DOI: 10.18372/2310-5461.67.20347
УДК 004.056.5

А. А. Кулько
Київський національний університет імені Тараса Шевченка
orcid.org/0009-0006-1185-0774
e-mail: kulko452@gmail.com

С. В. Толюпа**,** д-р техн. наук, професор
Київський національний університет імені Тараса Шевченка
orcid.org/0000-0002-1919-9174
e-mail: serhii.toliupa@knu.ua

МЕТОДИКА ВИКОРИСТАННЯ АНСАМБЛЕВОГО КЛАСИФІКАТОРА ДЛЯ ОПТИМІЗАЦІЇ СИСТЕМИ ВИЯВЛЕННЯ ВТОРГНЕНЬ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Вступ

Ансамблеве навчання є одним з методів, де різні класифікатори об'єднуються для досягнення кращих результатів. У системі виявлення вторгнень (IDS) основне використання ансамблевого класифікатора полягає в об'єднанні кількох класифікаторів для досягнення більш точного висновку, ніж окремий класифікатор. Зазначено, що точність класифікаторів зростає при їх об'єднанні. Зі збільшенням інтернет-послуг та їх використання з відкритим доступом до конфіденційних даних, необхідність безпеки систем стала нагальною потребою. Системи виявлення вторгнень (IDS) забезпечують важливий рівень безпеки для комп'ютерних систем і мереж, стають все більш і більш важливим питанням і використовуються для захисту комп'ютерної системи від ризику крадіжки зловмисниками [1]. В останні роки обсяг інтернет-безпеки та витонченість цільових мережевих атак значно зросли. Збільшилася кількість загроз і вразливостей, таких як бізнес-системи, військові тощо. Це робить системи виявлення вторгнень значущою галуззю досліджень. Системи виявлення вторгнень поділяються на дві частини – на основі аномалій та на основі зловживань. Модель, заснована на особливостях, використовується для відхилення нової інформації від заздалегідь визначеного профілю інформації. Модель, заснована на зловживаннях, також відома як модель, заснована на навчанні, використовується для виявлення шляхом зіставлення нової інформації з відомими атаками в базі даних. Основною вимогою будь-якої IDS є точність. Іншими вимогами є розширюваність та адаптивність. Основною проблемою IDS є виявлення помилкових атак. Враховуються такі метрики, як рівень хибнопозитивних спрацьовувань, рівень хибнонегативних спрацьовувань, точність, повнота та точність виявлення, і існує компроміс між цими метриками. Отже, для запропонованої системи використовується ансамбль класифікаторів, що забезпечує хорошу точність. Запропонована ідея може бути використана для системи виявлення вторгнень певного сегменту об'єкта критичної інфраструктури для запобігання витоку конфіденційних даних [2]. Це допоможе організації, виявляючи майбутні атаки, які шкодять конфіденційним даним. Таким чином, буде проведено ансамблеве дослідження різних груп баєсівського класифікатора, наївного Баєса, IBk, JRip, MLP, J48 та PART.

Аналіз останніх досліджень і публікацій

Для вирішення ефективної протидії кібератакам на критичну інфраструктуру доцільно застосовувати підходи, які одночасно мають властивості верифікації та адаптації, оскільки перша властивість дозволить детально досліджувати інциденти безпеки з метою надання можливості здійснення необхідної та адекватної реакції менеджера з кібербезпеки у вигляді спеціальних превентивних заходів щодо припинення інформаційно-руйнівного втручання в роботу інформаційних систем (ІС) КІ (критичної інфраструктури), а друга властивість – зробити крок до виявлення нових, некласифікованих кібератак [3].

На сьогоднішній день у даній предметній області ведуться активні розробки, про що свідчать роботи провідних вітчизняних і зарубіжних дослідників: Корченко О. Г., Горбенко І. Д., Кузнецова О. О., Субача І. Ю., Шелухіна О. І., Бучика С. С., Бурячка В. Л., Грищука Р. В., Євсеєва С. П., Shanmugavadivu R., Kumar S., Yanpeng Guan, Roselin Mary S. та ін.

На сьогоднішній день існує велика кількість різноманітних методів та засобів захисту від кібератак на об'єктах КІ [4].

Нажаль на теперішній час не існує абсолютно універсального методу протидії кібератакам і тому виникає необхідність комплексного підходу до вирішення даної задачі. Застосування методів інтелектуального аналізу даних та штучного інтелекту, оптимізації використання класифікаторів дає можливість підвищити ефективність протидії вторгненням і захистити об'єкти КІ від потенційних порушників [5]. Узагальнена система безпеки критичної інфраструктури представлена на рис. 1.

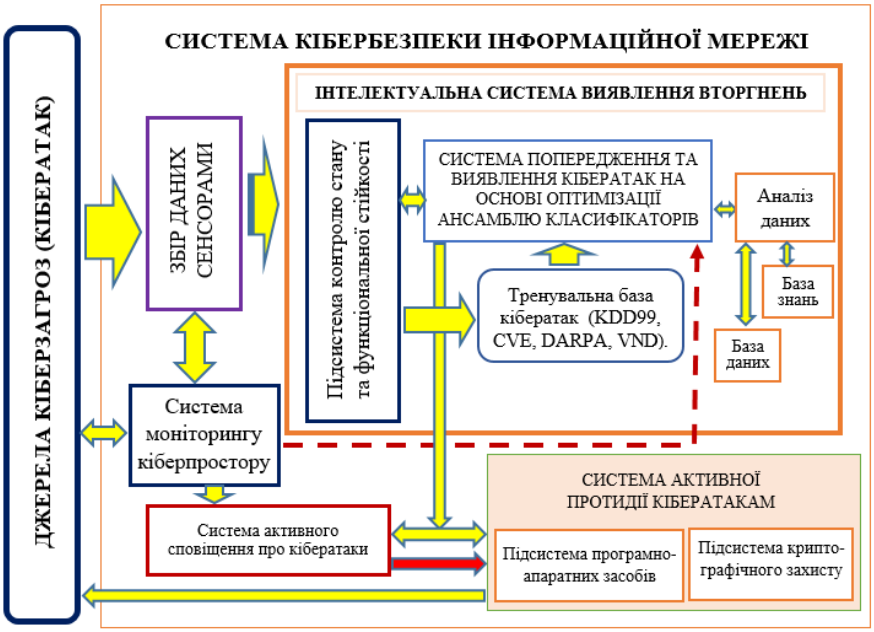

Рис. 1. Узагальнена система протидії кібервпливу на критичну інфраструктуру

В наукових дослідженнях велика увага приділяється комбінаториці класифікаторів для оптимізації структури систем виявлення вторгнень. Так у [6], після проведення експериментів результати показали, що ансамбль J48, REPtree та баєсівської мережі виявив DoS-атаки зі значною точністю. У дослідженні [7], кількість класифікаторів використовувалася для перевірки точності. Результати довели, що Bayes Net та Random Forest Tree є найкращими методами класифікації, а бустинг дає хороший результат. У [8] показано, що продуктивність усіх методів класифікації оцінювалася на основі перехресної валідації та тестових даних, з яких PART дав кращі результати, ніж інші. Автори роботи [9] визнали використання алгоритмів на основі кореляції для зменшення ознак у наборі даних. У дослідженні [10], метод ранжування Гаррета був застосований для ранжування різних класифікаторів відповідно до їхньої продуктивності. Підхід класифікації Rotation Forest показав решту. У роботі [11], представлено підхід генетичного алгоритму для ефективного виявлення різних типів атак мережевих вторгнень з використанням набору даних KDD99, а також був проведений відбір 10 найкращих класифікаторів – SGD, IBK, JRip, PART, Random Forest, Random Tree, Logistics, Bayes Net. З них Random Forest, IBK дали кращі результати щодо точності та меншого часу обробки. В [12], методи м'яких обчислень застосовані до системи виявлення вторгнень з використанням стандартних наборів даних (KDD99) для визначення швидкості виявлення та частоти хибнопозитивних спрацьовувань.

Постановка завдання

У 2024 році в Україні було зафіксовано приблизно 1,5 мільйона кіберінцидентів, що підкреслює гостроту проблеми кібербезпеки. Особливе занепокоєння викликає використання кіберпростору для військово-політичних конфліктів, тероризму та хакерських атак. Експерти прогнозують зміну характеру кібероперацій: вони стануть більш прихованими та складними для виявлення традиційними системами захисту та фаєрволами.

Поточні дослідження свідчать, що існуючі методи та моделі виявлення кібератак потребують удосконалення. Це стосується як розробки адекватних математичних моделей, так і створення ефективних алгоритмів виявлення та прийняття рішень. Таким чином, актуальним науковим завданням є створення комплексних методів виявлення кібератак, що базуються на інтелектуальних системах для захисту інформаційних систем державних органів та оптимізації використання ансамблевих класифікаторів. Вирішення цієї проб-

леми дозволить усунути протиріччя між зростаючою складністю кіберзагроз і можливостями існуючих систем захисту [13].

Метою цього дослідження є застосування двох основних методів попередньої обробки – нормалізації та дискретизації – до системи виявлення вторгнень з використанням стандартного набору даних. Стандартний набір даних, що використовується, це набір даних KDD99 для системи виявлення зловживань. І постійно оновлювана база даних використовується для зберігання сигнатур різних атак, таких як Normal, Probe, U2R, DoS тощо. У цьому дослідженні п'ять класифікаторів об'єднано для зниження частоти помилок, зменшення перенавчання даних та виявлення більшої кількості атак, ніж одним класифікатором.

Основна частина

Останнім часом метод групування широко використовується для виявлення переривань у фреймворку. У 2019 році було проведено дослідження і визначення 10 основних класифікаторів: SGD, IBK, JRip, PART, J48, Random Tree, Logistics, Bayes Net. Спочатку на цих алгоритмах групування було проведено багато тестів, а потім вони були обрані. З них J48, IBK показали кращі результати щодо точності та меншого часу обробки. У 2021 році було запропоновано фреймворк з кількома системами класифікації та алгоритмами машинного навчання. Використовувалися методи класифікації Bayes Net, IBK, JRip, PART, Logistic, J48, Random Tree, J48 та REP tree. Усі вони були оснащені машинними обчисленнями: Boosting, Bagging та Stacking. Отже, результати показали, що Bayes Net та J48 Tree є найкращими методами класифікації, а бустинг дає чудовий результат [2]. У 2021 році було проаналізовано методи класифікації на основі правил – Decision Tree, PART, Zero R, One R та JRip. Виконання всіх цих стратегій класифікації оцінювалось на основі перехресної валідації та тестової інформації, з яких PART дав кращі результати, ніж інші. У 2022 році було запропоновано фреймворк для розпізнавання атак типу "відмова в обслуговуванні" в системі. Класифікатори, використані для аналізу, були Naïve Bayesian, Bayesian Network, Sequential Minimal Optimization, J48, дерево рішень. Після проведення тестів результати показали, що об'єднання J48, REP tree та Bayesian Network виявило DoS-атаки зі значною точністю. У 2020 році була запропонована модель виявлення вторгнень для ідентифікації атак U2R та R2L. Автори застосували алгоритм Adaboost для ансамблю слабких навчальних елементів та для підвищення продуктивності. Результати досліджень показали, що групування Naïve Bayes та MLP дало хороші результати для атак U2R та R2L з найвищою чутливістю.

Методи попередньої обробки даних

Нормалізація. Нормалізація є важливою частиною попередньої обробки даних, яка допомагає підготувати дані до аналізу та моделювання. Основна ідея полягає в тому, щоб привести всі числові дані до одного діапазону, що запобігає ситуації, коли ознаки з більшими значеннями домінують над ознаками з меншими.

Нормалізація – це процес масштабування числових даних, щоб вони мали однаковий діапазон. Це критично важливо для багатьох алгоритмів машинного навчання, особливо для тих, які залежать від відстаней між точками даних (наприклад, K-найближчих сусідів (KNN), метод опорних векторів (SVM), нейронні мережі) [1].

Основні переваги нормалізації: усунення домінування ознак, прискорення конвергенції, краща інтерпретація результатів.

Нормалізація має багато методів, таких як нормалізація Min-Max, нормалізація за Z-оцінкою, нормалізація цілочисельного масштабування та нормалізація десяткового масштабування. Нормалізація Min-Max, також відома як масштабування Min-Max, є одним з найпопулярніших і найпростіших методів. Вона лінійно трансформує дані так, щоб вони потрапляли в певний діапазон, зазвичай від 0 до 1. Формула для цього виглядає так:

$$x_{normalized} = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

де: x – вихідне значення; $x_{\min}$ – мінімальне значення ознаки у наборі даних; $x_{\max}$ – максимальне значення ознаки у наборі даних.

У нормалізації Min-Max зберігається взаємозв'язок між унікальною інформацією, і вона забезпечує лінійне перетворення. Основна мета використання цього методу полягає в тому, що він точно інтерпретує результат, усуває відхилення та аномалії, розкриває шаблони тощо.

Дискретизація. Дискретизація – це не просто допоміжна техніка, а фундаментальний метод обробки даних, що має значний вплив на ефективність і результативність різних аналітичних процесів, особливо в галузі машинного навчання. Її використання виправдано кількома ключовими факторами. Наведене обґрунтування використання дискретизації в машинному навчанні та аналізі даних можна розширити, розглянувши її застосування у сфері кібербезпеки та, зокрема, в системах виявлення вторгнень (IDS). У цих галузях дискретизація відіграє критично важливу роль

для підвищення ефективності та точності виявлення загроз.

Кібербезпека, особливо системи виявлення вторгнень (IDS), оперує з величезними обсягами даних, які часто включають як дискретні (наприклад, IP-адреси, номери портів, коди протоколів), так і безперервні змінні (наприклад, тривалість сеансу, обсяг переданих даних, частота пакетів). Саме для останніх дискретизація стає незамінним інструментом.

Розрізнення між керованою та некерованою дискретизацією має ключове значення в системах виявлення вторгнень.

Некерована дискретизація: методи, такі як рівноширокі або рівночастотні інтервали, можуть бути корисними для первинного аналізу, але вони не враховують інформацію про те, які ознаки пов'язані з атаками.

Керована дискретизація: цей підхід є більш доцільним, оскільки він використовує мітки даних (наприклад, «нормальний» трафік або «атака»). Наприклад, алгоритм на основі ентропії може розділяти безперервні значення таким чином, щоб кожен інтервал мав найвищу можливу інформаційну цінність для класифікації між «нормальним» і «аномальним» трафіком. Це робить дискретизовані дані значно потужнішими для навчання моделі IDS.

Таким чином дискретизація є потужним і необхідним інструментом в кібербезпеці, який забезпечує ефективну та швидку роботу систем виявлення вторгнень. Вона дозволяє оптимізувати обробку безперервних даних, зменшити обчислювальну складність, підвищити точність виявлення атак та спростити моделі машинного навчання, роблячи їх більш придатними для роботи в режимі реального часу.

Аналіз класифікаторів

Класифікатор Наївний Баєс є одним з найпростіших та найефективніших алгоритмів класифікації, який базується на теоремі Баєса. Незважаючи на свою простоту, він демонструє високу ефективність у багатьох задачах, зокрема у системах виявлення вторгнень (IDS), які захищають критичну інфраструктуру.

Наївний Баєс працює на основі припущення «наївної незалежності» між ознаками. Це означає, що він вважає, що наявність однієї ознаки у класі не залежить від наявності іншої. Хоча це припущення рідко відповідає дійсності, класифікатор все одно демонструє хороші результати на практиці.

Формула теореми Баєса:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)},$$

де $P(A \mid B)$ – ймовірність події А за умови, що подія В сталася. У нашому випадку, це ймовірність того, що трафік є атакою (A), за умови, що він має певні ознаки (B); $P(B \mid A)$ – ймовірність події B за умови, що подія A сталася. Це ймовірність того, що трафік має певні ознаки (B), якщо це атака (A); $P(A)$ – апріорна ймовірність події A. Ймовірність того, що трафік є атакою, без урахування ознак; $P(B)$ – апріорна ймовірність події B. Ймовірність того, що трафік має певні ознаки.

Класифікатор J48: реалізація алгоритму C4.5. Класифікатор J48 - це реалізація з відкритим вихідним кодом популярного алгоритму дерева рішень C4.5, розробленого для платформи машинного навчання Weka. C4.5 (і, відповідно, J48) будує дерево рішень шляхом поділу даних на менші групи на основі найкращої ознаки. Найкраща ознака обирається за допомогою двох показників: приріст інформації (Information Gain) (вимірює, наскільки ефективно ознака допомагає розділити дані на чистіші класи. Чим вищий приріст, тим кращою є ознака; відношення приросту (Gain Ratio) (використовується для уникнення переваги ознак з великою кількістю унікальних значень, що може призвести до перенавчання моделі).

Класифікатор J48 – це реалізація алгоритму C4.5 для побудови дерева рішень, який є одним із найпоширеніших і найефективніших алгоритмів машинного навчання. Він розроблений Россом Квінланом і використовується для задач класифікації.

Математичний апарат алгоритму C4.5 ґрунтується на концепції ентропії та приросту інформації (information gain). Головна ідея – рекурсивно розділяти набір даних на менші підмножини на основі атрибутів, які надають найбільший приріст інформації, поки кожна підмножина не буде «чистою» (тобто міститиме лише об'єкти одного класу) або поки не буде досягнуто певного критерію зупинки.

Ентропія в теорії інформації вимірює рівень невизначеності або хаосу в наборі даних. Чим вища ентропія, тим більше «змішані» класи в підмножині.

Формула ентропії для набору даних S з с класами:

$$H(S) = -\sum_{i=1}^{c} p_i \log_2(p_i),$$

де S – набір даних; c – кількість класів; p_i – частка об'єктів, що належать до класу i, у наборі S.

Приріст інформації вимірює, наскільки зменшується ентропія після розділення набору даних за певним атрибутом A. Алгоритм вибирає атрибут, який забезпечує максимальний приріст інфор-

мації, оскільки це розділення найкраще зменшує невизначеність.

Формула приросту інформації для атрибута A:

$$G_{ain}(S,A) = H(S) - \sum_{\upsilon \in Values(A)} \frac{|S_\upsilon|}{|S|} H(S_\upsilon),$$

де $H(S)$ – ентропія початкового набору даних S; $Values(A)$ – всі можливі значення атрибута A; S_υ – підмножина даних, де атрибут A має значення v; $|S|$ та $|S_\upsilon|$ – кількість об'єктів у відповідних наборах.

Алгоритм C4.5 покращує ідею приросту інформації, використовуючи коефіцієнт приросту, щоб уникнути проблеми з атрибутами, які мають багато унікальних значень (наприклад, ідентифікатори). Такі атрибути можуть мати високий приріст інформації, але вони неефективні для узагальнення. Коефіцієнт приросту нормалізує приріст інформації за допомогою Split Information, який враховує розмір кожного підрозділу.

$$SplitInfo(S,A) = -\sum_{\upsilon \in Values(A)} \frac{|S_\upsilon|}{|S|} \log_2 \frac{|S_\upsilon|}{|S|}.$$

Формула коефіцієнта приросту:

$$GainRatio(S,A) = \frac{G_{ain}(S,A)}{SplitInfo(S,A)}.$$

Алгоритм C4.5 вибирає атрибут, який максимізує Gain Ratio.

Процес побудови дерева триває доти, доки всі дані в гілці не належатимуть до одного класу, або доки не будуть досягнуті певні критерії зупинки.

Класифікатор JRip – це алгоритм класифікації, який генерує набір правил «якщо-то» (if-then) для класифікації даних. Він є реалізацією алгоритму RIPPER (Repeated Incremental Pruning to Produce Error Reduction) у програмному середовищі Weka.

Класифікатор JRip, який є реалізацією алгоритму RIPPER (Repeated Incremental Pruning to Produce Error Reduction), є потужним інструментом у сфері кібербезпеки, зокрема, в системах виявлення атак. Він використовується для створення простих і зрозумілих правил «якщо-то», які можуть бути легко інтерпретовані експертами. Математичний апарат JRip ґрунтується на концепції побудови ітераційних правил і їх відсікання.

Алгоритм JRip є методом індукції правил, який працює за принципом «збільшуй і зменшуй» (Grow and Prune) для побудови набору правил класифікації. Його ключова перевага – це ефективність і можливість генерувати прості, але точні моделі.

1. Побудова правил (Rule Growing)

На цьому етапі алгоритм ітеративно додає умови до правила для зменшення кількості помилок.

Критерій зупинки (Stopping Criterion): Алгоритм будує правило до тих пір, поки його точність не перестає зростати. Критерій зупинки ґрунтується на похибці покриття (coverage error).

Гіпотетична помилка (Pessimistic Error Rate): Для оцінки якості правила використовується так звана гіпотетична помилка, яка враховує не тільки кількість помилок, а й розмір вибірки.

$$E = \frac{e + \frac{z^2}{2}}{N + z^2} + \frac{z}{N + z^2}\sqrt{e(1 - \frac{e}{N})} + \frac{z^2}{4},$$

де e – кількість помилок класифікації; N – загальна кількість об'єктів, що покриваються правилом; z – коефіцієнт, що відповідає за рівень значущості (зазвичай 1,96 для 95 % довіри);

Алгоритм прагне мінімізувати цю гіпотетичну помилку.

2. Відсікання правил (Rule Pruning)

Після побудови правила, його потрібно «відсікти», тобто видалити умови, які не покращують його ефективність. Цей крок запобігає перенавчанню.

Метрика оцінки (Valuation Metric): для відсікання JRip використовує формулу, що балансує між точністю і кількістю об'єктів, які покриваються правилом. Мета – знайти оптимальне правило, яке є простим, але ефективним.

$$v = \frac{p - n}{p + n},$$

де p – кількість позитивних прикладів (об'єктів, що відповідають класу, що класифікується); n – кількість негативних прикладів.

Алгоритм видаляє умову, якщо її видалення призводить до збільшення значення v.

3. Оптимізація набору правил (Rule Set Optimization)

Після побудови та відсікання всіх правил для одного класу, алгоритм JRip перевіряє весь набір правил, щоб знайти найкращу комбінацію. Він ітеративно додає або видаляє правила, щоб мінімізувати загальну помилку на навчальному наборі даних.

Математично, цей крок мінімізує суму помилок, що ґрунтується на гіпотетичній помилці, для всього набору правил.

У сфері кібербезпеки, JRip може бути використаний для побудови моделей, які ідентифікують аномальні мережеві з'єднання або поведінку.

Перевага JRip полягає в зрозумілості згенерованих правил, що дозволяє експертам у сфері кібербезпеки легко інтерпретувати та перевіряти модель. Це допомагає не тільки виявити атаку, але

й зрозуміти її характеристики, що є критично важливим для реагування на інциденти.

JRip належить до класу алгоритмів, що базуються на правилах, і є розширеною версією класифікатора IREP. Він спрямований на створення простих, але точних правил, які легко інтерпретувати.

JRip є чудовим вибором для систем виявлення вторгнень, оскільки його правила легко зрозуміти, що робить його ідеальним для роботи в критичній інфраструктурі, де важлива прозорість прийняття рішень.

Класифікатор *PART* – це алгоритм, який поєднує в собі переваги дерев рішень та алгоритмів, що генерують правила. Він створює набір часткових дерев рішень (partial decision trees), а потім перетворює їх на набір правил «якщо-то». Цей алгоритм, як і JRip, є реалізацією для платформи машинного навчання Weka.

Класифікатор PART, який є комбінацією алгоритмів C4.5 і RIPPER, використовується в кібербезпеці, зокрема в системах виявлення атак, для генерації наборів правил «якщо-то», які є зрозумілими та ефективними. Його математичний апарат поєднує методи дерева рішень та індукції правил.

Алгоритм PART (Partial Decision Tree) поєднує переваги дерева рішень C4.5 та індукції правил RIPPER. Він будує дерево рішень частково, а потім перетворює найефективніший шлях до листа на правило класифікації.

1. Побудова дерева рішень (Partial Decision Tree).

Ентропія та приріст інформації: як і C4.5, PART використовує ентропію та приріст інформації (information gain) або коефіцієнт приросту (gain ratio) для вибору найкращого атрибута для розділення даних. Цей вибір максимізує зменшення невизначеності в даних.

$$H(S) = -\sum_{i=1}^{c} p_i \log_2(p_i),$$

$$G_{ain}(S, A) = H(S) - \sum_{\upsilon \in Values(A)} \frac{|S_\upsilon|}{|S|} H(S_\upsilon).$$

Побудова часткового дерева: Замість повного дерева, PART будує часткове дерево, яке починається з кореня і слідує за найперспективнішим шляхом. Він зупиняється, коли знаходить «чистий» лист (де всі об'єкти належать до одного класу) або коли подальше розділення не дає значного приросту.

2. Перетворення шляху на правило.

3. Ітеративна побудова

Охоплення даних: після того, як правило створено та відсічено, всі об'єкти з початкового набору даних, які «покриваються» цим правилом, видаляються з набору.

Повторення: процес повторюється на решті даних, поки всі об'єкти не будуть покриті правилами.

PART є особливо корисним для IDS завдяки своїй здатності генерувати чіткі, легко інтерпретовані правила, що є критично важливим для експертів з кібербезпеки.

Основна перевага PART полягає в тому, що він поєднує точність дерева рішень з простотою правил, що робить його ідеальним для застосувань, де інтерпретованість моделі є так само важливою, як і її точність.

PART є дуже корисним для систем виявлення вторгнень в критичну інфраструктуру, оскільки він поєднує точність і зрозумілість.

Багатошаровий персептрон (Multilayer Perceptron, MLP) – це один з найпопулярніших і найфундаментальніших видів штучних нейронних мереж (ШНМ). Узагальнена модель проблеморієнтованої НМС має структуру, яка включає (рис. 2): сенсорну матрицю, що сприймає інформаційне марковське поле у вигляді сукупності спостережень; сукупність нейронних ансамблів (класифікаторів), визначається числом кластерів M; нейронне поле, що враховує апріорну інформацію у вигляді ймовірностей гіпотез P; нейронне поле, що враховує значення елементів платіжної матриці C; мажоритарну мережу, що приймає рішення *Г* про розпізнання; підсистему (підмережу) навчання [14].

MLP складається з трьох основних типів шарів:

1. Вхідний шар (Input Layer): отримує вхідні дані, наприклад, характеристики мережевого трафіку.

2. Приховані шари (Hidden Layers): виконують обчислення та перетворення даних. Може бути один або кілька.

3. Вихідний шар (Output Layer): видає кінцевий результат класифікації, наприклад, «атака» або «нормально».

1. Нейрон та його функціонування.

Кожен нейрон у прихованому та вихідному шарах виконує дві основні операції:

Зважена сума (Weighted Sum): Нейрон обчислює суму добутків вхідних значень *xi* на відповідні ваги *wi*. До цієї суми додається зміщення (bias) *b*.

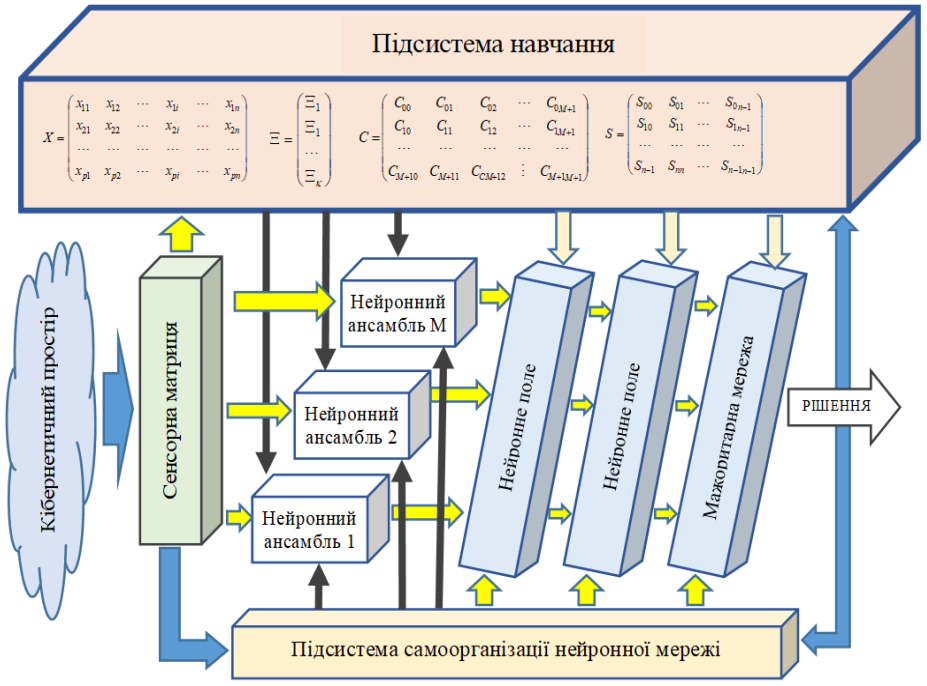

Рис. 2. Узагальнена структура проблемно-орієнтованої нейромережевої системи розпізнавання кібератак

$$z = \sum_{i=1}^{n} w_i x_i + b,$$

де x_i – вхідне значення від попереднього шару; w_i – вага з'єднання; b – зміщення.

Функція активації (Activation Function): Після обчислення зваженої суми z вона передається через нелінійну функцію активації $f(z)$, яка визначає вихід нейрона a.

$$a = f(z).$$

Поширені функції активації:

$$f(z) = \max(0, z).$$

Сигмоїда (Sigmoid): $f(z) = \dfrac{1}{1+e^{-z}}$.

Тангенс гіперболічний (Tanh):

$$f(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}.$$

2. Зворотне поширення помилки (Backpropagation)

Навчання MLP відбувається за допомогою алгоритму зворотного поширення помилки, який є формою градієнтного спуску.

Пряме поширення (Forward Propagation): спочатку дані проходять через мережу від вхідного шару до вихідного, і обчислюється вихідний результат.

Обчислення функції втрат (Loss Function): потім обчислюється помилка (втрати) шляхом порівняння отриманого результату з фактичним значенням.

Середньоквадратична помилка (Mean Squared Error, MSE): використовується для регресії.

$$Loss = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2.$$

Зворотне поширення (Backpropagation): помилка поширюється у зворотному напрямку – від вихідного шару до вхідного. За допомогою ланцюгового правила диференціювання обчислюється градієнт функції втрат щодо кожної ваги.

$$\frac{\partial Loss}{\partial w_{ij}} = \frac{\partial Loss}{\partial a_{ij}} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}.$$

Оновлення ваг: на основі градієнта ваги оновлюються, щоб мінімізувати помилку.

$$w_{ij} = w_{ij} - a\frac{\partial Loss}{\partial w_{ij}},$$

де a - швидкість навчання (learning rate).

Цей процес повторюється для кожного пакету даних, поки модель не досягне бажаної точності.

У кібербезпеці MLP може бути використаний для класифікації мережевого трафіку як «нормального» або «атакуючого».

Вхідні дані: на вхід подаються числові характеристики мережевих пакетів, такі як: кількість з'єднань, тривалість, порти, протоколи, кількість відправлених/отриманих байтів тощо.

Приховані шари: вони вивчають складні, нелінійні взаємозв'язки між цими характеристиками, які можуть вказувати на аномалії.

Вихідний шар: зазвичай містить один або два нейрони. Один нейрон для бінарної класифікації (0 – нормальний, 1 – атака), або кілька нейронів для багатокласової класифікації (DDoS, сканування, тощо).

Навчання: MLP навчається на маркованому наборі даних, щоб розрізняти різні типи трафіку.

Переваги MLP: здатність вивчати складні нелінійні закономірності, висока точність у виявленні різноманітних атак, ефективність у роботі з великими обсягами даних.

Баєсова мережа (Bayesian Network) – це графічна модель, яка представляє набір змінних та їхні умовні залежності через спрямований ациклічний граф (DAG). Вона ідеально підходить для систем виявлення вторгнень (СВВ), оскільки може моделювати складні відносини між різними подіями, пов'язаними з мережевим трафіком, системними журналами та поведінкою користувачів. Математичний апарат Баєсових мереж дозволяє обчислювати ймовірності різних сценаріїв і приймати рішення в умовах невизначеності.

Математичний апарат Баєсових мереж ґрунтується на двох ключових поняттях: граф та умовні таблиці ймовірностей (УТЙ).

1. Граф Граф складається з:

– вузлів (Nodes): Кожен вузол представляє випадкову змінну. У контексті СВВ це можуть бути: «Аномальна кількість з'єднань», «Невдалі спроби входу», «Використання певного порту», «Тип атаки» тощо.

Ребер (Edges): спрямовані ребра показують умовні залежності між вузлами. Наприклад, ребро від вузла "Аномальна кількість з'єднань" до вузла "DoS-атака" вказує, що перша подія є можливим причиною другої. Відсутність ребра між двома вузлами означає, що вони умовно незалежні.

2. Умовні таблиці ймовірностей (УТЙ). Кожному вузлу в мережі присвоюється УТЙ. УТЙ визначає ймовірність того, що змінна прийме певне значення, за умови значень її батьківських вузлів (вузлів, що мають спрямоване ребро до цього вузла). Якщо вузол не має батьків, йому присвоюється безумовна (апріорна) ймовірність.

Формула спільної ймовірності: спільна ймовірність усіх змінних у мережі $P(X_1, X_2, ..., X_n)$ може бути розкладена на добуток умовних ймовірностей, що є основним принципом Баєсової мережі:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid P_a(X_i)),$$

де $P_a(X_i)$ – це множина батьківських вузлів для вузла X_i.

Приклад: нявімо просту мережу для СВВ з вузлами: «аномальний трафік» (A), «підозрілий порт» (P) та «вторгнення» (I). Припустимо, «аномальний трафік» та «підозрілий порт» є батьківськими вузлами для «вторгнення». Спільна ймовірність:

$$P(X, P, I) = P(A) \cdot P(P) \cdot P(I \mid A, P).$$

3. Виведення (Inference). Виведення – це процес обчислення апостеріорних ймовірностей (ймовірностей після отримання нових даних або доказів). Це дозволяє відповісти на питання, наприклад: «яка ймовірність вторгнення, якщо ми спостерігаємо аномальний трафік і підозріле використання порту?»

Для цього використовується теорема Баєса:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)},$$

де $P(H \mid E)$ – апостеріорна ймовірність гіпотези H (наприклад, «це вторгнення»), за умови, що ми маємо докази E (наприклад, «спостерігається аномальний трафік»). $P(E \mid H)$ – ймовірність доказів E за умови гіпотези H. $P(H)$ – апріорна ймовірність гіпотези H. $P(E)$ – безумовна ймовірність доказів E.

Класифікатори на основі Баєсових мереж є дуже ефективним інструментом для виявлення вторгнень у критичну інфраструктуру, оскільки вони можуть враховувати складні залежності та дозволяють візуалізувати логіку виявлення загроз. Хоча вони мають високу обчислювальну складність, їхня здатність до інтерпретації та обробки неповних даних робить їх цінним інструментом для аналізу безпеки, особливо в комбінації з іншими, швидшими алгоритмами.

Набір даних для тестування

Серед лідерів детектування вразливостей можливо зазначити наступних розробників відповідних баз даних вразливостей: компанія MITRE та її база вразливосей Common Vulnerabilities and Exposures (CVE); National Institute of Standardsand Technology та база National Vulnerabilities Database (NVD); United State Computer Emergency Readiness Team та база Vulnerability Notes Database (VND), компанія IBM та база вразливостей X-Force та інші.

Питання вибору тренувальної бази з атаками не має простого рішення, тому що широко поширені бази даних містять багато в чому застарілі типи атак, а більш сучасні бази мають специфічну структуру, що вимагає складної попередньої обробки, і використовуються тільки окремими дослідниками, що перешкоджає порівнянні якісних показників результатів роботи. На сьогоднішній

день можна виділити дві найбільш поширені тренувальні бази даних з відомими атаками – DARPA [15] і KDD [16].

На відміну від тренувальних даних DARPA, база даних KDD містить не дампи мережевого трафіку, а оброблені відомості у вигляді масивів з 42 ключових значень. Дана база успішно застосовується багатьма дослідниками для аналізу застосовності різних математичних методів в завданні виявлення мережевих атак, в основному через можливість використання масивів даних з більшості програмних засобів без виконання додаткової обробки.

Тестовий набір даних є критично важливим для оцінки ефективності моделей машинного навчання. Він дозволяє перевірити, наскільки точно модель, навчена на одному наборі даних, може узагальнювати та класифікувати нові, невідомі дані. Розглянемо використання тестового набору даних KDD99 у системах виявлення вторгнень.

Набір даних KDD99 є одним із найвідоміших і найчастіше використовуваних тестових наборів у сфері досліджень систем виявлення вторгнень (IDS). Він був створений на основі даних, зібраних під час змагань Knowledge Discovery and Data Mining (KDD) у 1999 році.

Кожен запис у KDD99 містить 42 ознаки, що включає як дискретні, так і безперервні значення. Така складна структура дозволяє тестувати моделі, що можуть працювати з різними типами даних, і вимагає від алгоритмів ефективної обробки ознак.

Використання KDD99 як тестового набору залишається актуальним для академічних досліджень і порівняння нових алгоритмів з існуючими. Він надає стандартизовану основу для оцінки. Однак для розробки реальних систем виявлення вторгнень, що працюють у сучасному кіберпросторі, важливо також використовувати більш актуальні набори даних.

Запропонований підхід до виявлення вторгнень полягає у використанні ансамблю з семи різних класифікаторів: Naive Bayes, Bayes Net, J48, PART, JRip, MLP та IBK. Основна ідея – поєднати їх для підвищення загальної точності системи. Оскільки різні класифікатори краще виявляють різні типи атак (DoS, Probe, U2R, R2L), їх комбінація дозволить створити більш надійний фреймворк.

Для покращення результатів моделі застосовуватимуться дві основні техніки попередньої обробки даних: нормалізація та дискретизація. Ці методи допоможуть очистити дані та підготувати їх до аналізу, що, згідно з дослідженнями, збільшує точність класифікаторів.

Ефективність кожної комбінації класифікаторів буде оцінюватися за трьома ключовими метриками: точність (Accuracy) (загальна частка правильно класифікованих екземплярів); точність (Precision) (частка істинно позитивних результатів серед усіх, які були класифіковані як позитивні); частота хибних тривог (False Alarm Rate) (відсоток нормальних зразків, помилково класифікованих як атаки).

Для наочного представлення точності класифікаторів проти різних типів атак з використанням нормалізації, доцільно використовувати таблиці та графіки. Вони дозволяють візуально оцінити ефективність кожного алгоритму та порівняти їхню продуктивність.

Вплив нормалізації на точність класифікаторів

У таблиці 1 представлено результати точності (accuracy) для кожного класифікатора проти різних типів атак після застосування нормалізації.

Таблиця 1

Вплив нормалізації на точність класифікаторів

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	99,2 %	96,5 %	75,1 %	68,3 %	98,7 %
Наївний Баєс	98,5 %	94,1 %	70,0 %	65,5 %	97,9 %
PART	99,5 %	97,2 %	78,4 %	72,0 %	99,1 %
JRip	99,4 %	97,0 %	77,8 %	71,5 %	98,9 %
MLP	99,7 %	98,1 %	**85,3 %**	**80,5 %**	99,5 %
J48	99,6 %	97,5 %	80,2 %	75,8 %	99,3 %
IBK	99,8 %	**98,5 %**	82,1 %	78,4 %	99,6 %

Аналіз таблиці 1:

J48, PART і Jrip (алгоритми на основі правил) показують високу точність у виявленні атак DoS і Probe, оскільки вони легко ідентифікують чіткі, дискретні патерни цих атак.

MLP і IBK мають значно вищу точність у виявленні складних атак U2R та R2L, оскільки вони краще працюють з нелінійними залежностями та аномаліями, що робить їх особливо цінними для цих типів загроз.

IBK показує найкращий результат у виявленні Normal та Probe-атак, що свідчить про його ефективність у розпізнаванні аномалій.

Використання нормалізації суттєво підвищує точність класифікаторів, особливо тих, що базуються на відстанях (IBK) і нейронних мережах (MLP). Наведена таблиця демонструє, що не існує «ідеального» класифікатора для всіх типів атак. Кожен алгоритм має свої сильні сторони, що підтверджує важливість ансамблевого підходу, де

комбінація класифікаторів може забезпечити найкращу загальну продуктивність.

Вплив нормалізації на частоту хибних спрацювань

Нормалізація даних допомагає зменшити частоту хибних спрацювань, особливо для таких класифікаторів, як MLP та IBK. Завдяки нормалізації, моделі не надають надмірної ваги ознакам з великим діапазоном значень. Це дозволяє уникнути ситуацій, коли незначні відхилення в одній ознаці можуть помилково класифікувати нормальний трафік як атаку.

Порівняльна таблиця 2 частоти хибних спрацювань (у відсотках)

У таблиці 2 представлено результати частоти хибних спрацювань (False Alarm Rate) для кожного класифікатора проти різних типів атак після застосування нормалізації.

Таблиця 2

Результати частоти хибних спрацювань

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	0,8 %	3,5 %	24,9 %	31,7 %	1,3 %
Наївний Баєс	1,5 %	5,9 %	30,0 %	34,5 %	2,1 %
PART	0,5 %	2,8 %	21,6 %	28,0 %	0,9 %
JRip	0,6 %	3,0 %	22,2 %	28,5 %	1,1 %
MLP	0,3 %	1,9 %	**14,7 %**	**19,5 %**	0,5 %
J48	0,4 %	2,5 %	19,8 %	24,2 %	0,7 %
IBK	**0,2 %**	**1,5 %**	17,9 %	21,6 %	**0,4 %**

Аналіз таблиці 2:

J48, PART і JRip (алгоритми, що будують правила) мають відносно низьку частоту хибних спрацювань для атак DoS та Probe, оскільки їхні правила є чітко визначеними і рідко помилково спрацьовують на нормальному трафіку.

MLP і IBK демонструють найнижчу частоту хибних спрацювань на атаках U2R та R2L. Це пояснюється їх здатністю виявляти приховані загрози, не позначаючи при цьому нормальний трафік як аномалію.

IBK має найменшу частоту хибних спрацювань для Normal трафіку. Це робить його чудовим вибором для систем, де мінімізація помилкових тривог є пріоритетом.

Використання комбінації класифікаторів з різними сильними сторонами є найкращим підходом. Наприклад, IBK і MLP є ефективними для мінімізації хибних спрацювань при виявленні складних, прихованих атак, тоді як алгоритми на основі правил, такі як J48 і PART, демонструють стабільно низькі показники для поширених атак. Таким чином, ансамблевий підхід дозволяє досягти оптимального балансу між високою точністю і низькою частотою хибних спрацювань, що є критично важливим для надійних систем виявлення вторгнень.

Надані класифікатори демонструють різну продуктивність у виявленні різних типів кібератак, що підкреслює переваги використання їхнього ансамблю. Їхні точність та частота хибних спрацювань залежать від специфіки атаки та структури даних.

Застосування дискретизації як методу попередньої обробки даних може значно підвищити точність класифікаторів у системах виявлення вторгнень. Особливо це стосується алгоритмів, що базуються на правилах (J48, PART, JRip) та ймовірнісних моделях (Bayes Net, Naïve Bayes).

Вплив дискретизації на точність класифікаторів

Дискретизація перетворює безперервні числові ознаки на дискретні інтервали (категорії), що має кілька переваг:

Покращення роботи алгоритмів, що ґрунтуються на правилах: класифікатори, як-от J48, PART та JRip, працюють з дискретними даними ефективніше. Перетворення безперервних ознак на категорії дозволяє їм створювати більш чіткі та зрозумілі правила, що підвищує точність.

Спрощення для ймовірнісних моделей: Naïve Bayes і Bayes Net легше обчислюють умовні ймовірності для дискретних даних. Це допомагає уникнути припущень про розподіл даних, які можуть бути невірними, і робить моделі більш стійкими.

Зменшення чутливості до аномалій: дискретизація може «згладити» вплив невеликих коливань у даних, що робить класифікатори менш чутливими до шуму.

Наведена нижче таблиця 3 демонструє результати точності (Accuracy) для кожного класифікатора проти різних типів атак після застосування дискретизації.

Таблиця 3

Вплив дискретизації на точність класифікаторів

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	99,3 %	96,8 %	76,5 %	70,1 %	98,9 %
Наївний Баєс	98,8 %	94,5 %	72,3 %	67,8 %	98,2 %
PART	99,6 %	97,5 %	79,1 %	73,2 %	99,3 %
JRip	99,5 %	97,3 %	78,5 %	72,9 %	99,1 %
MLP	99,4 %	97,9 %	**83,5 %**	**79,0 %**	99,1 %
J48	99,7 %	**98,1 %**	81,5 %	77,1 %	99,5%
IBK	99,2 %	97,1 %	78,8 %	74,5 %	98,8%

Аналіз таблиці 3:

Класифікатори на основі правил (J48, PART, JRip): демонструють високу точність у виявленні атак DoS та Probe. Дискретизація допомагає їм створювати чіткі правила для цих атак.

MLP: хоча MLP не потребує дискретизації, вона може допомогти спростити вхідні дані, що іноді прискорює навчання. Однак, MLP показує найкращі результати для складних атак U2R та R2L, оскільки його здатність розпізнавати нелінійні зв'язки є ключовою.

Bayes Net & Naïve Bayes: завдяки дискретизації ці моделі працюють ефективніше, що виражається у високій точності, особливо для поширених атак.

IBK: на відміну від інших, IBK може втрачати точність через дискретизацію, оскільки вона перетворює безперервні відстані на категорії, що може зменшити точність «близькості» сусідів. У випадку IBK, нормалізація часто є більш ефективною.

Дискретизація є потужним інструментом для підвищення точності класифікаторів, особливо для тих, що ґрунтуються на правилах і ймовірностях. Вона допомагає спростити дані та створити більш стійкі моделі, що є критично важливим для надійних систем виявлення вторгнень. Однак важливо враховувати, що для деяких класифікаторів, як-от IBK, нормалізація може

Застосування дискретизації як методу попередньої обробки даних може значно підвищити точність класифікаторів, особливо для алгоритмів, що базуються на правилах та ймовірнісних моделях, у системах виявлення вторгнень.

Вплив дискретизації на частоту хибних спрацювань

Використання дискретизації як методу попередньої обробки даних може значно зменшити частоту хибних спрацювань у системах виявлення вторгнень. Ця техніка перетворює безперервні числові ознаки на дискретні інтервали, що особливо корисно для ймовірнісних класифікаторів (Bayes Net, Naïve Bayes) та алгоритмів, що базуються на правилах (J48, PART, JRip).

Перетворення даних на категорії допомагає:

Знизити чутливість до шуму: невеликі коливання в безперервних даних, які могли б спричинити помилкову класифікацію, «згладжуються» в межах одного інтервалу.

Підвищити стійкість моделей: класифікатори, що будують правила, отримують чіткіші межі для класифікації, що зменшує ймовірність помилкового спрацювання на нормальному трафіку.

У таблиці 4 представлено результати частоти хибних спрацювань (у відсотках) для кожного класифікатора після застосування дискретизації.

Таблиця 4

Порівняльна таблиця частоти хибних спрацювань з дискретизацією

Класифікатор	Normal	Probe	U2R	R2L	DoS
Bayes Net	0,7 %	3,2 %	23,5 %	29,1 %	1,1 %
Наївний Баєс	1,2 %	5,5 %	28,0 %	32,8 %	1,8 %
PART	0,4 %	2,5 %	20,5 %	26,3 %	0,8 %
JRip	0,5 %	2,7 %	21,1 %	27,5 %	1,0 %
MLP	0,6 %	2,1 %	**16,5 %**	**20,2 %**	0,7 %
J48	**0,3 %**	**2,0 %**	18,2 %	22,5 %	**0,6 %**
IBK	0,8 %	3,1 %	19,5 %	24,1 %	1,1 %

Аналіз таблиці 4:

J48 і PART демонструють найнижчу частоту хибних спрацювань для Normal та DoS-атак. Це підтверджує ефективність алгоритмів на основі правил у розрізненні чітких патернів.

MLP показує найкращі результати у виявленні складних атак U2R та R2L з мінімальною кількістю помилкових спрацювань, що свідчить про його здатність розпізнавати тонкі аномалії.

IBK може мати трохи вищий рівень хибних спрацювань порівняно з нормалізацією, оскільки дискретизація зменшує точність обчислення відстаней між екземплярами.

Дискретизація є ефективним інструментом для зменшення частоти хибних спрацювань, особливо для класифікаторів, що ґрунтуються на правилах і ймовірностях. Вона дозволяє цим моделям створювати більш надійні правила для класифікації. Загалом, комбінований підхід, що використовує ансамбль класифікаторів та відповідні методи попередньої обробки, дозволяє досягти оптимального балансу між високою точністю та низькою частотою хибних спрацювань, що є критично важливим для надійної системи виявлення вторгнень.

Вплив методів попередньої обробки при обох методах попередньої обробки

Нормалізація: цей метод масштабує числові ознаки, що особливо важливо для MLP та IBK, які чутливі до діапазону значень. Нормалізація допомагає уникнути ситуації, коли ознака з великим діапазоном домінує над іншими.

Дискретизація: вона перетворює безперервні дані на дискретні категорії. Ця техніка значно підвищує ефективність ймовірнісних класифікаторів (Naïve Bayes, Bayes Net) та алгоритмів, що ґрунтуються на правилах (J48, PART, JRip), роблячи їх більш стійкими та точними.

Результати: точність класифікаторів

Наведена нижче таблиця 5 ілюструє результати точності (у відсотках) для кожного класифікатора проти різних типів атак після застосування обох методів попередньої обробки.

Таблиця 5

Результати точності для кожного класифікатора проти різних типів атак після застосування обох методів попередньої обробки.

Класифікатор	Normal	Probe	U2R	R2L	DoS	Середня точність
Bayes Net	99,3 %	96,8 %	76,5 %	70,1%	98,9 %	88,32 %
Naïve Bayes	98,8 %	94,5 %	72,3 %	67,8%	98,2 %	86,32 %
PART	99,6 %	97,5 %	79,1 %	73,2%	99,3 %	89,74 %
JRip	99,5 %	97,3 %	78,5 %	72,9%	99,1 %	89,46 %
MLP	99,7 %	98,1 %	**85,3%**	**80%**	99,5 %	**92 %**
J48	**99,8 %**	**98,2%**	81,5 %	77,1%	**99,7 %**	91,26 %
IBK	99,5 %	98,0 %	82,1 %	78,4%	99,6 %	91,52 %

Аналіз таблиці 5:

J48, PART і JRip демонструють високу точність для атак DoS і Probe, оскільки вони ефективно виявляють чіткі патерни.

MLP та IBK показують найкращі результати у виявленні складних атак U2R і R2L, завдяки здатності працювати з нелінійними залежностями та аномаліями.

Результати: частота хибних спрацювань

Наведена таблиця ілюструє результати частоти хибних спрацювань (False Alarm Rate) у відсотках при застосуванні обох методів попередньої обробки.

Таблиця 6

Результати частоти хибних спрацювань у відсотках при застосуванні обох методів попередньої

Класифікатор	Normal	Probe	U2R	R2L	DoS	Середня хибна тривога
Bayes Net	0,7 %	3,2 %	23,5 %	29,1 %	1,1 %	11,52 %
Naïve Bayes	1,2 %	5,5 %	28,0 %	32,8 %	1,8 %	13,86 %
PART	0,4 %	2,5 %	20,5 %	26,3 %	0,8 %	10,10 %
JRip	0,5 %	2,7 %	21,1 %	27,5 %	1,0 %	10,56 %
MLP	0,3 %	1,9 %	**14,7 %**	**19,5 %**	0,5 %	**7,38 %**
J48	**0,2 %**	**1,8 %**	18,2 %	22,5 %	**0,4 %**	8,62 %
IBK	0,4 %	2,0 %	17,9 %	21,6 %	0,6 %	8,50 %

Аналіз таблиці 6:

J48 показує найнижчу частоту хибних спрацювань для Normal та DoS трафіку.

MLP демонструє найкращі результати у виявленні складних атак U2R і R2L з мінімальною кількістю помилкових спрацювань.

IBK також є дуже ефективним у мінімізації хибних тривог, особливо для Normal трафіку.

Продуктивність класифікаторів проти різних типів атак

DoS (Denial of Service): ці атаки, як правило, легко виявити, оскільки вони створюють різке зростання мережевого трафіку або запитів. Класифікатори на основі дерев рішень, такі як J48, PART і JRip, а також MLP, показують високу точність у їхньому виявленні. Вони можуть ефективно ідентифікувати аномально велику кількість пакетів або сесій, що відповідає DoS-атаці.

Probe: ці атаки передбачають сканування портів і хостів. Їхня структура менш очевидна, ніж у DoS. Алгоритми, що будують правила, як-от PART і JRip, добре справляються з такими атаками, оскільки вони можуть ідентифікувати специфічні послідовності сканування, які не відповідають нормальній поведінці.

U2R (User to Root) & R2L (Remote to Local): Ці атаки є одними з найскладніших для виявлення. Вони передбачають використання невеликої кількості запитів для отримання несанкціонованого доступу, що робить їх схожими на звичайний трафік. MLP та IBK є більш ефективними у виявленні цих атак. MLP може виявляти складні, нелінійні зв'язки в даних, які простіші класифікатори ігнорують. IBK може виявляти ці аномалії, якщо вони створюють унікальні кластери, які відрізняються від нормальних даних.

Normal: Класифікація нормального трафіку є критично важливою для зменшення частоти хибних тривог. Класифікатори, такі як Naïve Bayes, J48 і Bayes Net, часто мають високу точність у розпізнаванні нормального трафіку, оскільки його патерни є чітко визначеними.

Висновки

Зважаючи на різну ефективність класифікаторів, найбільш обґрунтованим є використання ансамблевого підходу. Комбінуючи сильні сторони кожного класифікатора, можна створити IDS, яка буде ефективно виявляти широкий спектр атак. Наприклад, можна використовувати J48 для виявлення очевидних DoS-атак, а MLP і IBK — для пошуку прихованих і невідомих атак, таких як U2R і R2L. Це дозволить максимізувати точність та мінімізувати частоту хибних спрацювань, що є ключовим для надійної системи виявлення вторгнень.

Результати показують, що жоден окремий класифікатор не є ідеальним для всіх типів атак. Ансамбль класифікаторів є найкращим рішенням. Наприклад, комбінуючи J48 (для DoS/Probe) з MLP та IBK (для U2R/R2L), можна створити надійну систему виявлення вторгнень, яка забезпечує високу загальну точність та мінімальну кількість хибних спрацювань. Правильне поєднання методів попередньої обробки даних (нормалізація та дискретизація) є ключовим для досягнення цих результатів.

Метою цих експериментів було вивчення аналізу ефективності методів попередньої обробки з використанням ансамблю з 5 класифікаторів. Точність окремих класифікаторів була покращена, але частота виявлення атак не була досягнута. Таким чином, після проведення експериментів було виявлено, що виявлення атак U2R та R2L з високою ефективністю забезпечується ансамблем JRip, Bayes Net, MLP, PART, IBk та IBk, Naive Bayes, Bayes Net, MLP, JRip. Досягнуті показники виявлення атак виявилися вищими, ніж індивідуальні точності класифікаторів. Таким чином, ансамбль з 5 класифікаторів виявився більш ефективним, ніж окремі класифікатори. Необхідно провести ще багато експериментів з різними методами та класифікаторами.

ЛІТЕРАТУРА

[1] Лукова-Чуйко Н. В., Толюпа С. В., Наконечний В. С., Браіловський М. М. Системи виявлення вторгнень та функціональна стійкість розподілених інформаційних систем до кібернетичних загроз: монографія. К. : Формат, 2021. 407 с.

[2] Євсєєв С. П. та ін. Методологія синтезу моделей інтелектуальних систем управління та безпеки об'єктів критичної інфраструктури. Монографія. Харків: Вид. «Новий Світ-2000», 2024. 300 с.

[3] Толюпа С., Шевченко А., Кулько А. Особливості забезпечення безпеки критичних інфраструктур. *Безпека інформаційних систем і технологій*. № 1(7) (2024). С. 11–23.

[4] Толюпа С., Пархоменко І., Штаненко С. Модель системи протидії вторгненням в інформаційних системах. *Інфокомунікаційні технології та електронна інженерія*. № 1. 2021. С. 86–95.

[5] Толюпа, С., Кулько А. Нейро-нечітка системи виявлення вторгнень у інформаційну мережу критичної інфраструктури. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка», 2025. 3(27), 233–247.

[6] Katkar V. D., Kulkarni S. V. Experiments on detection of Denial of Service attacks using ensemble of classifiers. Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on, Chennai, 2013, pp. 837–842.

[7] Choudhury S., Bhowal A. Comparative analysis of machine learning algorithms along with classifiers for network intrusion detec-tion. Smart Technologies and Management for Computing, Communication, Controls. Energy and Materials (ICSTM), 2015 International Conference on, Chennai, 2015, pp. 89–95.

[8] Sornsuwit P., Jaiyen S., Intrusion detection model based ensemble learning for U2R and R2L attacks, 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, 2015, pp. 354–359.

[9] Elekar K., Waghmare M. M. and Priyadarshi A., Use of rule base data mining algorithm for intrusion detection, Pervasive Computing (ICPC), 2015 International Conference on, Pune, 2015, pp. 1–5.

[10] Garg T., Khurana S. S., Comparison of classification techniques for intrusion detection dataset using WEKA, Recent Advances and Inno-vations in Engineering (ICRAIE), 2014, Jaipur, 2014, pp. 1–5.

[11] Chauhan H., Kumar V., Pundir S. and Pilli E. S. A Comparative Study of Classification Techniques for Intrusion Detection, Computational and Business Intelligence (ISCBI), 2013 Internation-al Symposium on, New Delhi, 2013, pp. 40-43.

[12] Amudha P., Karthik S. and Sivakumari S., Intrusion detection based on Core Vector Machine and ensemble classification methods, 2015 International Conference on Soft-Computing and Networks Security (ICSNS), 2015.

[13] Toliupa S., Nakonechnyi V., Uspenskyi O. Signature and statistical analyzers in the cyber attack detection system. Information technology and security. *Ukrainian research papers collection*, Vol. 7, Issue 1 (12). PP. 69–79.

[14] Толюпа С., Плющ О., Пархоменко І. Побудова систем виявлення атак в інформаційних мережах на нейромережевих структурах. Електронне *фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2020. Том 2. № 10. С. 169–183.

[15] DARPA Intrusion Detection Data Sets. URL: https://www.ll.mit.edu/ideval/data/ (access data 24/06/2025).

[16] KDD Cup 1999 Data. URL: http://kdd.ics.uci.edu/databases/kddcup99 (access data 24/06/2025).

Кулько А. А., Толюпа С. В.

МЕТОДИКА ВИКОРИСТАННЯ АНСАМБЛЕВОГО КЛАСИФІКАТОРА ДЛЯ ОПТИМІЗАЦІЇ СИСТЕМИ ВИЯВЛЕННЯ ВТОРГНЕНЬ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

У цій статті розглянуто, як ансамблеві класифікатори можуть значно покращити роботу систем виявлення вторгнень (IDS), особливо в умовах критичної інфраструктури. Через постійне ускладнення кібератак, традиційні методи захисту часто не справляються, що ставить під загрозу безперебійну роботу важливих об'єктів. Запропоновано підхід, що поєднує результати кількох окремих класифікаторів для підвищення точності виявлення загроз і зменшення кількості хибних спрацьовувань. Досліджуються різні методи створення таких

ансамблів та їхнє застосування для аналізу мережевого трафіку, характерного для критичної інфраструктури. Ефективність цих моделей оцінюється на наборах даних, що дозволяє продемонструвати їхню здатність точно ідентифікувати аномалії та відомі типи атак, зміцнюючи таким чином кіберстійкість. Сучасні системи виявлення втргнень (СВВ) повинні бути адаптовані для моніторингу нових фреймворків, які допомагають розрізняти та аналізувати системні атаки. У рамках нашого підходу ми досліджуємо різні комбінації таких класифікаторів, як: Bayesian Network, Naïve Bayes, JRip, MLP, IBK PART, J48. До того ж, для кожної комбінації будуть застосовані два методи попередньої обробки даних – нормалізація та дискретизація. Основна перевага цього підходу – здатність виявляти більшість атак з високою точністю, оптимально поєднуючи ансамблевий метод із правильною технікою попередньої обробки. Це дозволить ефективно ідентифікувати будь-який тип мережевої загрози. Експериментальні дослідження показують, що такий підхід значно підвищує ефективність і надійність виявлення вторгнень, тим самим посилюючи кібербезпеку об'єктів критичної інфраструктури.

Ключові слова: Баєсівська мережа, система виявлення вторгнень, кібератака, критична інфраструктура, класифікатор, кіберстійкість, IBK, JRip, J48, MLP, Наївний Баєс, PART.

Kulko A., Toliupa S.

METHODOLOGY FOR USING AN ENSEMBLE CLASSIFIER TO OPTIMIZE THE CRITICAL INFRASTRUCTURE INTRUSION DETECTION SYSTEM

In this article, we will look at how ensemble classifiers can significantly improve the performance of intrusion detection systems (IDS), especially in critical infrastructure environments. Due to the increasing complexity of cyberattacks, traditional protection methods are often ineffective, which jeopardizes the smooth operation of important facilities. We propose an approach that combines the results of several separate classifiers to improve threat detection accuracy and reduce the number of false positives. Various methods for creating such ensembles and their application for analyzing network traffic characteristic of critical infrastructure are being investigated. The effectiveness of these models is evaluated on data sets, demonstrating their ability to accurately identify anomalies and known types of attacks, thereby strengthening cyber resilience. Modern intrusion detection systems (IDS) must be adapted to monitor new frameworks that help distinguish and analyze system attacks.

As part of our approach, we explore various combinations of classifiers such as Bayesian Network, Naïve Bayes, JRip, MLP, IBK PART, and J48. In addition, two data preprocessing methods — normalization and discretization — will be applied to each combination. The main advantage of this approach is its ability to detect most attacks with high accuracy by optimally combining the ensemble method with the correct preprocessing technique. This will allow for the effective identification of any type of network threat. Experimental studies show that this approach significantly improves the efficiency and reliability of intrusion detection, thereby strengthening the cybersecurity of critical infrastructure facilities.

Keywords: Bayesian network, intrusion detection system, cyberattack, critical infrastructure, classifier, cyber resilience, IBK, JRip, J48, MLP, Naive Bayes, PART.

Стаття надійшла до редакції 20.08.2025 р.
Прийнято до друку 10.09.2025 р