і низька адаптація до динамічних змін структур вихідних даних та зовнішніх дій, що у своїй послідовності призводить до зниження достовірності ідентифікації [13].

Одним із ефективних методів машинного навчання є використання дерев рішень (ДР, Decision

Trees), які успішно застосовують у багатьох областях, вирішуючи завдання класифікації, регресії та вибору підмножини ознак ідентифікації [2, 3, 14].

ДР є популярними моделями машинного навчання. Для побудови дерева на кожному внутрішньому вузлі необхідно знайти таку умову, яка б розбивала множину, що асоціюється з цим вузлом на підмножини. В якості перевірки повинен бути вибраний один з атрибутів. Загальне правило для вибору атрибуту можна сформулювати таким чином: вибраний атрибут повинен розбити множину так, щоб отримані у результаті підмножини складалися з об'єктів, що належать до одного класу, або були максимально наближені до цього, тобто кількість об'єктів з інших класів («домішок») в кожній з цих множин була якомога меншою. На вхід алгоритму можна подавати всі існуючі атрибути, алгоритм сам вибере найбільш значущі і вони будуть використані для побудови дерева [15].

ДР є наочними та дозволяють зрозуміти та пояснити, чому конкретний об'єкт відноситься до того чи іншого класу. Дають змогу відображати правила з бази даних природною мовою. Дозволяють створювати моделі у тих галузях, де аналітику досить складно формалізувати знаннями. Більше того, їх індукційний процес є зазвичай швидким , вимагає невеликих обчислювальних ресурсів.

Більшість алгоритмів побудови ДР мають можливість спеціальної обробки пропущених значень, працюють і з числовими та з категоріальними типами даних. ДР, на відміну багатьох методів, будують непараметричні моделі і здатні обробляти категоріальні значення.

Однак алгоритм ДР має ряд недоліків. Основний з них – це застосування експертного методу при формуванні ймовірностей настання того чи іншого сценарію. Це пов’язано не лише із суб’єктивізмом аналітика-експерта, але і з неможливістю, в ряді випадків, точно спрогнозувати багатьох ймовірностей настання сценарії, наприклад використання нового сценарію кібератаки. Особливо це актуально при ідентифікації стану КС, оскільки тут існує чимало факторів, які важко передбачити. Цей недолік можна зменшити, застосовуючи нейронні мережі, теорію нечітких множин та відповідне програмне забезпечення [16].

Крім того, якщо значення атрибутів знаходяться на межі розмежування класів, то більш ефективним є також використання апарату нечіткої математики та ДР, який пристосований для ідентифікації стану КС в цих умовах [17].

Підходи та методи

Нечітка логіка – це математичний підхід, що базується на понятті нечіткої множини і вивчає певні процеси та явища з функцією приналежності елемента до множини. В основі цього підходу лежить теорія, викладена в серії робіт Л. Заде в 1965-1973 роках. Саме в його роботах розглядаються елементи множини, для яких функція приналежності представляє собою не жорсткий поріг (належить / не належить), а плавну сигмоїду (часто спрощено ламану лінію), що проходить через всі значення від нуля до одиниці [18]. Таким чином, даний метод дозволяє більш точно описувати комплексні явища, події, величини завдяки можливості описати їх не лише одним усередненим числом, а логічним відображенням, яке є найбільш близьким до того, як його розуміють з сприйняття зовнішнього світу. Тобто теорія нечітких множин дозволяє отримати більш точне відображення реального світу в математичних моделях.

Нечіткі підхід є одним із розширень класичної моделі ДР [19]. На відміну від простих дерев рішень, де кожен приклад належить конкретному вузлу, нечітке дерево рішень використовує ступінь приналежності. Нечіткий підхід краще адаптований до умов реального світу. Різні бази нечітких правил можна з легкістю об'єднати. Крім того, поняття нечіткої множини цілком узгоджується з інтуїтивними уявленнями про навколишній світ, так як велика частина використовуваних понять за своєю природою є нечіткою та розмитою.

Комбінація нечітких систем та дерев рішень дає нечітке дерево рішень (НДР, Fuzzy Decision Trees). При використанні нечіткого підходу у кінцевій оцінці поряд із числовим результатом буде одержано і розподіл його істинності у вигляді відповідної функції приналежності. Причому характер функції (ступінь розмитості) приналежності характеризуватиме і ступінь нечіткості рішення та може інтерпретуватися як функція розподілу оцінок істинності прийнятого рішення. Таким чином, при використанні НДР не втрачаються знання про те, що об'єкт може мати властивості як однієї ознаки, так і іншої в тій чи іншій мірі.

Перевагами НДР є висока точність класифікації, яка отримана за рахунок використанням нечіткої логіки, швидкий процес навчання та простота результату для інтерпретації. Нечіткий підхід, інтегруючи поняття нечітких множин та ентропії стає все більш популярним при аналізі даних, особливо які знаходяться на межі розмежування, при вирішенні проблем невизначеності, шуму і неточності даних. Здатність алгоритму видавати для нового об'єкту ступінь приналежності до класу дозволяє управляти порогом прийняття рішень для класифікації.

Відмінною рисою НДР є те, що для кожного атрибута необхідно виділити кілька його лінгвістичних значень і визначити ступінь приналежності прикладів (об’єктів) до них. Замість кількості прикладів конкретного вузла НДР групує їх ступінь приналежності.

Процес побудови НДР починається з фазифікації атрибутів x_i вихідних даних, побудови функцій приналежності та визначення ступеня приналежності об’єктів (прикладів) до кожного із них. Надалі процес повторюється ітераційно та починається зі знаходження атрибуту умови, що визначає поточний вузол дерева (при цьому спочатку визначається кореневий вузол дерева):

1. Визначити значення загальної ентропії, яка дає середню кількість інформації для визначення класу об'єкта з множини:

$$E(S^N) = -\sum_i \frac{P_i^N}{P^N} \cdot \log_2 \frac{P_i^N}{P^N}, \qquad (1)$$

де P^N – коефіцієнт що містить загальні характеристики прикладів вузла N.

2. Визначити коефіцієнти P_i та ентропію для розбиття за кожним атрибутом A

$$P_i^N = \sum_{S^N} \min(\mu_N(D_j), \mu_i(D_j)); \quad (2)$$

$$E(S^N, A) = \sum_j \frac{P^{N|j}}{P^N} \cdot E(S^{N|j}), \quad (3)$$

де вузол $N \mid j$ – дочірній для вузла N.

3. Обчислити інформаційний приріст для кожного із атрибутів

$$G(S^N, A) = E(S^N) - E(S^N, A). \quad (4)$$

4. Визначити атрибут A^x з максимальним приростом інформації G:

$$A^{\times} = \arg\max_A G(S, A). \quad (5)$$

5. Використати атрибут A^x у якості вузла дерева N та видалити його із списку атрибутів, які ще не було оброблено.

6. Вузол N розбити на кілька підсистем N_j. Перерахувати ступінь приналежності μ атрибутів до наступного ярусу:

$$\mu_{N|j}(e_k) = min(\mu_{N|j}(D_k), \mu_{N|j}(D_k, a_j)). \quad (6)$$

7. Якщо всі приклади підвузла $N|j$ мають ступінь приналежності, що дорівнює нулю, то такий підвузол видалити так як не існує жодного прикладу, де присутня така комбінація лінгвістичних значень.

8. Якщо не досягнуто максимальної глибини дерева або усі атрибути не використано, то повернутися до п. 1 інакше закінчити побудову дерева.

Таким чином, в режимі навчання формується база правил та будується НДР. У кожному вузлу дерева під час навчання формуються функції приналежності атрибутів до наступного ярусу ДР. При цьому, на відміну від звичайних дерев рішень, кількість листків (дуг) НДР може бути більшою ніж дві дуги.

Належність до цільового класу для нового об'єкту визначається формулою:

$$\sigma_j = \frac{\sum_l \sum_k P_k^l \cdot \mu_l(D_j) \cdot \chi_k}{\sum_l (\mu_l(D_j) \cdot \sum_k P_k^l)}, \quad (7)$$

де P_i^k – коефіцієнт співвідношення прикладів листа дерева для значення цільового класу k, $\mu_l(D_j)$ – ступінь приналежності прикладу до вузла l, χ_k – приналежність значення цільового класу k до позитивного значення результату класифікації.

Експериментальні дослідження та оцінка ефективності

Однією із проблем класифікаторів на основі нечітких дерев рішень є автоматизація процесу фазифікації атрибутів вихідних даних та налаштування НДР.

В даній роботі запропоновано метод побудови НДР, який відрізняється від відомих методів наявністю автоматизованої спеціальної процедури фазифікації атрибутів вихідних даних та побудови функції приналежності:

1. Нехай об'єкт спостереження може знаходитися в одному із двох станів H_1 та H_2. Стан об'єкту визначається значеннями p атрибутів (показників) x_1, x_2, … x_n. Передбачається, що числові значення атрибутів, які контролюються, є нормально розподіленими випадковими величинами, причому їх математичні сподівання визначаються вектором

$$M = (m_1, m_2, \dots m_p).$$

2. Для кожного атрибуту x_i розраховуються характеристики: математичне очікування m, дисперсія D, середнє квадратичне відхилення σ.

3. Якщо отримане значення дисперсії $D \le 5$, то фазифікація атрибуту відбувається за рахунок побудови трьох функцій приналежності μ_1, μ_2, μ_3 (рис. 1):

$$\mu_1(x_i, m_i, \sigma_i) = \begin{cases} 1, & x_i < m_i - 3\sigma_i, \\ \dfrac{m_i - \sigma_i - x_i}{2\sigma_i}, & m_i - 3\sigma_i \le x_i < m_i - \sigma_i, \\ 0, & m_i - \sigma_i \le x_i; \end{cases} \quad (8)$$

$$\mu_2(x_i, m_i, \sigma_i) = \begin{cases} 0, & x_i < m_i - 3\sigma_i, \\ \dfrac{x_i - m_i + 3\sigma_i}{2\sigma_i}, & m_i - 3\sigma_i \le x_i < m_i - \sigma_i, \\ 1, & m_i - \sigma_i \le x_i < m_i + \sigma_i \\ \dfrac{m_i + 3\sigma_i - x_i}{2\sigma_i}, & m_i + \sigma_i \le x_i < m_i + 3\sigma_i, \\ 0, & m_i + 3\sigma_i \le x_i; \end{cases} \quad (9)$$

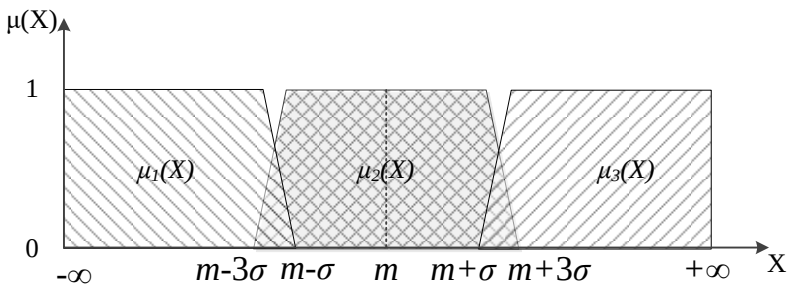

Рис. 1. Графічне подання функцій приналежності атрибуту (D ≤ 5)

$$\mu_3(x_i, m_i, \sigma_i) = \begin{cases} 0, & x_i < m_i + \sigma_i, \\ \dfrac{x_i - m_i - \sigma_i}{2\sigma_i}, & m_i + \sigma_i \le x_i < m_i + 3\sigma_i, \\ 1, & m_i + 3\sigma_i \le x_i. \end{cases} \quad (10)$$

4. Якщо отримане значення дисперсії D>5, то фазифікація вихідних даних відбувається за рахунок кластеризації атрибутів, наприклад алгоритмом DBSCAN. При цьому кількість нечітких множин визначається кількістю кластерів.

5. Для кожного кластеру будується трапецієвидна функція приналежності. При цьому, на відрізку [b;c] де функція приналежності дорівнює одиниці знаходиться 80% елементів. Відповідно, на відрізках [a;b] та [c;d] (часткова належність) знаходиться 20% елементів кластеру (рис.2).

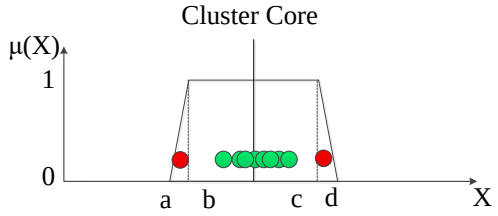

Рис. 2. Приклад побудови функції приналежності при кластеризації атрибутів

Проведені дослідження надали можливість оцінити показники якості класифікації на основі дерев рішень з багатовимірними вузлами (MD DT), запропонованого авторами в [3] та методу класифікації даних на основі нечітких дерев рішень зі спеціальною процедурою формування нечітких множин і функцій приналежності (FDT) [20]. Результати оцінки наведено в табл. 1 (Accuracy (ACC), Precision (PPV), Recall (TPR), Probability of false alarm (fall-out, FPR), Miss rate (FNR)).

Таблиця 1 – **Показники якості класифікації на основі методів MD DT та FDT, зі спеціальною процедурою формування нечітких множин і функцій приналежності**

Показники якості	Навчальний набір даних		Тестовий набір даних	
	MD DT	**FDT**	MD DT	FDT
ACC	1	**0,9997**	0,91	0,932
PPV	1	**0,9997**	0,86	0,889
TPR	1	**1**	0,972	0,987
FPR	0	**0,0007**	0,154	0,123
FNR	0	**0**	0,028	0,013

Як видно із таблиці, використання процедури формування нечітких множин на етапі навчання моделі надає можливість підвищити точність ідентифікації (при цьому, помилка на навчальній вибірці складає 0,03% на тестовій - 6,8%).

Крім того виконано оцінку методів класифікації на основі MD DT та FDT (рис. 3, 4) з використанням ROC-кривої.

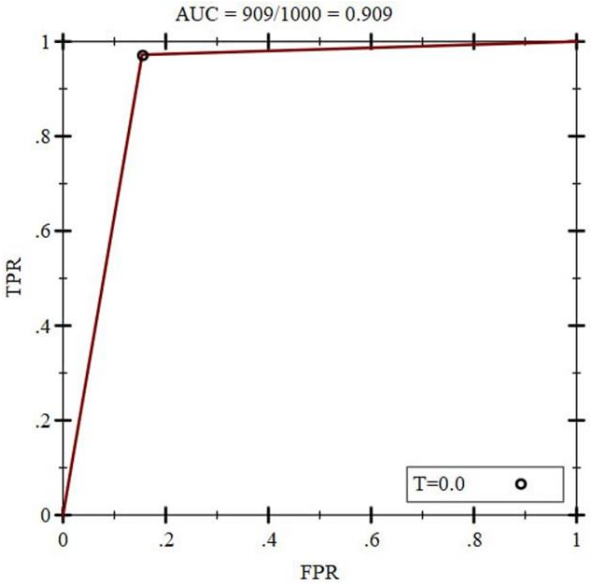

Рис. 3. Roc крива якості класифікації на основі MD DT

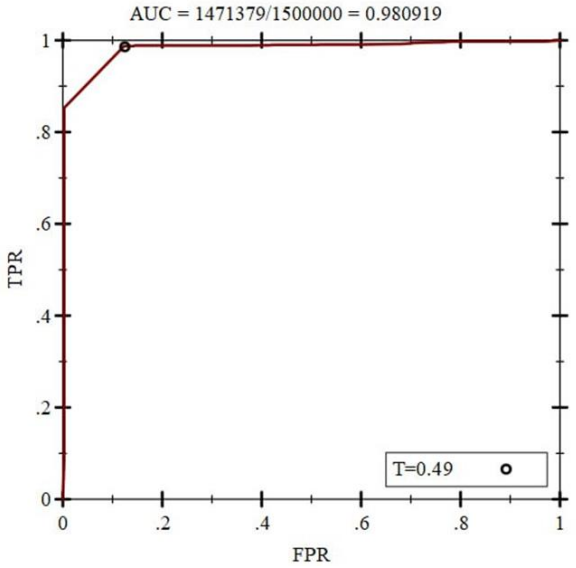

Рис. 4. Roc крива якості класифікації на основі FDT

Як видно із рис.3, методи класифікації на основі MD DT та FDT є якісними. Метод на основі нечітких дерев рішень зі спеціальною процедурою формування нечітких множин і функцій приналежності показав кращий результат.

Також досліджено на класифікатори на основі модифікацій алгоритмів SVM [19] та KNN [20] та виконано порівняльний аналіз. Результати дослідження наведено на рис 5, 6. Рис. 5 містить помилки класифікації на етапі навчання (Bias), рис. 6 – на етапі тестування моделі (Variance).

Як видно із рис. 5, 6 використання методу класифікації даних на основі нечітких дерев рішень зі спеціальною процедурою формування нечітких множин і функцій приналежності дозволило підвищити точність класифікації на етапі тестування на 2,3% за рахунок формування нечітких множин та зменшення кількість ярусів дерева рішень.

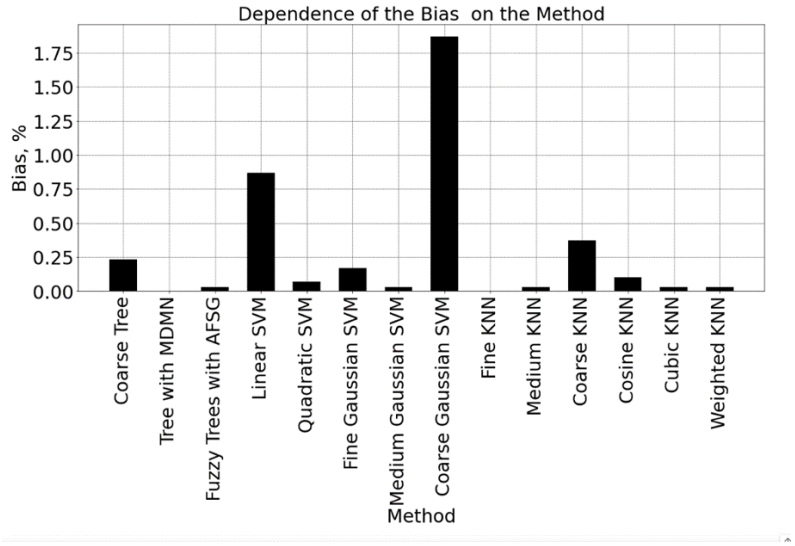

Рис. 5. Помилки класифікації на етапі навчання моделі (Bias)

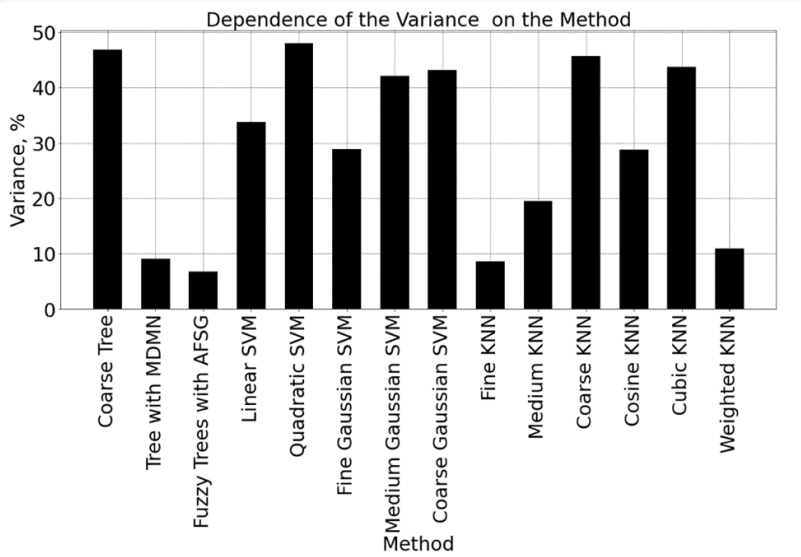

Рис. 6. Помилки класифікації на етапі тестування моделі (Variance)

Висновки

Таким чином, у роботі вирішено завдання підвищення оперативності та точності ідентифікації стану функціонування КС.

Наукова новизна отриманих результатів полягає в тому, що вперше запропоновано метод ідентифікації стану КС на основі НДР, який відрізняється від відомих методів побудови НДР наявністю спеціальної процедури фазифікації атрибутів вихідних даних та побудови функції приналежності.

При цьому фазифікація атрибутів є автоматизованою та відбувається за рахунок статистичного аналізу атрибутів аналізу або їх кластеризації, наприклад, алгоритмом DBSCAN.

Результати моделювання показали, що запропонований метод надає можливість зменшити кількість ярусів дерева рішень, що дозволяє підвищити оперативність ідентифікації стану комп'ютерної системи. Використання процедури формування нечітких множин на етапі навчання моделі надає можливість підвищити точність ідентифікації (при цьому, помилка на навчальній вибірці складає 0,03% на тестовій - 6,8%).

Недоліком даного методу є необхідність використання більшої кількості обчислювальних ресурсів при побудові НДР.

СПИСОК ЛІТЕРАТУРИ

1. О. Вдовенко. 25 мільйонів кібератак щомісяця. Як Росія намагається зашкодити Україні в цифровому просторі. URL: https://detector.media/infospace/article/204308/2022-10-29-25-milyoniv-kiberatak-shchomisyatsya-yak-rosiya-namagaietsya-zashkodyty-ukraini-v-tsyfrovomu-prostori/.
2. S. Y. Gavrylenko & I. V. Sheverdin, "Development Of Method To Identify The Computer System State Based On The «Isolation Forest» Algorithm". Radio Electronics, Computer Science, Control, 2021(1), pp. 105–116, doi:10.15588/1607-3274-2021-1-11

3. S.Y. Gavrylenko, V.V. Chelak and S.G. Semenov Development of Method for Identification the Computer System State based on the Decision Tree with Multi-Dimensional Nodes, Radio Electronics, Computer Science, Control (RECSC).– No. 2 (2022).– 2022.– pp.113-121, doi: 10.15588/1607-3274-2022-2-11.
4. Leon Reznik, "Computer Security with Artificial Intelligence, Machine Learning, and Data Science Combination," in *Intelligent Security Systems: How Artificial Intelligence, Machine Learning and Data Science Work For and Against Computer Security* , IEEE, 2022, pp.1-56, doi: 10.1002/9781119771579.ch1.
5. Amarudin, R. Ferdiana and Widyawan, "A Systematic Literature Review of Intrusion Detection System for Network Security: Research Trends, Datasets and Methods," *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, 2020, pp. 1-6, doi: 10.1109/ICICoS51170.2020.9299068.
6. J. J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: A review", *Computers & security*, 2011, vol. 30, no. 6-7, pp. 353-375, doi:10.1016/j.cose.2011.05.008.
7. J. Tai, I. Alsmadi, Y. Zhang and F. Qiao, "Machine Learning Methods for Anomaly Detection in Industrial Control Systems," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 2333-2339, doi: 10.1109/BigData50022.2020.9378018.
8. S. Sun and R. Huang, "An adaptive k-nearest neighbor algorithm," 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 91-94, doi: 10.1109/FSKD.2010.5569740.
9. S. Malhotra, V. Bali and K. K. Paliwal, "Genetic programming and K-nearest neighbour classifier based intrusion detection model," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, 2017, pp. 42-46, doi: 10.1109/CONFLUENCE.2017.7943121
10. P. Gattineni and G. R. S. Dharan, "Intrusion Detection Mechanisms: SVM, random forest, and extreme learning machine (ELM)," *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 273-276, doi: 10.1109/ICIRCA51532.2021.9544551.
11. L. Lin, R. Zuo, S. Yang and Z. Zhang, "SVM ensemble for anomaly detection based on rotation forest," *2012 Third International Conference on Intelligent Control and Information Processing*, 2012, pp. 150-153, doi: 10.1109/ICICIP.2012.6391455.
12. M. I. Sayed, I. M. Sayem, S. Saha and A. Haque, "A Multi-Classifier for DDoS Attacks Using Stacking Ensemble Deep Neural Network," *2022 International Wireless Communications and Mobile Computing (IWCMC)*, 2022, pp. 1125-1130, doi: 10.1109/IWCMC55113.2022.9824189.
13. S. Gavrylenko, V. Chelak and O. Hornostal, "Research of Intelligent Data Analysis Methods for Identification of Computer System State," 2020 XXX International Scientific Symposium 'Metrology and Metrology Assurance (MMA), Sozopol, Bulgaria, 2020, pp. 1-5, doi: 10.1109/MMA49863.2020.9254252.
14. N. Kumar, H. Akash, R. A. Prataap, G. Srinath and C. Mala, "Intelligent Intrusion Detection System using Decision Tree Classifier and Bootstrap Aggregation," *2018 8th International Symposium on Embedded Computing and System Design (ISED)*, 2018, pp. 199-203, doi: 10.1109/ISED.2018.8704056.
15. J. Ye et al., "A Chi-MIC Based Adaptive Multi-Branch Decision Tree," in IEEE Access, vol. 9, pp. 78962-78972, 2021, doi: 10.1109/ACCESS.2021.3077125.
16. F. Q. Li, S. -L. Wang, A. W. -C. Liew, W. Ding and G. -S. Liu, "Large-Scale Malicious Software Classification With Fuzzified Features and Boosted Fuzzy Random Forest," in IEEE Transactions on Fuzzy Systems, vol. 29, no. 11, pp. 3205-3218, Nov. 2021, doi: 10.1109/TFUZZ.2020.3016023.
17. S. Semenov, O. Sira , S. Gavrylenko, N. Kuchuk. Identification of the state of an object under conditions of fuzzy input data, *Eastern-European Journal of Enterprise Technologies***,** 2019, Vol 1, No 4 (97), pp.22-29, doi: 10.15587/1729-4061.2019.157085
18. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений, М.: Мир, 1976, 166 с.
19. Sumalatha and P. U. Sankar, "Fuzzy random decision tree (FRDT) framework for privacy preserving data mining," *2016 SAI Computing Conference (SAI)*, 2016, pp. 195-202, doi: 10.1109/SAI.2016.7555982.
20. and O. Hornostal, "Construction Method Of Fuzzy Decision Trees For Identification The Computer System State," *2022 XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA)*, Sozopol, Bulgaria, 2022, pp. 1-5, doi:10.1109/MMA55579.2022.9992878

Received (Надійшла) 12.01.2023
Accepted for publication (Прийнята до друку) 01.03.2023

Development of method base on fuzzy decision trees for identification of the computer systems state

S. Gavrylenko, V. Chelak

Abstract. The **subject** of research is the methods and means of identifying the state of the computer system. The **purpose** of the article is to improve the quality of data classification by developing a method for identifying the state of the computer system. **Task**: to investigate methods of identifying the state of a computer system and to develop a method of classifying the state of a computer system for the purpose of data protection. **Methods used**: artificial intelligence methods, machine learning, decision tree methods .The following results were obtained: methods of identifying the state of the computer system KNN (k Nearest Neighbors), support vector method (SVM), neural networks, decision trees were investigated. **The results were obtained**: a method of identifying the state of the computer system based on of fuzzy decision trees, which differs from the known methods of fuzzy decision trees by the presence of a special procedure for falsification the attributes of source data and constructing membership function was proposed. The software was developed, in which the proposed method of solving the problem of identifying the state of the computer system was implemented and investigated. **Conclusions**. The scientific novelty of the obtained results lies in the study of methods for identifying the state of the computer system, the development of a method based on fuzzy decision trees, the assessment of the quality of the model at the stage of training and testing, and the performance of a comparative analysis.

Keywords: computer system, state identification, data processing, machine learning, fuzzy logic, fuzzy decision trees.