ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «**ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ В МЕРЕЖАХ І ПІДВИЩЕННЯ ЇХНЬОЇ БЕЗПЕКИ»**

на здобуття освітнього ступеня магістра
зі спеціальності 122 Комп'ютерні науки
(код, найменування спеціальності)
освітньо-професійної програми Комп'ютерні науки
(назва)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Денис ДАНИЛЬЧУК
(підпис) *(Ім'я, ПРІЗВИЩЕ здобувача)*

Виконав:
здобувач вищої освіти
група КНДМ-63 Денис ДАНИЛЬЧУК

Керівник:
науковий ступінь, вчене звання Володимир ВАСИЛЕНКО
к.т.н., доцент

Рецензент:
науковий ступінь, вчене звання
_____
(Ім'я, ПРІЗВИЩЕ)

Київ 2023

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут інформаційних технологій

Кафедра Комп'ютерних наук

Ступінь вищої освіти Магістр

Спеціальність Комп'ютерні науки

Освітньо-професійна програма Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедрою Комп'ютерних наук

_____ Віктор ВИШНІВСЬКИЙ
«_____» _____ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Данильчуку Денису Миколайовичу
(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Використання штучного інтелекту для виявлення аномалій в мережах і підвищення їхньої безпеки

керівник кваліфікаційної роботи Володимир ВАСИЛЕНКО к.т.н., доцент,
(Ім'я, ПРІЗВИЩЕ науковий ступінь, вчене звання)
затверджені наказом Державного університету інформаційно-комунікаційних технологій від «19» 10.2023р. №145

2. Строк подання кваліфікаційної роботи «29» грудня 2023р.

3. Вихідні дані до кваліфікаційної роботи: науково-технічна література, налаштування комп'ютерних мереж, властивості штучного інтелекту.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

Дослідження принципів побудови комп'ютерних мереж
Аналіз технологій машинного навчання та можливості застосування в мережах для пошуку аномалій
Розробка вимог до системи виявлення аномалій та навчання штучного інтелекту

5. Перелік графічного матеріалу: *презентація*

1. Розвиток комп’ютерних мереж
2. Розвиток штучного інтелекту
3. Архітектура системи пошуку аномалій в мережі та штучного інтелекту

6. Дата видачі завдання«19» жовтня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз наявної науково-технічної літератури	19.10-05.11.23	
2	Вивчення матеріалів про роботу штучного інтелекту	05.11-12.11.23	
3	Дослідження методів навчання штучного інтелекту	13.11-19.11.23	
4	Дослідження технологій машинного навчання	20.11-25.11.23	
5	Аналіз особливостей машинного навчання	27.11-03.12.23	
6	Застосування машинного навчання в мережах для виявлення аномалій	04.12-10.12.23	
7	Оформлення роботи: вступ, висновки, реферат	11.12-20.12.23	
8	Розробка демонстраційних матеріалів	21.12-29.12.23	

Здобувач вищої освіти _____ Денис ДАНИЛЬЧУК
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Керівник
кваліфікаційної роботи _____ Володимир ВАСИЛЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 65 стор., 14 джерел.

Мета і завдання дослідження: Метою дослідження є дослідження використання ШІ для виявлення аномалій в інформаційних мережах. Для досягнення цієї мети необхідно вирішити наступні завдання:

- Провести аналіз методів виявлення аномалій в інформаційних мережах за допомогою ШІ.
- Провести експериментальні дослідження ефективності методів виявлення аномалій в інформаційних мережах за допомогою ШІ.

Методи дослідження: Для досягнення поставлених завдань будуть використовуватися наступні методи дослідження:

- Аналіз наукової літератури та нормативно-правових документів.
- Експериментальні дослідження.

Об'єкт і предмет дослідження: Об'єктом дослідження є інформаційні мережі. Предметом дослідження є методи виявлення аномалій в інформаційних мережах за допомогою ШІ.

Актуальність дослідження: Актуальність дослідження обумовлена посиленням загроз інформаційній безпеці. Аномалія може бути першим сигналом про те, що система зазнала атаки або стала жертвою помилки програмного забезпечення. Виявлення аномалій дозволяє своєчасно усунути загрозу і запобігти її поширенню.

ШІ дозволяє виявляти аномалії, які неможливо виявити за допомогою традиційних методів. У зв'язку з цим використання ШІ для виявлення аномалій є перспективним напрямком досліджень в області інформаційної безпеки.

Структура дипломного дослідження: Дипломне дослідження складається з вступу, трьох розділів, висновків, списку використаної літератури та додатків.

ABSTRACT

Text part of the qualification work for obtaining a Master's degree: 91 pages, 14 sources.

Purpose and objectives of the research: The purpose of the research is to investigate the use of AI for detecting anomalies in information networks. To achieve this goal, the following tasks need to be solved:

Conduct an analysis of methods for detecting anomalies in information networks using AI.

Conduct experimental studies on the effectiveness of methods for detecting anomalies in information networks using AI.

Research methods: To achieve the set goals, the following research methods will be used:

Analysis of scientific literature and normative-legal documents.

Experimental research.

Object and subject of research: The object of the research is information networks. The subject of the research is methods for detecting anomalies in information networks using AI.

Relevance of the research: The relevance of the research is due to the increasing threats to information security. An anomaly can be the first signal that the system has been attacked or has become a victim of a software error. Detecting anomalies allows timely elimination of the threat and prevention of its spread.

AI allows detecting anomalies that are impossible to detect using traditional methods. In this regard, the use of AI for detecting anomalies is a promising direction of research in the field of information security.

Structure of the diploma research: The diploma research consists of an introduction, three chapters, conclusions, a list of used literature, and appendices.

ЗМІСТ

ВСТУП ... 11

1 ТЕОРЕТИЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ ... 12

1.1 Визначення та класифікація аномалій в мережах ... 12

1.1.1 Внутрішні аномалії ... 13

1.1.2 Зовнішні аномалії ... 14

1.1.3 Активні аномалії ... 14

1.1.4 Пасивні аномалії ... 15

1.1.5 Виявлення активних і пасивних аномалій ... 16

1.1.6 Відомі аномалії ... 16

1.1.7 Невідомі аномалії ... 17

1.1.8 Методи виявлення аномалій ... 18

1.2 Штучний інтелект та машинне навчання в контексті виявлення аномалій ... 20

1.2.1 Види алгоритмів машинного навчання для виявлення аномалій ... 21

1.2.2 Глибоке навчання та нейронні мережі для аналізу мережевого трафіку ... 24

1.2.3 Перспективи ШІ та МЛ для виявлення аномалій у мережі ... 25

1.3 Відкриті дані і програмні засоби для виявлення аномалій ... 26

1.3.1 Відомі програми для аналізу мережевого трафіку ... 27

1.3.2 Вільно доступні набори даних для досліджень у галузі виявлення аномалій ... 29

ВИСНОВОК ... 31

2 ПРАКТИЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ ... 32

2.1 Збір та обробка мережевих даних ... 32

2.1.1 Засоби збору мережевого трафіку ... 34

2.1.2 Попередня обробка даних перед аналізом ... 36

2.2 Розробка та навчання моделей для виявлення аномалій ... 37

2.2.1 Вибір алгоритмів машинного навчання ... 40

2.2.2 Підготовка навчального набору даних ... 42

2.2.3 Тестування та підтримка моделі після навчання 44
2.3 Практичні випробування на реальних мережах 45
2.3.1 Використання розроблених моделей для виявлення аномалій в реальних мережах 46
2.3.2 Аналіз результатів та вдосконалення системи 47
ВИСНОВОК 48
З ЗАСТОСУВАННЯ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ В ПРАКТИЦІ 51
3.1 Розробка системи для виявлення мережевих аномалій. 53
3.2 Навчання штучного інтелекту 54
ВИСНОВОК 56
ВИСНОВКИ 58
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ 60
ДОДАТОК А 62
ДОДАТОК Б 64

СПИСОК СКОРОЧЕНЬ

SVM – Опорновекторні машини.

NN – Нейронні мережі.

ШІ – Штучний інтелект.

МЛ – машинне навчання.

IPS – Cisco Intrusion Prevention System.

TD – Cisco Traffic Director.

IDS – Системи виявлення вторгнень.

TMS – Системи управління трафіком.

C&C – Command and Control.

ВСТУП

У сучасному світі інформаційні мережі є невід'ємною частиною нашого життя. Вони використовуються для передачі даних, зберігання інформації, здійснення комунікацій та ін. Зростання використання інформаційних мереж призвело до зростання кількості загроз їхній безпеці.

Однією з найпоширеніших загроз інформаційним мережам є аномалії. Виявлення аномалій є важливим завданням для забезпечення безпеки інформаційних мереж. Аномалія може бути першим сигналом про те, що система зазнала атаки або стала жертвою помилки програмного забезпечення. Виявлення аномалій дозволяє своєчасно усунути загрозу і запобігти її поширенню.

Для виявлення аномалій в інформаційних мережах використовуються різні методи. Одним з перспективних методів є використання штучного інтелекту (ШІ). ШІ дозволяє виявляти аномалії, які неможливо виявити за допомогою традиційних методів.

Метою даного дипломного дослідження є дослідження використання ШІ для виявлення аномалій в інформаційних мережах. У роботі будуть розглянуті основні методи виявлення аномалій за допомогою ШІ, а також будуть представлені результати експериментальних досліджень ефективності цих методів.

1 ТЕОРЕТИЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ

1.1 Визначення та класифікація аномалій в мережах

Аномалія в мережі - це подія, яка не відповідає очікуваній поведінці мережі. Аномалії можуть бути викликані різними факторами, такими як:

- Зловмисні дії: Атаки хакерів, шкідливе програмне забезпечення, витоки даних тощо[5];
- Фізичні проблеми: Пошкодження обладнання, перевантаження мережі, збої в електропостачанні тощо;
- Помилка програмного забезпечення: Нестабільність програмного забезпечення, помилки в коді тощо;
- Людський фактор: Помилка оператора, помилка користувача тощо.

Аномалії в мережах можуть мати серйозні наслідки, такі як:

- Втрата даних: Атаки хакерів можуть призвести до витоку конфіденційної інформації або даних користувачів;
- Падіння мережі: Фізичні проблеми або збої в програмному забезпеченні можуть призвести до повного або часткового відключення мережі;
- Пошкодження обладнання: Зловмисні дії або помилки в програмному забезпеченні можуть призвести до пошкодження обладнання мережі.

Виявлення аномалій в мережах є важливою задачею, яка дозволяє захистити мережу від атак, забезпечити її надійність і безперебійність роботи.

Існує багато різних методів виявлення аномалій в мережах. Загалом, ці методи можна розділити на дві категорії:

- Статистичні методи: Ці методи використовують статистичні характеристики мережі для визначення аномалій. Наприклад, можна аналізувати розподіл трафіку в мережі, щоб визначити аномалії у вигляді раптових змін у трафіку;

- Методи машинного навчання: Ці методи використовують алгоритми машинного навчання для навчання моделі поведінки мережі. Потім ця модель використовується для визначення аномалій, які не відповідають моделі.

Вибір методу виявлення аномалій залежить від конкретних потреб мережі. Наприклад, для виявлення зловмисних дій в мережі можуть бути ефективними статистичні методи, які аналізують трафік мережі. Для виявлення фізичних проблем в мережі можуть бути ефективними методи машинного навчання, які навчаються на даних про стан мережі.

Виявлення аномалій в мережах є складною задачею, оскільки аномалії можуть бути різноманітними і маскуватися під звичайну поведінку мережі. Тому для ефективного виявлення аномалій необхідно використовувати комплексний підхід, який включає в себе кілька різних методів.

1.1.1 Внутрішні аномалії

Внутрішні аномалії - це аномалії, які виникають всередині мережі. Вони можуть бути викликані такими факторами, як:

- Зловмисні дії;
- Фізичні проблеми;
- Помилка програмного забезпечення.

Внутрішні аномалії можуть бути виявлені шляхом моніторингу мережі на наявність таких ознак, як:

- Незвичайний трафік: Раптові зміни в кількості трафіку, незвичайні напрямки трафіку, незвичайні типи трафіку;
- Помилкові дані: Помилкові записи в журналах, неправильні значення в метричних даних;
- Нестабільність роботи: Зниження продуктивності, збої в роботі мережі.

1.1.2 Зовнішні аномалії

Зовнішні аномалії - це аномалії, які виникають поза мережею. Вони можуть бути викликані такими факторами, як:

- Зловмисні дії: Атаки ззовні мережі, наприклад, DDoS-атаки.
- Фізичні проблеми: Пошкодження інфраструктури, яка обслуговує мережу, наприклад, лінії зв'язку, сервери тощо.
- Помилка програмного забезпечення: Помилки в програмному забезпеченні, яке обслуговує мережу, наприклад, в операційній системі, веб-сервері тощо.

Зовнішні аномалії можуть бути виявлені шляхом моніторингу мережі на наявність таких ознак, як:

- Зниження продуктивності: Зниження продуктивності мережі, викликане підвищеним навантаженням ззовні.
- Збої в роботі: Збої в роботі мережі, викликані фізичними проблемами або помилками програмного забезпечення.
- Незвичайні дії: Незвичайні дії в мережі, наприклад, спроби доступу до заборонених ресурсів, спроби входу в систему з неправильних IP-адрес тощо.

1.1.3 Активні аномалії

Активні аномалії - це аномалії, які виникають в результаті активних дій користувачів або пристроїв. Вони можуть бути викликані такими факторами, як:

- Зловмисні дії: Атаки хакерів, шкідливе програмне забезпечення, витоки даних тощо.
- Людський фактор: Помилка оператора, помилка користувача тощо.

Активні аномалії можуть бути виявлені шляхом моніторингу мережі на наявність таких ознак, як:

- Незвичайний трафік: Раптові зміни в кількості трафіку, незвичайні напрямки трафіку, незвичайні типи трафіку.

- Незвичайні дії: Незвичайні дії в мережі, наприклад, спроби доступу до заборонених ресурсів, спроби входу в систему з неправильних IP-адрес тощо.

Приклади активних аномалій:

- Зловмисна атака: Хакер може спробувати отримати доступ до конфіденційної інформації, використовуючи незвичайні методи, такі як використання невідомого протоколу або спроба отримати доступ до забороненого ресурсу.
- Шкідливе програмне забезпечення: Шкідливе програмне забезпечення може спробувати завантажити файли або відправити повідомлення з комп'ютера користувача, використовуючи незвичайні протоколи або напрямки.
- Людська помилка: Оператор може випадково видалити важливі дані або змінити конфігурацію мережі, що призведе до непередбачених наслідків.

1.1.4 Пасивні аномалії

Пасивні аномалії - це аномалії, які виникають в результаті природних процесів або помилок в роботі обладнання. Вони можуть бути викликані такими факторами, як:

- Фізичні проблеми: Пошкодження обладнання, перевантаження мережі, збої в електропостачанні тощо.
- Помилка програмного забезпечення: Нестабільність програмного забезпечення, помилки в коді тощо.

Пасивні аномалії можуть бути виявлені шляхом моніторингу мережі на наявність таких ознак, як:

- Зниження продуктивності: Зниження продуктивності мережі, викликане фізичними проблемами або помилками програмного забезпечення.
- Збої в роботі: Збої в роботі мережі, викликані фізичними проблемами або помилками програмного забезпечення.
- Незвичайні значення: Неправильні значення в метричних даних, які можуть вказувати на проблеми з обладнанням або програмним забезпеченням.

Приклади пасивних аномалій:

- Пошкодження обладнання: Згоряння кабелю або поломка сервера можуть призвести до зниження продуктивності або збоїв в роботі мережі.
- Перевантаження мережі: Підвищене навантаження на мережу, наприклад, в результаті проведення онлайн-заходу, може призвести до зниження продуктивності або збоїв в роботі мережі.
- Помилка програмного забезпечення: Нестабільна операційна система або помилки в програмному забезпеченні можуть призвести до збоїв в роботі мережі.

1.1.5 Виявлення активних і пасивних аномалій

Активні аномалії часто легше виявити, ніж пасивні аномалії. Це пов'язано з тим, що активні аномалії, як правило, викликають раптові і значні зміни в поведінці мережі. Пасивні аномалії, з іншого боку, можуть бути менш помітними і вимагають більш детального аналізу даних.

Для виявлення активних аномалій часто використовуються статистичні методи, які аналізують розподіл трафіку в мережі. Наприклад, можна використовувати метод середнього квадратичного відхилення для виявлення раптових змін у трафіку.

Для виявлення пасивних аномалій часто використовуються методи машинного навчання, які навчаються на даних про стан мережі. Потім ця модель використовується для визначення аномалій, які не відповідають моделі.

Для ефективного виявлення аномалій в мережі необхідно використовувати комплексний підхід, який включає в себе кілька різних методів.

1.1.6 Відомі аномалії

Відомі аномалії - це аномалії, які вже відомі і описані. Вони можуть бути викликані такими факторами, як:

- Зловмисні дії: Відомі типи зловмисних атак, наприклад, DDoS-атаки, фішингові атаки, атаки на веб-сайти тощо.[4]
- Фізичні проблеми: Відомі типи фізичних проблем, наприклад, перевантаження мережі, збої в електропостачанні тощо.
- Помилка програмного забезпечення: Відомі типи помилок програмного забезпечення, наприклад, нестабільність операційної системи, помилки в коді тощо.

Відомі аномалії можна виявити, використовуючи статистичні методи або методи машинного навчання. Для статистичних методів можна використовувати відомі статистичні моделі для виявлення аномалій, які не відповідають моделі. Для методів машинного навчання можна використовувати відомі набори даних аномалій для навчання моделі виявлення аномалій.

Приклади відомих аномалій:

- DDoS-атака: DDoS-атака - це атака, яка спрямована на перевантаження мережі або сервера. DDoS-атаки можуть бути виявлені шляхом моніторингу мережі на наявність раптового і значного підвищення навантаження.
- Фішингова атака: Фішингова атака - це атака, яка спрямована на виманювання конфіденційної інформації, наприклад, паролів або номерів кредитних карток. Фішингові атаки можуть бути виявлені шляхом моніторингу мережі на наявність незвичайних електронних листів або веб-сайтів, які можуть містити шкідливе програмне забезпечення.
- Атака на веб-сайт: Атака на веб-сайт - це атака, яка спрямована на пошкодження або виведення з ладу веб-сайту. Атаки на веб-сайти можуть бути виявлені шляхом моніторингу веб-сайту на наявність незвичайної поведінки, наприклад, збільшення кількості помилок або зниження продуктивності.

1.1.7 Невідомі аномалії

Невідомі аномалії - це аномалії, які не відомі і не описані. Вони можуть бути викликані такими факторами, як:

- Нові зловмисні атаки: Зловмисники постійно розробляють нові типи атак, які можуть бути невідомі для систем виявлення аномалій.
- Нові фізичні проблеми: Можуть виникати нові типи фізичних проблем, які не були враховані при розробці систем виявлення аномалій.
- Нові помилки програмного забезпечення: Можуть виникати нові помилки програмного забезпечення, які не були враховані при розробці систем виявлення аномалій.

Невідомі аномалії можуть бути дуже складними для виявлення. Це пов'язано з тим, що системи виявлення аномалій, як правило, навчаються на відомих наборах даних аномалій. Коли виникає невідома аномалія, вона може не відповідати моделі, на якій навчена система виявлення аномалій.

Методи виявлення аномалій поділяються на дві категорії: виявлення на основі сигнатур та виявлення на основі статистики. Сигнатурний метод виявляє аномалію по відомих ознаках –«сигнатурах». Недоліком цього методу є властивість виявляти відомі заздалегідь типи аномалій. Тому даний метод не може бути застосований для ідентифікації невідомих аномалій.[1]

Для виявлення невідомих аномалій можна використовувати такі методи, як:

- Методи машинного навчання, які здатні навчатися на невідомому наборі даних.
- Методи, які використовують експертні знання для виявлення аномалій, які не відповідають очікуваній поведінці мережі.

Використання комплексного підходу, який включає в себе кілька різних методів, може підвищити ефективність виявлення невідомих аномалій.

1.1.8 Методи виявлення аномалій

Існує багато різних методів виявлення аномалій в мережах. Загалом, ці методи можна розділити на дві категорії:

- Статистичні методи: Ці методи використовують статистичні характеристики мережі для визначення аномалій. Наприклад, можна аналізувати

розподіл трафіку в мережі, щоб визначити аномалії у вигляді раптових змін у трафіку.

- Методи машинного навчання: Ці методи використовують алгоритми машинного навчання для навчання моделі поведінки мережі. Потім ця модель використовується для визначення аномалій, які не відповідають моделі.

Одним з методів, що використовують малохвильове перетворення для виявлення аномалій трафіку є метод з кореляцією адрес для певної кількості точок відбору. Сигнал може бути визначений на відповідному часовому масштабі та на певних позиціях часового масштабу, крім того існує можливість відображення частотних та часових компонент одночасно.[1]

Вибір методу виявлення аномалій залежить від конкретних потреб мережі. Наприклад, для виявлення зловмисних дій в мережі можуть бути ефективними статистичні методи, які аналізують трафік мережі. Для виявлення фізичних проблем в мережі можуть бути ефективними методи машинного навчання, які навчаються на даних про стан мережі.

Виявлення аномалій в мережах є складною задачею, оскільки аномалії можуть бути різноманітними і маскуватися під звичайну поведінку мережі. Тому для ефективного виявлення аномалій необхідно використовувати комплексний підхід, який включає в себе кілька різних методів.

1.2 Штучний інтелект та машинне навчання в контексті виявлення аномалій

Штучний інтелект (ШІ) та машинне навчання (МЛ) є потужними інструментами, які можуть бути використані для виявлення аномалій у мережі. Аномалії можуть бути ознакою потенційних проблем, таких як атаки, збої обладнання або помилки програмного забезпечення. ШІ та МЛ можуть допомогти виявити ці аномалії, навіть якщо вони не є очевидними для людини.

Існує кілька різних підходів до використання ШІ та МЛ для виявлення аномалій у мережі. Один підхід полягає в тому, щоб навчити модель на наборі даних, який містить як нормальні, так і аномальні дані. Модель потім може використовуватися для класифікації нових даних як нормальних або аномальних.

Інший підхід полягає в тому, щоб використовувати ШІ для виявлення аномалій у поведінці мережі. Наприклад, ШІ може використовуватися для виявлення аномалій у трафіку мережі, таких як раптові зміни в інтенсивності трафіку або незвичні шаблони трафіку.

ШІ та МЛ можуть бути ефективними інструментами для виявлення аномалій у мережі. Однак важливо вибрати правильний підхід для конкретної мережі.

Ось деякі конкретні приклади того, як ШІ та МЛ можуть використовуватися для виявлення аномалій у мережі:

- Виявлення атак: ШІ та МЛ можуть використовуватися для виявлення атак на мережу, таких як DoS-атаки, атаки типу "людина посередині" та атаки типу "відвідувач-хакер". Наприклад, ШІ можна використовувати для виявлення аномалій у трафіку мережі, які можуть бути ознакою атаки.
- Виявлення збоїв обладнання: ШІ та МЛ можуть використовуватися для виявлення збоїв обладнання, таких як несправності мережевих пристроїв або несправності ліній зв'язку. Наприклад, ШІ можна використовувати для виявлення аномалій у даних, що надходять від мережевих пристроїв, які можуть бути ознакою несправності.

- Виявлення помилок програмного забезпечення: ШІ та МЛ можуть використовуватися для виявлення помилок програмного забезпечення, які можуть призвести до проблем у мережі. Наприклад, ШІ можна використовувати для виявлення аномалій у поведінці мережі, які можуть бути ознакою помилки програмного забезпечення.

ШІ та МЛ можуть бути ефективними інструментами для виявлення аномалій у мережі, але важливо вибрати правильний підхід для конкретної мережі. Ось деякі фактори, які слід враховувати при виборі підходу:

- Тип мережі: Різні типи мереж мають різні характеристики та потреби. Наприклад, мережі з великою кількістю даних можуть вимагати іншого підходу, ніж мережі з невеликою кількістю даних.
- Тип аномалій: Різні типи аномалій вимагають різних підходів. Наприклад, виявлення атак вимагає іншого підходу, ніж виявлення збоїв обладнання.
- Вимоги до продуктивності: Деякі підходи вимагають більшої продуктивності, ніж інші. Наприклад, підходи, які використовують навчання на великих даних, можуть бути менш продуктивними, ніж підходи, які використовують навчання на невеликих наборах даних.

ШІ та МЛ є швидко розвиваючі області, і в майбутньому вони, ймовірно, будуть відігравати все більш важливу роль у виявленні аномалій у мережі[7].

1.2.1 Види алгоритмів машинного навчання для виявлення аномалій

Існує два основних типи алгоритмів машинного навчання для виявлення аномалій:

- Алгоритми контрольованого навчання: Ці алгоритми вимагають набору даних, який містить як нормальні, так і аномальні дані. Алгоритм використовується для навчання на цьому наборі даних, а потім використовується для класифікації нових даних як нормальних або аномальних.

- Алгоритми неконтрольованого навчання: Ці алгоритми не вимагають набору даних, який містить аномальні дані. Алгоритм використовується для виявлення аномалій, використовуючи статистичні методи або моделі машинного навчання.

Алгоритми контрольованого навчання

Алгоритми контрольованого навчання є найбільш поширеним типом алгоритмів для виявлення аномалій. Вони є ефективними для виявлення аномалій, які мають чіткі відмінності від нормальних даних.[9]

Деякі з найпопулярніших алгоритмів контрольованого навчання для виявлення аномалій включають:

- Дерево рішень: Дерево рішень - це алгоритм, який використовує правила для класифікації даних. Дерево рішень можна використовувати для навчання на наборі даних, який містить як нормальні, так і аномальні дані. Після навчання дерево рішень може використовуватися для класифікації нових даних як нормальних або аномальних.
- Опорновекторні машини: Опорновекторні машини (SVM) - це алгоритм, який використовує гіперплощину для класифікації даних. SVM можна використовувати для навчання на наборі даних, який містить як нормальні, так і аномальні дані. Після навчання SVM може використовуватися для класифікації нових даних як нормальних або аномальних.
- Регресійний аналіз: Регресійний аналіз - це метод, який використовується для прогнозування значення змінної на основі інших змінних. Регресійний аналіз можна використовувати для навчання на наборі даних, який містить як нормальні, так і аномальні дані. Після навчання регресійний аналіз може використовуватися для прогнозування, чи є нові дані аномальними.

Алгоритми неконтрольованого навчання

Алгоритми неконтрольованого навчання є більш складними, ніж алгоритми контрольованого навчання. Вони можуть бути ефективними для виявлення аномалій, які не мають чітких відмінностей від нормальних даних.[8]

Деякі з найпопулярніших алгоритмів неконтрольованого навчання для виявлення аномалій включають:

- Метод k-середніх: Метод k-середніх - це алгоритм, який поділяє дані на k кластерів. Дані, які не належать до жодного кластера, вважаються аномальними.
- Ієрархічна кластеризація: Ієрархічна кластеризація - це алгоритм, який використовує ієрархію для поділу даних на кластери. Дані, які не належать до жодного кластера, вважаються аномальними.
- Метод ексцесів: Метод ексцесів використовує статистичні методи для виявлення аномалій. Метод ексцесів шукає дані, які мають значні відхилення від нормального розподілу.
- Метод лінійної дисперсії: Метод лінійної дисперсії використовує статистичні методи для виявлення аномалій. Метод лінійної дисперсії шукає дані, які мають значні відхилення від лінійної регресії.

Вибір алгоритму

Вибір алгоритму машинного навчання для виявлення аномалій залежить від кількох факторів, включаючи:

- Тип аномалій, які потрібно виявити: Деякі алгоритми краще підходять для виявлення певних типів аномалій, ніж інші.
- Розмір набору даних: Деякі алгоритми вимагають великих наборів даних для навчання, ніж інші.
- Вимоги до продуктивності: Деякі алгоритми є більш продуктивними, ніж інші.

А що якщо сприйняти завдання знаходження аномалій як нове завдання машинного навчання (відмінне від класифікації та кластеризації)?!

- Найпопулярніші алгоритми (є реалізація навіть у scikit-learn) тут:
- Метод опорних векторів для одного класу (OneClassSVM)
- Ізолюючий ліс (IsolationForest)
- Еліпсоїдна апроксимація даних (EllipticEnvelope) [6]

Важливо протестувати різні алгоритми на конкретному наборі даних, щоб визначити, який алгоритм забезпечує найкращу продуктивність для даної задачі.

1.2.2 Глибоке навчання та нейронні мережі для аналізу мережевого трафіку

Глибоке навчання та нейронні мережі (NN) є потужними інструментами, які можна використовувати для аналізу мережевого трафіку. NN можуть навчитися виявляти аномалії в мережевому трафіку, навіть якщо ці аномалії не є очевидними для людини.

Глибоке навчання — це тип машинного навчання, який використовує багатошарові нейронні мережі для навчання на великих обсягах даних. NN можуть навчитися виявляти складні закономірності в даних, які неможливо виявити за допомогою традиційних методів.

NN для аналізу мережевого трафіку можна використовувати для виявлення різних типів аномалій, включаючи:

- Атаки: NN можуть навчитися виявляти ознаки атак на мережу, таких як DoS-атаки, атаки типу "людина посередині" та атаки типу "відвідувач-хакер".
- Збої обладнання: NN можуть навчитися виявляти ознаки збоїв обладнання, таких як несправності мережевих пристроїв або несправності ліній зв'язку.
- Помилки програмного забезпечення: NN можуть навчитися виявляти ознаки помилок програмного забезпечення, які можуть призвести до проблем у мережі.

Переваги використання NN для аналізу мережевого трафіку:

- NN можуть навчитися виявляти складні аномалії, які неможливо виявити за допомогою традиційних методів.
- NN можуть бути ефективними для виявлення нових типів аномалій, які не були відомі раніше.
- NN можуть бути більш точними та чутливими, ніж традиційні методи.

Недоліки використання NN для аналізу мережевого трафіку:

- NN вимагають великих наборів даних для навчання.
- NN можуть бути складними для розробки та налаштування.

1.2.3 Перспективи ШІ та МЛ для виявлення аномалій у мережі

ШІ та МЛ мають потенціал для радикального поліпшення виявлення аномалій у мережі. Ці технології можуть допомогти виявити аномалії, які неможливо виявити за допомогою традиційних методів.

Ось деякі конкретні перспективи ШІ та МЛ для виявлення аномалій у мережі:

- Виявлення нових типів аномалій: ШІ та МЛ можуть допомогти виявити нові типи аномалій, які не були відомі раніше. Це може бути особливо корисно для виявлення нових видів атак або інших загроз.
- Покращення точності та чутливості: ШІ та МЛ можуть допомогти покращити точність та чутливість виявлення аномалій. Це може призвести до того, що аномалії будуть виявлені швидше та з більшою впевненістю.
- Зменшення витрат: ШІ та МЛ можуть допомогти зменшити витрати на виявлення аномалій. Це може бути досягнуто за рахунок автоматизації завдань, які в даний час виконуються вручну, а також за рахунок покращення ефективності виявлення аномалій.

Однак існують також деякі виклики, які необхідно подолати, перш ніж ШІ та МЛ зможуть повністю реалізувати свій потенціал у виявленні аномалій у мережі. Одним з викликів є необхідність великих наборів даних для навчання моделей ШІ та МЛ. Іншим викликом є необхідність розробки моделей, які є достатньо точними та чутливими, щоб виявляти широкий спектр аномалій.

Незважаючи на ці виклики, ШІ та МЛ є перспективними технологіями, які мають потенціал для радикального поліпшення виявлення аномалій у мережі.

1.3 Відкриті дані і програмні засоби для виявлення аномалій

Існує безліч наборів даних відкритого доступу, які можуть бути використані для навчання моделей виявлення аномалій. Ось деякі приклади:

- NIDS Dataset - це набір даних, який містить записи про трафік мережі, класифіковані як нормальні або аномальні.
- UNSW-NB15 Dataset - це набір даних, який містить записи про трафік мережі, класифіковані як нормальні або аномальні. Цей набір даних більший, ніж NIDS Dataset, і він містить записи про більш широкий спектр аномалій.
- KDD Cup 99 Dataset - це набір даних, який містить записи про трафік мережі, класифіковані як нормальні або аномальні. Цей набір даних є одним із найстаріших наборів даних для виявлення аномалій, і він містить записи про широкий спектр аномалій.

Програмні засоби для виявлення аномалій

Існує також ряд програмних засобів, які можна використовувати для виявлення аномалій. Ось деякі приклади:

- Anomaly Detection Toolkit (ADT) - це відкрите програмне забезпечення, яке можна використовувати для виявлення аномалій у різних типах даних.
- Snort - це система інспектування пакетів, яка може використовуватися для виявлення аномалій у мережному трафіку.
- Suricata - це ще одна система інспектування пакетів, яка може використовуватися для виявлення аномалій у мережному трафіку.

Ці програмні засоби можуть використовуватися для виявлення різних типів аномалій, таких як:

- Незвичайний трафік мережі
- Зловмисна активність
- Ошибки обладнання
- Вади безпеки

Вони можуть бути використані для різних цілей, таких як:

- Захист мережі

- Діагностика обладнання
- Виявлення помилок
- Покращення ефективності

Відкриті дані і програмні засоби для виявлення аномалій можуть бути цінним ресурсом для розробників і дослідників. Вони дозволяють швидко і легко розпочати роботу з виявлення аномалій і можуть бути використані для навчання моделей виявлення аномалій, які можуть бути використані для вирішення різних завдань.

1.3.1 Відомі програми для аналізу мережевого трафіку

Аналіз мережевого трафіку є важливим завданням для забезпечення безпеки та продуктивності мережі. Для виконання цього завдання існує широкий спектр програмного забезпечення, яке можна використовувати.

Відомі програми для аналізу мережевого трафіку можна розділити на кілька категорій:

- Сніфери: Сніфери - це програми, які перехоплюють та записують мережевий трафік. Вони можуть використовуватися для виявлення аномалій у трафіку, а також для аналізу трафіку для інших цілей, таких як моніторинг продуктивності або збирання даних.
- Системи виявлення вторгнень (IDS): IDS - це системи, які використовують аналіз мережевого трафіку для виявлення атак. Вони можуть використовуватися для виявлення різних типів атак, включаючи DoS-атаки, атаки типу "людина посередині" та атаки типу "відвідувач-хакер".
- Системи запобігання вторгненням (IPS): IPS - це системи, які використовують аналіз мережевого трафіку для запобігання атакам. Вони можуть використовуватися для блокування атак або для попередження систем безпеки про потенційні атаки.
- Системи управління трафіком (TMS): TMS - це системи, які використовуються для управління мережевим трафіком. Вони можуть

використовуватися для моніторингу трафіку, для визначення проблем із пропускною здатністю та для оптимізації використання мережі.

Ось деякі з найвідоміших програм для аналізу мережевого трафіку:

- Wireshark: Wireshark - це відкритий вихідний код сніфер, який є одним з найпопулярніших інструментів для аналізу мережевого трафіку. Він може використовуватися для перехоплення та запису мережевого трафіку, а також для аналізу трафіку за допомогою широкого спектру фільтрів та інструментів.
- Nmap: Nmap - це безкоштовний і відкритий вихідний код сканер портів, який можна використовувати для виявлення відкритих портів на серверах та інших пристроях. Він також може використовуватися для виявлення інших проблем із мережею, таких як небезпеки безпеки та проблеми з конфігурацією.
- Snort: Snort - це відкритий вихідний код IDS, який є одним з найпопулярніших інструментів для виявлення атак. Він може використовуватися для виявлення різних типів атак, включаючи DoS-атаки, атаки типу "людина посередині" та атаки типу "відвідувач-хакер".
- Suricata: Suricata - це відкритий вихідний код IDS, який є одним із найновіших і найпотужніших інструментів для виявлення атак. Він може використовуватися для виявлення різних типів атак, включаючи DoS-атаки, атаки типу "людина посередині" та атаки типу "відвідувач-хакер".
- Palo Alto Networks WildFire: WildFire - це платна IDS, яка використовує аналіз поведінки для виявлення атак. Він може використовуватися для виявлення навіть найновіших і найскладніших атак.
- Cisco Intrusion Prevention System (IPS): Cisco IPS - це платна IPS, яка використовує аналіз поведінки для запобігання атакам. Він може використовуватися для блокування атак або для попередження систем безпеки про потенційні атаки.
- Cisco Traffic Director (TD): Cisco TD - це платна TMS, яка використовується для управління мережевим трафіком. Він може використовуватися для моніторингу трафіку, для визначення проблем із пропускною здатністю та для оптимізації використання мережі.

При виборі програми для аналізу мережевого трафіку слід враховувати такі фактори:

- Функціональні можливості: Які функції необхідні для вашого конкретного завдання?
- Ціна: Скільки ви готові витратити на програмне забезпечення?
- Простота використання: Наскільки легко використовувати програмне забезпечення?
- Підтримка: Чи доступна підтримка програмного забезпечення?

Підводячи підсумок, можна сказати, що на ринку існує широкий спектр програмного забезпечення для аналізу мережевого трафіку. При виборі програми слід враховувати свої конкретні потреби та вимоги.

1.3.2 Вільно доступні набори даних для досліджень у галузі виявлення аномалій

Виявлення аномалій є важливим завданням у багатьох галузях, включаючи безпеку, медицину та промисловість. Для дослідження та розробки методів виявлення аномалій необхідні набори даних, які містять як нормальні, так і аномальні дані.

Відкриті інструменти та бібліотеки для ШІ є цінним ресурсом для розробників і дослідників ШІ. Вони пропонують широкий спектр функцій і можливостей, які можуть використовуватися для вирішення широкого спектру завдань.

Однією з ключових переваг відкритих інструментів та бібліотек є їхня доступність. Вони часто доступні безкоштовно або за низькою ціною, що робить їх доступними для широкого кола користувачів.

Іншою перевагою відкритих інструментів та бібліотек є їхня гнучкість. Вони можуть бути використані для вирішення широкого спектру завдань, що робить їх цінним ресурсом для розробників і дослідників ШІ.

Деякі з найпопулярніших відкритих інструментів та бібліотек для ШІ включають:[13]

- TensorFlow — фреймворк для машинного навчання, розроблений компанією Google. TensorFlow використовується для розробки та впровадження моделей машинного навчання для широкого спектру завдань, включаючи розпізнавання образів, розпізнавання мови та природний мовний оброб.[8]
- PyTorch — фреймворк для машинного навчання, розроблений компанією Facebook. PyTorch використовується для розробки та впровадження моделей машинного навчання для широкого спектру завдань, включаючи розпізнавання образів, розпізнавання мови.
- Scikit-learn — бібліотека для машинного навчання, написана на Python. Scikit-learn використовується для розробки та впровадження моделей машинного навчання для широкого спектру завдань, включаючи класифікацію, регресію та кластеризацію.[12]

Ці інструменти та бібліотеки можуть бути використані для вирішення широкого спектру завдань ШІ, включаючи:

- Розпізнавання образів: Відкриті інструменти та бібліотеки для ШІ можуть бути використані для розробки моделей розпізнавання образів, які можуть ідентифікувати об'єкти на зображеннях.
- Розпізнавання мови: Відкриті інструменти та бібліотеки для ШІ можуть бути використані для розробки моделей розпізнавання мови, які можуть розуміти людську мову.
- Глибоке навчання: Відкриті інструменти та бібліотеки для ШІ можуть бути використані для розробки та впровадження моделей глибокого навчання, які можуть вирішувати складні завдання, які неможливо вирішити за допомогою традиційних методів.[11]

При виборі відкритих інструментів та бібліотек для ШІ слід враховувати такі фактори:

- Функціональні можливості: Які функції необхідні для вашого конкретного завдання?

- Відповідність вимогам: Чи відповідають інструменти та бібліотеки вашим технічним вимогам?
- Стабільність: Чи є інструменти та бібліотеки стабільними та добре підтримуються?

ВИСНОВОК

Аномалія в мережі - це подія, яка не відповідає очікуваній поведінці мережі. Існують внутрішні та зовнішні, активні та пасивні, відомі та невідомі аномалії.

Штучний інтелект (ШІ) та машинне навчання (МЛ) є потужними інструментами, які можуть бути використані для виявлення аномалій у мережі

Відкриті інструменти та бібліотеки для ШІ є цінним ресурсом для розробників і дослідників ШІ. Вони пропонують широкий спектр функцій і можливостей, які можуть використовуватися для вирішення широкого спектру завдань. При виборі відкритих інструментів та бібліотек для ШІ слід враховувати такі фактори: функціональні можливості, відповідність вимогам та стабільність.

2 ПРАКТИЧНИЙ АСПЕКТ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ

2.1 Збір та обробка мережевих даних

Збір та обробка мережевих даних є важливими завданнями для забезпечення безпеки та продуктивності мережі. Для виконання цих завдань існує широкий спектр інструментів та методів, які можна використовувати.

Збір мережевих даних - це процес отримання даних з мережі. Для цього можна використовувати різні методи, включаючи:

- Сніфери: Сніфери - це програми, які перехоплюють та записують мережевий трафік. Вони можуть використовуватися для збору даних про весь мережевий трафік або про певний тип трафіку.
- Системи виявлення вторгнень (IDS): IDS - це системи, які використовують аналіз мережевого трафіку для виявлення атак. Вони можуть генерувати дані про виявлені атаки, які можна використовувати для аналізу та виявлення тенденцій.
- Системи управління трафіком (TMS): TMS - це системи, які використовуються для управління мережевим трафіком. Вони можуть генерувати дані про використання мережі, які можна використовувати для аналізу та оптимізації продуктивності мережі.

Обробка мережевих даних - це процес очищення, форматування та аналізу даних, отриманих з мережі. Для цього можна використовувати різні методи, включаючи:

- Очищення даних: Це процес видалення шуму та непотрібних даних з набору даних. Це важливий крок, оскільки він може покращити точність аналізу даних.

- Форматування даних: Це процес приведення даних до стандартного формату, який може бути оброблений певним програмним забезпеченням або алгоритмом.
- Аналіз даних: Це процес виявлення закономірностей та тенденцій у наборі даних. Це може бути використано для виявлення атак, проблем із мережею або інших проблем.

Збір та обробка мережевих даних можуть використовуватися для вирішення широкого спектру завдань, включаючи:

- Забезпечення безпеки: Збір та обробка мережевих даних можуть використовуватися для виявлення атак на мережу. Це може допомогти захистити мережу від шкідливих програм, хакерських атак та інших загроз.
- Оптимізація продуктивності: Збір та обробка мережевих даних можуть використовуватися для аналізу використання мережі. Це може допомогти оптимізувати пропускну здатність мережі та забезпечити її ефективне використання.
- Діагностика проблем: Збір та обробка мережевих даних можуть використовуватися для виявлення проблем із мережею. Це може допомогти усунути проблеми та забезпечити надійність мережі.

Приклади збору та обробки мережевих даних

Ось кілька прикладів того, як можна використовувати збір та обробку мережевих даних:

- Компанія може використовувати сніфер для збору даних про весь мережевий трафік. Ці дані можна використовувати для виявлення атак на мережу, таких як DoS-атаки або атаки типу "людина посередині".
- Система виявлення вторгнень (IDS) може генерувати дані про виявлені атаки. Ці дані можна використовувати для аналізу та виявлення тенденцій у атаках.
- Система управління трафіком (TMS) може генерувати дані про використання мережі. Ці дані можна використовувати для аналізу та оптимізації продуктивності мережі.

2.1.1 Засоби збору мережевого трафіку

Мережевий трафік - це дані, які передаються по мережі. Він може бути використаний для різних цілей, таких як:

- Забезпечення безпеки: Мережевий трафік може бути використаний для виявлення атак на мережу.
- Оптимізація продуктивності: Мережевий трафік може бути використаний для аналізу використання мережі та виявлення проблем із мережею.
- Діагностика проблем: Мережевий трафік може бути використаний для виявлення проблем із мережею.

Для збору мережевого трафіку використовуються спеціальні інструменти, які називаються засобами збору мережевого трафіку.

Основні типи засобів збору мережевого трафіку

Існує два основних типи засобів збору мережевого трафіку:

- Сніфери - це програми, які перехоплюють та записують мережевий трафік. Вони можуть використовуватися для збору даних про весь мережевий трафік або про певний тип трафіку.
- Системи виявлення вторгнень (IDS) - це системи, які використовують аналіз мережевого трафіку для виявлення атак. Вони можуть генерувати дані про виявлені атаки, які можуть бути використані для аналізу та виявлення тенденцій.

Сніфери

Сніфери - це найпоширеніший тип засобів збору мережевого трафіку. Вони можуть бути використані для збору даних про весь мережевий трафік або про певний тип трафіку, наприклад, про трафік, який проходить через певний порт або між певними комп'ютерами.

Сніфери можуть бути встановлені на різних пристроях, таких як комп'ютери, роутери та комутатори.

Системи виявлення вторгнень (IDS)

IDS - це системи, які використовуються для захисту мережі від атак. Вони використовують аналіз мережевого трафіку для виявлення атак.

IDS можуть генерувати дані про виявлені атаки, які можуть бути використані для аналізу та виявлення тенденцій.

Додаткові типи засобів збору мережевого трафіку

Окрім сніферів та IDS, існують також інші типи засобів збору мережевого трафіку, такі як:

- Системи управління мережею (NMS) - це системи, які використовуються для управління мережею. Вони можуть генерувати дані про використання мережі, які можуть бути використані для аналізу та оптимізації продуктивності мережі.
- Системи моніторингу мережі (NMS) - це системи, які використовуються для моніторингу мережі. Вони можуть генерувати дані про стан мережі, які можуть бути використані для виявлення проблем із мережею.

Вибір засобів збору мережевого трафіку

При виборі засобів збору мережевого трафіку необхідно враховувати такі фактори:

- Цілі збору даних - для яких цілей збираються дані?
- Тип даних, які необхідно зібрати - які дані необхідно зібрати?
- Розмір мережі - який розмір мережі необхідно моніторити?
- Бюджет - який бюджет виділений на закупівлю засобів збору мережевого трафіку?

Завдання засобів збору мережевого трафіку

Засоби збору мережевого трафіку можуть виконувати такі завдання:

- Збір даних - засоби збору мережевого трафіку можуть збирати дані про мережевий трафік.
- Фільтрація даних - засоби збору мережевого трафіку можуть фільтрувати дані, щоб залишити тільки необхідні дані.
- Архівація даних - засоби збору мережевого трафіку можуть архівувати дані для подальшого аналізу.
- Аналіз даних - засоби збору мережевого трафіку можуть аналізувати дані для виявлення аномалій або тенденцій.[10]

2.1.2 Попередня обробка даних перед аналізом

Попередня обробка даних - це важливий етап процесу аналізу даних. Вона включає в себе ряд завдань, спрямованих на підготовку даних до аналізу, таких як очищення, форматування та стандартизація.

Очищення даних.

Очищення даних - це процес видалення шуму та непотрібних даних з набору даних. Шум може включати в себе помилки, пропуски даних та неправильні дані. Непотрібні дані - це дані, які не є важливими для аналізу.

Очищення даних може бути виконано за допомогою різних методів, таких як:

- Видалення дублікатів: Видалення дублікатів даних, які повторюються в наборі даних.
- Видалення помилок: Виправлення або видалення помилок у даних.
- Заповнення пропусків: Заповнення пропусків у даних.
- Вилучення непотрібних даних: Видалення даних, які не є важливими для аналізу.

Форматування даних.

Форматування даних - це процес приведення даних до стандартного формату, який може бути оброблений певним програмним забезпеченням або алгоритмом.

Форматування даних може бути виконано за допомогою різних методів, таких як:

- Перетворення даних до одного формату: Перетворення даних з одного формату в інший, наприклад, з текстового формату в числовий формат.
- Нормування даних: Приведення даних до одного масштабу, наприклад, до діапазону від 0 до 1.
- Категоризація даних: Приведення даних до категорій, наприклад, до категорій "нормальні" або "аномальні".

Стандартизація даних.

Стандартизація даних - це процес приведення даних до єдиного стандарту. Це може бути корисно для порівняння даних з різних джерел.

Стандартизація даних може бути виконана за допомогою різних методів, таких як:

- Приведення даних до одного масштабу: Приведення даних до одного масштабу, наприклад, до діапазону від 0 до 1.
- Згладжування даних: Зменшення коливань даних, наприклад, за допомогою середнього зважування.

Вплив попередньої обробки даних на аналіз

Попередня обробка даних може мати значний вплив на результати аналізу. Неякісна попередня обробка даних може привести до спотворення результатів аналізу.

Тому важливо ретельно виконувати попередню обробку даних, використовуючи правильні методи та інструменти.

2.2 Розробка та навчання моделей для виявлення аномалій

Виявлення аномалій є завданням, яке полягає в тому, щоб ідентифікувати дані, які відрізняються від нормального шаблону. Це може бути корисно для різних цілей, таких як виявлення атак, діагностика проблем та виявлення тенденцій.

Існує два основних підходи до розробки та навчання моделей для виявлення аномалій:

- Статистичний підхід - цей підхід ґрунтується на тому, що аномалії є рідкісними випадками, які відхиляються від нормального розподілу даних.
- Метод машинного навчання - цей підхід використовує алгоритми машинного навчання для навчання моделі, яка може виявляти аномалії.

Статистичний підхід.

При статистичному підході аномалії виявляються шляхом порівняння даних з нормальним розподілом. Для цього можна використовувати такі методи, як:

- Критерій хі-квадрат - цей критерій використовується для порівняння даних з гіпотетичним розподілом.

- Тест Колмогорова-Смірнова - цей тест використовується для порівняння даних з нормальним розподілом.
- Тест Андерсона-Дарлинга - цей тест використовується для порівняння даних з різними розподілами.

Кластерний аналіз.

Суть даної групи методів полягає в розбитті безлічі спостережуваних векторів-властивостей системи на кластери, серед яких виділяють кластери нормального поведінки [3]. У кожному конкретному методі кластерного аналізу використовується своя метрика, яка дозволяє оцінювати приналежність спостережуваного вектора властивостей системи одному з кластерів або вихід за межі відомих кластерів. Методи кластерного аналізу можна розділити на 3 класи:

- ієрархічні методи,
- методи розбиття,
- комбіновані методи.

Результатом роботи ієрархічного методу є ієрархія кластерів (таксономія). Самі кластери можуть бути отримані шляхом розбиття дерева на основі певного критерію. Результатом роботи алгоритму, заснованого на методі розділення, є безліч кластерів, не пов'язаних між собою певною ієрархією. Комбіновані методи використовують ієрархічні методи, і методи розбиття. Найбільш часто використовується послідовно спочатку метод розбиття, а потім будується ієрархія на підставі отриманих кластерів. Даний алгоритм є найкращім з даної групи за рахунок двоступеневої кластеризації великих обсягів даних, що працює на обмеженому обсязі пам'яті, є локальним алгоритмом. Перевагою даної групи методів є те, що він є адаптивним. Недоліком є сильна залежність результату від вибору кількості кластерів і початкового розташування кластерів.[2]

Метод машинного навчання.

При методі машинного навчання аномалії виявляються шляхом навчання моделі на наборі даних, який містить як нормальні, так і аномальні дані. Модель навчається розпізнавати відмінності між нормальним і аномальними даними.

Для навчання моделей для виявлення аномалій можуть використовуватися різні алгоритми машинного навчання, такі як:

- Дерева рішень - дерева рішень можуть використовуватися для навчання моделі, яка може класифікувати дані як нормальні або аномальні.
- Навчання на підтримках - навчання на підтримках може використовуватися для навчання моделі, яка може виявляти аномалії на основі того, як часто дані зустрічаються в наборі даних.
- Нейронні мережі - нейронні мережі можуть використовуватися для навчання моделі, яка може виявляти аномалії на основі складних взаємозв'язків між даними.

Вибір підходу.

При виборі підходу до розробки та навчання моделей для виявлення аномалій необхідно враховувати такі фактори:

- Тип даних - які дані необхідно аналізувати?
- Кількість даних - скільки даних доступні для навчання?
- Точність - наскільки важливою є точність виявлення аномалій?
- Швидкість - наскільки важливою є швидкість виявлення аномалій?

Навчання моделі.

Після вибору підходу до розробки моделі необхідно навчити модель на наборі даних, який містить як нормальні, так і аномальні дані.

Для навчання моделі необхідно визначити такі параметри:

- Параметри алгоритму машинного навчання - деякі алгоритми машинного навчання мають параметри, які необхідно налаштувати для досягнення найкращої точності.
- Розмір набору даних - розмір набору даних впливає на точність моделі.
- Метод навчання - метод навчання впливає на час навчання моделі.

Використання моделі

Після навчання моделі її можна використовувати для виявлення аномалій у нових даних.

Для виявлення аномалій у нових даних необхідно визначити такі параметри:

- Порогове значення - порогове значення визначає, які дані вважаються аномальними.
- Метод виявлення - метод виявлення визначає, як аномалії будуть виявлені.

2.2.1 Вибір алгоритмів машинного навчання

Алгоритми машинного навчання - це методи, які дозволяють машинам навчатися на даних і робити прогнози. Існує безліч різних алгоритмів машинного навчання, кожен з яких має свої переваги та недоліки.

При виборі алгоритму машинного навчання необхідно враховувати такі фактори:

- Тип задачі - який тип задачі необхідно вирішити?
- Тип даних - які дані доступні для навчання?
- Важливість точності - наскільки важливою є точність прогнозів?
- Важливість швидкості - наскільки важливою є швидкість прогнозів?

Типи задач машинного навчання

Задачі машинного навчання можна розділити на такі категорії:

- Класифікація - це задача, яка полягає в тому, щоб класифікувати дані в одну з декількох категорій. Наприклад, можна використовувати класифікацію для виявлення спаму або для визначення виду рослини.
- Регресія - це задача, яка полягає в тому, щоб передбачити значення вихідної змінної на основі значення вхідних змінних. Наприклад, можна використовувати регресію для прогнозування продажів або для прогнозування погоди.
- Асоціативне навчання - це задача, яка полягає в тому, щоб знайти закономірності між даними. Наприклад, можна використовувати асоціативне навчання для виявлення тенденцій у продажах або для виявлення груп подібних клієнтів.

- Управління портфелем - це задача, яка полягає в тому, щоб розподілити ресурси між різними альтернативами. Наприклад, можна використовувати управління портфелем для розподілу інвестицій або для розподілу ресурсів між різними проектами.

Типи даних.

Дані, які використовуються для навчання алгоритмів машинного навчання, можуть бути дискретними або неперервними. Дискретні дані можуть приймати лише обмежене число значень, наприклад, стать людини може бути "чоловік" або "жінка". Неперервні дані можуть приймати будь-яке значення в заданому діапазоні, наприклад, температура повітря може бути будь-яким числом між -100 і 100 градусами Цельсія.

Важливість точності.

Точність прогнозів є важливим фактором, який необхідно враховувати при виборі алгоритму машинного навчання. Однак, важливо також враховувати важливість швидкості прогнозів. Якщо прогнози необхідні в реальному часі, то алгоритм повинен бути досить швидким, щоб обробляти дані вчасно.

Важливість швидкості.

Швидкість прогнозів також є важливим фактором, який необхідно враховувати при виборі алгоритму машинного навчання. Якщо прогнози не потрібні в реальному часі, то можна використовувати більш повільні алгоритми, які можуть забезпечити більш високу точність.

Популярні алгоритми машинного навчання.

Ось деякі з найбільш популярних алгоритмів машинного навчання:

- Дерева рішень - це простий і ефективний алгоритм, який може використовуватися для вирішення різних задач машинного навчання.
- Навчання на підтримках - це алгоритм, який ґрунтується на тому, як часто дані зустрічаються в наборі даних.
- Навчання на помилках - це алгоритм, який ґрунтується на тому, як часто дані класифікуються неправильно.

- Нейронні мережі - це складні алгоритми, які можуть забезпечувати високу точність, але вони також можуть бути повільними і складними у використанні.
- Статистичні методи - це методи, які ґрунтуються на статистиці для прогнозування даних.

2.2.2 Підготовка навчального набору даних

Навчання алгоритму машинного навчання відбувається на наборі даних, який містить як вхідні дані, так і вихідні дані. Вхідні дані - це дані, на основі яких алгоритм навчається робити прогнози. Вихідні дані - це дані, які алгоритм повинен прогнозувати.

Підготовка навчального набору даних є важливим етапом процесу машинного навчання. Вона включає в себе такі завдання:

- Збір даних - необхідно зібрати дані, які будуть використовуватися для навчання алгоритму.
- Очищення даних - необхідно очистити дані від помилок і непотрібних даних.
- Форматування даних - необхідно привести дані до формату, який може бути використаний алгоритмом.
- Створення наборів даних - необхідно розділити дані на навчальний і тестовий набори.

Збір даних.

Першим кроком у підготовці навчального набору даних є збір даних. Дані можна зібрати з різних джерел, таких як:

- Власні дані - можна використовувати власні дані, які були зібрані в рамках бізнесу або організації.
- Опубліковані дані - можна використовувати дані, які були опубліковані в Інтернеті або в інших джерелах.

- Дані, зібрані за допомогою датчиків - можна використовувати дані, які були зібрані за допомогою датчиків, таких як камери, мікрофони або сенсори.

Очищення даних.

Після того, як дані були зібрані, їх необхідно очистити від помилок і непотрібних даних. Помилки можуть виникнути під час збору даних або в результаті невірного введення даних. Непотрібні дані - це дані, які не є важливими для задачі, для якої збираються дані.

Для очищення даних можна використовувати різні методи, такі як:

- Видалення дублікатів - видалення дублікатів даних, які повторюються в наборі даних.
- Видалення помилок - виправлення або видалення помилок у даних.
- Заповнення пропусків - заповнення пропусків у даних.
- Видалення непотрібних даних - видалення даних, які не є важливими для задачі.

Форматування даних.

Після того, як дані були очищені, їх необхідно привести до формату, який може бути використаний алгоритмом. Алгоритми машинного навчання вимагають, щоб дані були представлені у певному форматі. Наприклад, деякі алгоритми вимагають, щоб дані були представлені у вигляді числових значень, а інші алгоритми можуть вимагати, щоб дані були представлені у вигляді текстових значень.

Для форматування даних можна використовувати різні методи, такі як:

- Перетворення даних у числовий формат - перетворення даних з текстового формату в числовий формат.
- Нормування даних - приведення даних до одного масштабу.
- Категоризація даних - приведення даних до категорій.

Створення наборів даних

Після того, як дані були очищені та відформатовані, їх необхідно розділити на навчальний і тестовий набори. Навчальний набір даних використовується для

навчання алгоритму, а тестовий набір даних використовується для оцінки точності алгоритму.

Найбільш поширеним співвідношенням між навчальним і тестовим наборами даних є співвідношення 80/20, тобто 80% даних використовуються для навчання алгоритму, а 20% даних використовуються для оцінки точності алгоритму.

2.2.3 Тестування та підтримка моделі після навчання

Після того, як модель навчена на навчальному наборі даних, її необхідно протестувати на тестовому наборі даних. Тестовий набір даних не використовувався для навчання моделі, тому він дає незалежну оцінку точності моделі.

Для тестування моделі можна використовувати такі метрики, як:

- Точність - це частка правильних прогнозів, зроблених моделлю.
- Акуратність - це частка правильних прогнозів, зроблених моделлю, серед всіх прогнозів, зроблених моделлю.
- F-міра - це комбінована метрика точності та акуратності.

Підтримка моделі після навчання

Після того, як модель протестована і визнана придатною для використання, її необхідно підтримувати. Підтримка моделі включає в себе такі завдання:

- Оновлення моделі - модель може бути оновлена новими даними, щоб забезпечити її актуальність.
- Контроль якості - модель повинна регулярно перевірятися, щоб переконатися, що вона все ще працює належним чином.
- Доступність - модель повинна бути доступна для використання в потрібний час.

Оновлення моделі.

З часом дані, на яких була навчена модель, можуть стати застарілими. У цьому випадку модель може бути оновлена новими даними, щоб забезпечити її актуальність.

Оновлення моделі може бути виконано за допомогою таких методів, як:

- Перенавчання - модель може бути навчена заново на новому наборі даних.
- Додаткове навчання - модель може бути дообучена на новому наборі даних.

Контроль якості.

Модель повинна регулярно перевірятися, щоб переконатися, що вона все ще працює належним чином. Це можна зробити за допомогою таких методів, як:

- Тестування на новому наборі даних - модель може бути протестована на новому наборі даних, щоб оцінити її точність.
- Аналіз помилок - помилки, зроблені моделлю, можуть бути проаналізовані, щоб виявити проблеми з моделлю.

Доступність

Модель повинна бути доступна для використання в потрібний час. Це означає, що модель повинна бути правильно розгорнута і підтримуватися.

2.3 Практичні випробування на реальних мережах

Після того, як модель навчена і протестована, її необхідно протестувати на реальній мережі. Це дозволяє оцінити точність моделі в реальних умовах.

Практичні випробування на реальних мережах можуть бути виконані за допомогою таких методів, як:

- Впровадження моделі в реальну мережу - модель може бути впроваджена в реальну мережу і використовуватися для виявлення аномалій.
- Симуляція реального трафіку - реальний трафік може бути продемонстрований моделі за допомогою симулятора.

Впровадження моделі в реальну мережу

Найбільш точним методом практичних випробувань є впровадження моделі в реальну мережу. Це дозволяє оцінити точність моделі в реальних умовах, з урахуванням всіх факторів, які впливають на мережу.

Однак, впровадження моделі в реальну мережу може бути складним і дорогим завданням. Крім того, впровадження моделі може порушити роботу мережі.

Симуляція реального трафіку.

Альтернативним методом практичних випробувань є симуляція реального трафіку. Це дозволяє оцінити точність моделі в умовах, які є близькими до реальних.

Симуляція реального трафіку може бути виконана за допомогою симулятора. Симулятор може генерувати трафік, який є схожим на реальний трафік.

Однак, симуляція реального трафіку може не враховувати всі фактори, які впливають на реальну мережу.

Вибір методу практичних випробувань

Вибір методу практичних випробувань залежить від таких факторів, як:

- Точність - важливо вибрати метод, який дозволить оцінити точність моделі в реальних умовах.
- Витрати - важливо вибрати метод, який буде економічно ефективним.
- Вплив на мережу - важливо вибрати метод, який не буде порушувати роботу мережі.

2.3.1 Використання розроблених моделей для виявлення аномалій в реальних мережах

Після того, як модель навчена, протестована і впроваджена в реальну мережу, вона може використовуватися для виявлення аномалій.

Модель може використовуватися для виявлення аномалій у реальному часі або постфактум.

Виявлення аномалій у реальному часі

Виявлення аномалій у реальному часі означає, що модель використовується для виявлення аномалій, коли вони відбуваються. Це дозволяє швидко реагувати на аномалії і запобігти їхньому поширенню.

Для виявлення аномалій у реальному часі модель повинна бути швидкою і ефективною. Крім того, модель повинна бути здатна обробляти великі обсяги даних.

Виявлення аномалій постфактум.

Виявлення аномалій постфактум означає, що модель використовується для виявлення аномалій, які відбулися в минулому. Це дозволяє аналізувати аномалії і виявляти їхні причини.

Для виявлення аномалій постфактум модель може використовуватися для аналізу архівних даних. Крім того, модель може використовуватися для аналізу даних, які були зібрані за допомогою засобів моніторингу мережі.

Оцінка результатів виявлення аномалій.

Результати виявлення аномалій необхідно оцінювати, щоб переконатися, що модель ефективно виявляє аномалії.

Для оцінки результатів виявлення аномалій можна використовувати такі метрики, як:

- Точність - це частка правильних прогнозів, зроблених моделлю.
- Акуратність - це частка правильних прогнозів, зроблених моделлю, серед всіх прогнозів, зроблених моделлю.
- F-міра - це комбінована метрика точності та акуратності.

Крім того, результати виявлення аномалій можуть бути оцінені за допомогою експертного аналізу. Експерти можуть оцінити, чи є аномалії, виявлені моделлю, реальними аномаліями.

2.3.2 Аналіз результатів та вдосконалення системи

Після того, як модель для виявлення аномалій впроваджена в реальну мережу, необхідно регулярно аналізувати її результати. Це дозволяє виявити проблеми з моделлю і вдосконалити її.

Аналіз результатів може бути виконано за допомогою таких методів, як:

- Оцінка метрик точності - оцінка метрик точності дозволяє визначити, наскільки добре модель виявляє аномалії.
- Аналіз помилок - аналіз помилок дозволяє визначити, які типи аномалій модель виявляє погано.
- Експертний аналіз - експертний аналіз дозволяє оцінити, чи є аномалії, виявлені моделлю, реальними аномаліями.

На основі результатів аналізу можна вдосконалити модель за допомогою таких методів, як:

- Зміна параметрів моделі - зміна параметрів моделі може дозволити поліпшити її точність.
- Доповнення набору даних - доповнення набору даних новими даними може дозволити моделі краще навчитися виявляти аномалії.
- Використання інших алгоритмів машинного навчання - використання інших алгоритмів машинного навчання може дозволити поліпшити точність моделі.

ВИСНОВОК

Збір та обробка мережевих даних є важливими завданнями для забезпечення безпеки та продуктивності мережі. Для виконання цих завдань існує широкий спектр інструментів та методів, які можна використовувати.

Засоби збору мережевого трафіку є важливим інструментом для забезпечення безпеки та продуктивності мережі. Вони можуть бути використані для різних цілей, таких як виявлення атак, аналіз використання мережі та виявлення проблем із мережею.

Попередня обробка даних - це важливий етап процесу аналізу даних. Вона включає в себе ряд завдань, спрямованих на підготовку даних до аналізу, таких як очищення, форматування та стандартизація. Виконання попередньої обробки даних є необхідним для отримання точних і достовірних результатів аналізу.

Розробка та навчання моделей для виявлення аномалій є складним завданням, яке вимагає знання статистики, машинного навчання та даних, які необхідно аналізувати.

Вибір алгоритму машинного навчання є складним завданням, яке вимагає знання задачі, даних і алгоритмів машинного навчання.

Підготовка навчального набору даних є важливим етапом процесу машинного навчання. Вона впливає на точність алгоритму, який буде навчений на цьому наборі даних.

Тестування і підтримка моделі після навчання є важливими завданнями, які забезпечують ефективність і надійність моделі.

Практичні випробування на реальних мережах є важливим етапом процесу розробки моделей для виявлення аномалій. Вони дозволяють оцінити точність моделі в реальних умовах.

Використання розроблених моделей для виявлення аномалій в реальних мережах є складним завданням. Воно вимагає, щоб модель була точною, ефективною і здатною обробляти великі обсяги даних. Крім того, результати виявлення аномалій необхідно оцінювати, щоб переконатися, що модель ефективно виявляє аномалії.

Аналіз результатів та вдосконалення системи є важливими етапами процесу розробки та впровадження моделей для виявлення аномалій. Вони дозволяють забезпечити ефективність і надійність системи.

Ось деякі конкретні приклади того, як можна вдосконалити систему виявлення аномалій:

- Можна використовувати більш складні алгоритми машинного навчання, такі як нейронні мережі. Нейронні мережі можуть забезпечити більш високу точність, ніж прості алгоритми, такі як дерева рішень.
- Можна використовувати більш великі набори даних для навчання моделі. Більші набори даних дозволяють моделі краще навчитися виявляти аномалії.

- Можна використовувати методи машинного навчання, які є адаптивними до змін. Адаптивні методи можуть поліпшити точність моделі, якщо дані в мережі змінюються.

Вдосконалення системи виявлення аномалій є постійним процесом. Інженери та дослідники постійно працюють над розробкою нових методів та технологій, які можуть допомогти забезпечити ефективність і надійність цих систем.

3 ЗАСТОСУВАННЯ ВИКОРИСТАННЯ ШІ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ В ПРАКТИЦІ

Безпека корпоративних мереж є одним з найважливіших завдань для будь-якої компанії. Атаки на корпоративні мережі можуть призвести до крадіжки даних, порушення роботи бізнесу та інших серйозних наслідків.

Штучний інтелект (ШІ) може використовуватися для підвищення безпеки корпоративних мереж за допомогою таких методів:

- Виявлення аномалій. ШІ може використовуватися для виявлення аномалій в поведінці користувачів, мережевих пристроїв та інших елементів корпоративної мережі. Аномальні події можуть бути ознакою атаки, тому їх виявлення дозволяє швидко реагувати на загрозу.
- Автоматизація захисту. ШІ може використовуватися для автоматизації завдань з безпеки, таких як сканування на вразливості, управління доступом та реагування на інциденти. Автоматизація дозволяє зменшити навантаження на персонал з безпеки та підвищити ефективність захисту.
- Розуміння загроз. ШІ може використовуватися для аналізу даних про загрози, щоб краще зрозуміти їхні мотиви та методи. Це дозволяє розробляти більш ефективні заходи захисту.

Виявлення аномалій

Одним з найважливіших завдань з безпеки корпоративних мереж є виявлення аномалій. Аномальні події можуть бути ознакою атаки, тому їх виявлення дозволяє швидко реагувати на загрозу.

ШІ може використовуватися для виявлення аномалій в поведінці користувачів, мережевих пристроїв та інших елементів корпоративної мережі. Наприклад, ШІ може використовуватися для виявлення наступних аномалій:

- Незвичайний трафік. ШІ може використовуватися для виявлення незвичайного трафіку в мережі, такого як різке збільшення кількості пакетів або незвичайні напрямки трафіку.

- Зловмисна активність. ШІ може використовуватися для виявлення зловмисної активності, такої як спроби входу в систему з невідомих пристроїв або запуск шкідливого коду.
- Вразливості. ШІ може використовуватися для виявлення вразливостей в мережевих пристроях та системах, які можуть бути використані для атаки.

Автоматизація захисту

ШІ також може використовуватися для автоматизації завдань з безпеки, таких як сканування на вразливості, управління доступом та реагування на інциденти. Автоматизація дозволяє зменшити навантаження на персонал з безпеки та підвищити ефективність захисту.

Наприклад, ШІ може використовуватися для автоматизації наступних завдань:

- Сканування на вразливості. ШІ може використовуватися для автоматичного сканування мережевих пристроїв та систем на наявність вразливостей.
- Управління доступом. ШІ може використовуватися для автоматичного управління доступом до ресурсів мережі, наприклад, на основі ролей користувачів.
- Рішення інцидентів. ШІ може використовуватися для автоматичного аналізу інцидентів безпеки та прийняття рішень про те, як на них реагувати.

Розуміння загроз

ШІ також може використовуватися для аналізу даних про загрози, щоб краще зрозуміти їхні мотиви та методи. Це дозволяє розробляти більш ефективні заходи захисту.

Наприклад, ШІ може використовуватися для наступних завдань:

- Аналіз даних про загрози. ШІ може використовуватися для аналізу даних про загрози, таких як відкриті джерела, звіти про безпеку та записи з мережевих пристроїв.
- Профілювання загроз. ШІ може використовуватися для створення профілів загроз, які описують їхні мотиви, методи та уразливості.

- Розробка заходів захисту. ШІ може використовуватися для розробки заходів захисту, які враховують профілі загроз.

Використання ШІ для підвищення безпеки корпоративних мереж є перспективним напрямком. ШІ дозволяє автоматизувати завдання з безпеки, підвищити ефективність виявлення загроз та краще зрозуміти їхні мотиви та методи. Це дозволяє компаніям краще захищатися від атак та мінімізувати ризики для своєї діяльності.

3.1 Розробка системи для виявлення мережевих аномалій.

Мережеві аномалії можуть бути викликані багатьма факторами, такими як шкідливі програми, хакерські атаки, або просто несправності обладнання. Виявлення мережевих аномалій є важливим завданням для забезпечення безпеки мережі. Сама система представлена в додатку А. Один із способів розробити таку систему - використовувати інструменти, такі як Scapy. Scapy - це потужна бібліотека для створення і аналізу мережевих пакетів.

Ця система працює наступним чином:

- Вона використовує бібліотеку Scapy для перехоплення пакетів з мережі.
- Для кожного перехопленого пакета вона перевіряє, чи є ознаки аномалії.
- Якщо аномалія виявлена, вона записується у файл з міткою аномалії в окремому стовпці.

У цьому коді є три правила виявлення аномалій:

- Висока інтенсивність трафіку від однієї IP-адреси. Це може бути ознакою атаки або інфільтрації.
- Використання нестандартних портів. Це може бути ознакою шкідливої діяльності, наприклад, запуску сервера C&C.
- Невдалі спроби входу. Це може бути ознакою спроби зламу облікового запису.

Ці правила можна налаштувати відповідно до потреб конкретної мережі. Наприклад, можна збільшити або зменшити значення порогів для кожного правила.

Крім того, можна додати інші правила виявлення аномалій. Наприклад, можна перевіряти, чи містить пакет певну шкідливу інформацію.

3.2 Навчання штучного інтелекту

Після того, як система зібрала достатньо даних для аналізу, вона може розпочати навчання штучного інтелекту. Для цього використовується алгоритм машинного навчання, який навчений розпізнавати аномалії в мережевому трафіку. Сам код штучного інтелекту представлений в додатку Б.

У даному випадку використовується алгоритм Random Forest Classifier, який є ансамблевим методом машинного навчання, що використовує групу дерев рішень для класифікації даних.

Далі наведено покроковий процес навчання штучного інтелекту:

- Завантаження даних

Першим кроком є завантаження даних, які будуть використовуватися для навчання. У даному випадку дані були зібрані системою і містяться в файлі data.csv.

- Поділення даних на навчальний та тестовий набори

Дані необхідно розділити на навчальний та тестовий набори. Навчальний набір використовується для навчання моделі, а тестовий набір - для оцінки точності моделі.

У даному випадку 30% даних використовуються для навчання, а 70% - для тестування.

- Створення моделі

Наступним кроком є створення моделі машинного навчання. У даному випадку використовується модель Random Forest Classifier.

- Навчання моделі

Модель навчається на навчальному наборі даних.

- Прогнозування на тестових даних

Після навчання модель використовується для прогнозування на тестових даних.

- Оцінка моделі.

Точність моделі оцінюється за допомогою метрик класифікації, таких як точність, чутливість і специфічність.

У даному випадку використовуються наступні метрики:

- Точність - це частка правильних прогнозів.
- Чутливість - це частка позитивних прогнозів, які були правильними.
- Специфічність - це частка негативних прогнозів, які були правильними.

Якщо точність моделі є високою, то модель добре навчається і може робити точні прогнози.

Приклад

Розглянемо приклад, як штучний інтелект може бути використаний для виявлення аномалій в мережевому трафіку.

Давайте припустимо, що система зібрала дані про мережевий трафік за останній місяць. Ці дані містять інформацію про IP-адреси, порти, типи протоколів і обсяг трафіку.

Система може використовувати ці дані для навчання моделі Random Forest Classifier. Модель може бути навчена розпізнавати аномалії в мережевому трафіку, такі як:

- Зростання обсягу трафіку з однієї IP-адреси.
- Зростання кількості пакетів, які надходять з одного порту.
- Зростання кількості пакетів, які використовують незвичайний протокол.

Після навчання модель може використовуватися для виявлення аномалій в реальному часі.

Наприклад, якщо модель прогнозує, що обсяг трафіку з однієї IP-адреси перевищує норму, то система може надіслати попередження адміністратору мережі.

Удосконалення точності моделі

Точність моделі машинного навчання може бути удосконалена за допомогою наступних методів:

- Збільшення кількості даних. Чим більше даних використовується для навчання, тим точнішою буде модель.
- Використання різних алгоритмів машинного навчання. Деякі алгоритми машинного навчання можуть бути більш точними для певних наборів даних.
- Ручна настройка параметрів моделі. Деякі параметри моделі можуть бути налаштовані для підвищення точності.

У випадку з виявленням аномалій в мережевому трафіку, точність моделі може бути удосконалена за допомогою наступних методів:

- Включення додаткових характеристик. У дані для навчання можна включити додаткові характеристики, такі як час доби, день тижня або тип пристрою, з якого надходить трафік.
- Використання методів машинного навчання, які є більш чутливими до аномалій. Деякі алгоритми машинного навчання, такі як Isolation Forest, можуть бути більш чутливими до аномалій, ніж інші.

Вибір методу удосконалення точності моделі залежить від конкретних потреб системи.

ВИСНОВОК

У цьому розділі було розглянуто застосування штучного інтелекту для виявлення аномалій в мережевому трафіку. Було показано, як система може збирати дані про мережевий трафік, навчати модель машинного навчання на цих даних та використовувати модель для виявлення аномалій в реальному часі.

Штучний інтелект має потенціал стати ефективним методом виявлення аномалій в мережевому трафіку. Він може забезпечити високу точність виявлення, а також може бути використаний для виявлення аномалій, які не можуть бути виявлені за допомогою традиційних методів.

Однак, існує також ряд проблем, які необхідно вирішити, щоб штучний інтелект міг бути ефективно використаний для виявлення аномалій в мережевому трафіку. Однією з проблем є необхідність великої кількості даних для навчання моделі. Іншою проблемою є необхідність удосконалення алгоритмів машинного навчання, щоб вони були більш чутливими до аномалій.

Незважаючи на ці проблеми, штучний інтелект є перспективним напрямком досліджень у галузі виявлення аномалій в мережевому трафіку. Він має потенціал поліпшити ефективність і безпеку мереж.

Рекомендації

На основі проведених досліджень можна зробити наступні рекомендації:

- Для підвищення точності моделі машинного навчання необхідно використовувати більше даних для навчання.
- Для підвищення чутливості моделі машинного навчання до аномалій необхідно використовувати методи машинного навчання, які є більш чутливими до аномалій.
- Для підвищення ефективності системи виявлення аномалій в мережевому трафіку необхідно розробити алгоритми, які можуть обробляти великі обсяги даних в реальному часі.

Впровадження цих рекомендацій дозволить поліпшити ефективність і безпеку систем виявлення аномалій в мережевому трафіку.

ВИСНОВКИ

У ході дослідження було розглянуто використання штучного інтелекту для виявлення аномалій в мережах і підвищення їхньої безпеки. Було встановлено, що штучний інтелект має значний потенціал для вирішення цієї проблеми, оскільки він дозволяє автоматизувати процес виявлення аномалій і підвищити його ефективність.

У роботі було розглянуто основні підходи до використання штучного інтелекту для виявлення аномалій в мережах. Було показано, що найефективнішими підходами є ті, які використовують машинне навчання для навчання на наборах даних, що містять нормальні та аномальні зразки.

У роботі також було розглянуто деякі практичні аспекти використання штучного інтелекту для виявлення аномалій в мережах. Було показано, що для ефективної реалізації таких систем необхідно враховувати такі фактори, як складність мережі, типи загроз, що розглядаються, і доступні ресурси.

На основі проведеного дослідження можна зробити такі висновки:

- Штучний інтелект є ефективним інструментом для виявлення аномалій в мережах.
- Машине навчання є найефективнішим підходом до використання штучного інтелекту для цієї мети.
- Для ефективної реалізації систем виявлення аномалій на основі штучного інтелекту необхідно враховувати такі фактори, як складність мережі, типи загроз, що розглядаються, і доступні ресурси.

У подальших дослідженнях у цьому напрямку необхідно:

- Розробити більш ефективні алгоритми машинного навчання для виявлення аномалій в мережах.

- Розробити методи для адаптації систем виявлення аномалій до змін в мережі.
- Розробити методи для підвищення точності виявлення аномалій в умовах обмежених ресурсів.

Впровадження систем виявлення аномалій на основі штучного інтелекту може істотно підвищити безпеку мереж. Такі системи дозволяють автоматизувати процес виявлення аномалій і підвищити його ефективність, що дозволяє значно зменшити ризик успішного проведення атак на мережу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ілона Лагун, Андрій Лагун / «Використання дискретного малохвильового перетворення для виявлення аномалій мережевого трафіку» // Національний університет «Львівська політехніка» // УДК 681.325 // 2011;
2. І.В. Рубан, В.О. Мартовицький, С.О. Партика / «Класифікація методів виявлення аномалій в інформаційних системах» // УДК 004.056.5 // Системи озброєння і військова техніка, 2016, № 3(47);
3. Y. Frank Jou, Fengmin Gong, Chandru Sargor, Shyhtsun FelixWu, and CleavelandW Rance, “Architecture design of a scalable intrusion detection system for the emerging network infrastructure.” // Technical Report CDRL A005, Dept. of Computer Science, North Carolina State University, Releigh, N.C, USA, April 1997.
4. https://hostiq.ua/blog/ukr/internet-phishing/
5. https://hyperhost.ua/info/ru/vidyi-atak-na-sayt-k-chemu-sleduet-byit-gotovyim
6. https://learn.ztu.edu.ua/pluginfile.php/269205/mod_resource/content/1/%D0%A8%D0%86_%D0%9B_8%20%28%D0%94%D0%95%D0%A2%D0%95%D0%9A%D0%A2_%D0%90%D0%9D%D0%9E%D0%9C%29.pdf
7. Навчальний посібник «Методи та системи штучного інтелекту» Лубко Д.В. Шаров С.В.//Напрямки використання штучного інтелекту//2019 -ст. 16-25.
8. Комплексна платформа машинного навчання з відкритим кодом. TensorFlow. / [Режим доступу – електронне джерело]:
9. https://www.tensorflow.org/;
10. Елена В. Карачанская, Надежда И. Соседова / «Метод виявлення аномалій мережевого трафіка, оснований на йому самопідній структурі» // Безпека інформаційних технологій = IT Security, Том 26, № 1 (2019);
11. И.М. Ажмухамедов, А.Н. Марьенков / «Пошук та оцінка аномалій мережевого трафіку на основі циклічного аналізу» // 2012;

12. Houssam Zenati, Chuan-Sheng F, Bruno Lecouat, GauravManek, Vijay Ramaseshan Chandrasekhar / «Efficient gan-based anomaly detection» // Workshop track - ICLR 2018;

13. https://scikit-learn.org/stable/

14. https://www.unite.ai/uk/10-best-machine-learning-software/

ДОДАТОК А

```
import os
from scapy.all import sniff, IP, TCP
from collections import Counter
import pandas as pd
from datetime import datetime

# Лічильники для IP-адрес та портів
ip_counter = Counter()
port_counter = Counter()
failed_logins = Counter()

def analyze_packet(packet):
    if IP in packet:
        source_ip = packet[IP].src
        ip_counter[source_ip] += 1

        # Аномалія - більше ніж 100 пакетів від однієї IP
        if ip_counter[source_ip] > 100:
            record_anomaly(source_ip, "High traffic")

        # Перевірка на нестандартні порти (припустимо, що стандартні - це 80 та 443)
        if TCP in packet:
            dst_port = packet[TCP].dport
            if dst_port not in [80, 443]:
                port_counter[dst_port] += 1
                if port_counter[dst_port] > 50:
                    record_anomaly(source_ip, f"Unusual port usage: {dst_port}")
```

```
        # Припустимо, що невдалі входи позначені у даних пакетів (це спрощення)
        # В реальних умовах це може бути складніше
        if "Failed login" in str(packet):
            failed_logins[source_ip] += 1
            if failed_logins[source_ip] > 10:
                record_anomaly(source_ip, "Multiple failed login attempts")

def record_anomaly(source_ip, anomaly_type):
    current_time = datetime.now().strftime("%Y-%m-%d %H:%M:%S")
    data = {
        "Timestamp": [current_time],
        "Source_IP": [source_ip],
        "Anomaly": [anomaly_type]
    }
    df = pd.DataFrame(data)

    # Створити файл, якщо він не існує
    if not os.path.exists("traffic_anomalies.csv"):
        df.to_csv("traffic_anomalies.csv", index=False)

    # Додати дані до файлу
    df.to_csv("traffic_anomalies.csv", mode='a', header=False, index=False)
    print(f"Anomaly from {source_ip} recorded: {anomaly_type}")

# Запуск перехоплення пакетів
print("Starting traffic monitoring...")
sniff(prn=analyze_packet, filter="ip", store=False)
```

ДОДАТОК Б

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

# Завантаження даних з файлу
file_path = 'data.csv'
data = pd.read_csv(file_path)

# Поділ даних на характеристики (X) та мітки (y)
X = data.drop(['Number', 'Timestamp', 'Source_IP', 'Destination_IP', 'Anomaly'],
axis=1)
y = data['Anomaly']

# Розділення на навчальний та тестовий набори
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Створення моделі та її навчання
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Прогнозування на тестових даних
predictions = model.predict(X_test)

# Оцінка моделі
print(classification_report(y_test, predictions))
```