МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

15.03 — КМР. 1636–“С” 2020.10.29. 010 ПЗ

АВРАМЕНКА БОГДАНА ТАРАСОВИЧА

2024 р.

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ

Завідувач кафедри

__ _____
(назва кафедри)

_____ _ ___
(підпис) (ПІБ)

“___” _____ 2024р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА
на тему

Система аналізу мережевого обладнання для забезпечення кібербезпеки в офісах

Спеціальність 122 – «Комп’ютерні науки»
ОП – «Інформаційні управляючі системи та технології»

Гарант освітньої програми

___ ____ _____ Голуб Б.Л.
(науковий ступінь та вчене звання) (підпис) (ПІБ)

Керівник бакалаврської кваліфікаційної роботи

__ _____ _Панкратьєв В.О.)

Виконав _____ ____Авраменко.Б.Т. _____
(підпис) (ПІБ студента)

КИЇВ - 2024

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
Факультет інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____

____

____ _____ _____ __.
(науковий ступінь, вчене звання) (підпис) (ПІБ)
"____" _____20_____р.

З А В Д А Н Н Я
на виконання магістерської кваліфікаційної роботи студенту
Авраменко Богдану Тарасовичу
(прізвище, ім'я, по батькові)

Спеціальність 122 – «Комп'ютерні науки»
1.Тема **магістерської** кваліфікаційної роботи ___ «Система аналізу мережевого обладнання для забезпечення кібербезпеки в офісах »_____
_____
затверджена наказом ректора НУБіП України від "____"_____20___ р. №_____
2.Термін подання завершеної роботи на кафедру _____
(рік, місяць, число)
3.Вихідні дані до бакалаврської кваліфікаційної роботи
4.Перелік питань, які потрібно розробити:
1, Аналыз проблемної області
2. Моделювання предметної області
3. Проектування програмної системи
4. Впровадження та експлуатації системи

Дата видачі завдання "_____"_____20___ р.

Керівник бакалаврської кваліфікаційної роботи _____ ____ _____
(підпис) (прізвище та ініціали)

Завдання прийняв до виконання _____ ____ Авраменко.Б.Т. _____
(підпис) (прізвище та ініціали студента)

АНОТАЦІЯ

Пояснювальна записка до магістерської роботи "Система аналізу мережевого обладнання для забезпечення кібербезпеки в офісах" складається з комплексних досліджень, розробок та аналізу, спрямованих на вдосконалення практики мережевої

безпеки та управління нею в офісних середовищах.

Об'єктом дослідження є мережева інфраструктура та пов'язане з нею обладнання, з акцентом на його роботу, моніторинг та управління кібербезпекою. Предметом дослідження є застосування аналітичних методів та автоматизованих систем для виявлення аномалій, класифікації подій та забезпечення кібербезпеки мережевого обладнання.

Метою кваліфікаційної магістерської роботи є розробка ефективної системи аналізу мережевих даних для підвищення ефективності виявлення загроз безпеці, оптимізації процесів моніторингу та підтримки прийняття рішень за рахунок автоматизації та методів інтелектуального аналізу.

Дослідження включало як емпіричні, так і теоретичні методи, інтегруючи машинне навчання, видобуток асоціативних правил та методи кластеризації. Практична реалізація ґрунтувалася на застосуванні цих аналітичних методів до реальних мережевих наборів даних для виявлення закономірностей, класифікації подій та надання дієвих висновків.

Під час роботи було вивчено існуючі системи аналізу мережевих даних, виявлено їхні сильні та слабкі сторони. Ці висновки лягли в основу розробки запропонованої системи, яка включає архітектуру, пристосовану для обробки великих обсягів мережевих даних і забезпечення виявлення загроз у реальному часі. Компоненти системи були розроблені з урахуванням масштабованості та ефективності, підтримуючи такі завдання, як кластеризація мережевих подій, виявлення аномалій та автоматизація реагування на інциденти кібербезпеки.

Результати роботи мають практичне застосування для мережевих адміністраторів та команд кібербезпеки. Розроблені моделі та аналітичні методи можуть бути інтегровані в існуючі системи моніторингу мереж для підвищення точності та швидкості виявлення загроз. Отримані результати сприяють розвитку автоматизації процесів кібербезпеки та оптимізації управління мережевою інфраструктурою.

ABSTRACT

The explanatory note to the master's thesis “Network Equipment Analysis System for Cybersecurity in Offices” consists of comprehensive research, development, and analysis

aimed at improving network security practices and management in office environments.

The object of research is network infrastructure and related equipment, with a focus on its operation, monitoring and cybersecurity management. The subject of the study is the application of analytical methods and automated systems to detect anomalies, classify events and ensure cybersecurity of network equipment.

The purpose of the master's thesis is to develop an effective network data analysis system to improve the efficiency of detecting security threats, optimize monitoring processes and support decision-making through automation and intelligent analysis methods.

The research included both empirical and theoretical methods, integrating machine learning, association rule mining, and clustering techniques. Practical implementation was based on applying these analytical methods to real network datasets to identify patterns, classify events, and provide actionable insights.

During the work, existing systems for analyzing network data were studied, and their strengths and weaknesses were identified. These conclusions formed the basis for the development of the proposed system, which includes an architecture adapted to process large amounts of network data and provide real-time threat detection. The system components were designed with scalability and efficiency in mind, supporting tasks such as network event clustering, anomaly detection, and automated cybersecurity incident response.

The results have practical applications for network administrators and cybersecurity teams. The developed models and analytical methods can be integrated into existing network monitoring systems to improve the accuracy and speed of threat detection. The obtained results contribute to the development of automation of cybersecurity processes and optimization of network infrastructure management.

Зміст

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

СУБД – Система управління базами даних.
HTML – HyperText Markup Language (Мова розмітки гіпертексту).
CSS – Cascading Style Sheets (Каскадні таблиці стилів).
JavaScript – Мова програмування для створення динамічного контенту веб-сторінок.
API – Application Programming Interface (Інтерфейс прикладного програмування).
UI – User Interface (Інтерфейс користувача).
SQL – Structured Query Language (Мова структурованих запитів).
UML – Unified Modeling Language (Уніфікована мова моделювання).
ОС – Операційна система.
ІС – Інформаційна система.

DBSCAN – Density-Based Spatial Clustering of Applications with Noise (Алгоритм кластеризації на основі густини).
FP-Growth – Frequent Pattern Growth (Алгоритм для пошуку частих патернів).
ECLAT – Equivalence Class Transformation (Метод пошуку частих наборів даних).
Apriori – Алгоритм пошуку асоціативних правил.
AIS – Artificial Immune System (Штучна імунна система для виявлення аномалій).
Mean-Shift – Метод кластеризації, що ідентифікує області з високою густотою.

ВСТУП

Актуальність теми.

Інтернет розвивається досить швидко.

Створення локальних мереж і підключення їх до глобальної мережі стає обов'язковим, як для великих підприємств і для домашніх користувачів. Зараз важко уявити будь-яке підприємство без будь-якої комп’ютерної техніки, адже вона полегшує виконання більшості сучасних завдань. Комп’ютери, будь то звичайні смартфони чи ноутбуки, чи потужні сервери, обробляють і передають багато інформації. Таке спілкування можливе завдяки комунікаційним мережам.

Комп’ютерні мережі, зокрема, складаються із середовища передачі даних, такого як: вита пара, волоконно-оптичний кабель, коаксіальний кабель, радіохвилі та активне або пасивне мережеве обладнання, таке як комутатор, маршрутизатор.

Мережеве обладнання випускається дуже великою кількістю виробників, таких як Juniper, Cisco, HP, Extreme, Mikrotik та багато інших. Всі пристрої відрізняються один від одного не тільки зовнішнім виглядом і ціною, але і підтримуваним функціоналом, інтерфейсом командного рядка або веб-інтерфейсом (при необхідності). Щоб розібратися в такій кількості мережевого обладнання і

підтримуваних ним технологіях, потрібно витратити багато часу на читання технічної документації або проходження спеціальних курсів. Однак навіть підготовленому фахівцеві необхідно докласти чимало зусиль для налаштування того чи іншого обладнання, ситуація ускладнюється ще й кількістю цього обладнання, від кількох десятків у межах якогось звичайного підприємства до тисяч у випадку з інтернет-провайдерами.

В даний час існують готові рішення у вигляді систем управління або обліку мережевого обладнання, але жодна система не поєднує обидві ці функції і не має ряду інших недоліків. Тому було прийнято рішення розробити власну систему, яка б відповідала всім потребам і не мала недоліків інших систем.

Об'єкт дослідження– мережеве обладнання, його управління та облік.

Мета кваліфікаційної бакалаврської роботи – спрощення процесу управління різноманітним мережевим обладнанням за рахунок автоматизації та уніфікації, вдосконалення облікового процесу.

Цілі дослідження:

1. Виконати рецензію за темою кваліфікаційної бакалаврської роботи роботи. Зокрема, проаналізувати сучасне мережеве обладнання.

2. Проаналізувати системи управління та обліку мережевого обладнання.

3. Встановіть вимоги та спроектуйте систему управління та обліку мережевого обладнання.

4. Здійснити програмну реалізацію системи контролю та обліку мережевого обладнання.

Наукове значення полягає в застосуванні методів проектування та розробки програмного забезпечення, а також методів програмного забезпечення для розробки нової системи управління та обліку мережевого обладнання.

Практичне значення. Розроблений програмний продукт дозволить автоматизувати та уніфікувати процес управління мережевим обладнанням та полегшить завдання його обліку.

У процесі роботи було проведено аналіз існуючих мережевих систем, визначено їх переваги та недоліки, що було корисно при розробці та проектуванні власної системи, її архітектури та компонентів.

Методи та засоби дослідження– методи програмування розробки програмного забезпечення, Python, HTML, CSS.

Результати кваліфікаційної бакалаврської роботи роботи можуть бути використані для уніфікованого управління мережевими ресурсами.

1.1 Актуальність теми

У сучасному взаємопов'язаному цифровому середовищі офіси значною мірою покладаються на надійну та безпечну мережеву інфраструктуру для підтримки роботи, захисту конфіденційних даних та забезпечення безперебійної комунікації. Моя система аналізу мережевого обладнання спрямована на вирішення критично важливих проблем у захисті офісного середовища від зростаючих кіберзагроз. Ця тема є особливо актуальною через зростаючу складність мережевих систем та підвищені ризики, пов'язані з кібератаками, спрямованими на вразливе мережеве обладнання.

Порушення кібербезпеки, такі як атаки з вимогою викупу, витік даних та несанкціонований доступ, часто використовують вразливості мережевого обладнання, включаючи маршрутизатори, комутатори та брандмауери. Ці пристрої слугують точками входу для зловмисників, що робить їх аналіз та моніторинг життєво важливим компонентом стратегії захисту організації. Зосереджуючись на розробці системи аналізу мережевого обладнання, ця робота сприяє виявленню та зменшенню ризиків, забезпечуючи захист критичної інфраструктури в офісних середовищах.

Сучасні офіси включають в себе передові технології, такі як пристрої Інтернету речей, хмарні обчислення та системи віддаленого доступу, які ще більше збільшують поверхню атаки. Актуальність цієї теми полягає в її відповідності сучасним технологічним тенденціям і необхідності захисту цих інновацій. Добре розроблена система аналізу використовує сучасні алгоритми аналізу для проактивного виявлення потенційних вразливостей і реагування на нові загрози.Ця тема має пряме практичне значення для покращення кібербезпеки в офісних мережах. Аналіз мережевого обладнання дозволяє IT-адміністраторам

Моніторити та керувати продуктивністю пристроїв для виявлення аномалій або порушень в режимі реального часу.

Виявляти застаріле або вразливе програмне забезпечення та конфігурації, які можуть наражати мережу на загрози.

Підвищити відповідність нормам захисту даних, забезпечивши безпеку мережевих налаштувань.

Вирішуючи ці завдання, система посилить загальний рівень безпеки офісів, зменшить час простою та запобігатиме фінансовим і репутаційним збиткам.

Це дослідження не лише збагачує академічне розуміння систем мережевої безпеки, але й пропонує практичні рекомендації для IT-фахівців. Оскільки кіберзагрози

стають все більш витонченими, ця робота надає організаціям інструменти для ефективного захисту своїх операцій.
Аналіз мережевого обладнання для кібербезпеки в офісах є дуже впливовою темою. Вона поєднує академічні дослідження з реальними застосуваннями, пропонуючи значущий внесок у критично важливу сферу кібербезпеки.

1.2 Об'єкт і предмет дослідження

Об'єктом дослідження є процеси моніторингу та аналізу даних у великомасштабних інформаційних системах. Ці системи необхідні для підтримки контролю над мережевими ресурсами та виявлення аномалій, які можуть свідчити про загрози кібербезпеці. З огляду на стрімке зростання обсягів мережевого трафіку і даних, потреба в передових методах автоматизації моніторингу та аналізу є критично важливою. Це включає використання сучасних технологій для виявлення та класифікації інцидентів, пов'язаних з безпекою, в режимі реального часу, підвищуючи загальну безпеку системи та її стійкість до потенційних кібератак.

Предметом дослідження є використання передових аналітичних методів, таких як класифікація, асоціативний аналіз і кластерний аналіз для оцінки даних, отриманих від мережевих датчиків. Ці датчики відстежують різні аспекти, такі як транспортні потоки, контроль доступу та параметри навколишнього середовища (наприклад, температуру, рух). Аналізуючи дані, отримані від датчиків, стає можливим виявити закономірності та залежності, виявити аномальну поведінку та згрупувати схожі інциденти для більш структурованої та проактивної системи управління загрозами.

Це дослідження підкреслює важливість розробки точних і надійних методів аналізу мережевих даних, особливо в середовищах з високим ризиком кіберзагроз. Динамічна природа сучасних мережевих інфраструктур вимагає адаптивних аналітичних підходів, здатних реагувати на постійні зміни в потоках даних. Тому це дослідження спрямоване на створення методів, які забезпечують не лише ефективне виявлення аномалій у реальному часі, але й покращують класифікацію та розуміння інцидентів, пов'язаних з кібербезпекою.

Результати таких досліджень є практичними та впливовими, оскільки вони сприяють підвищенню стабільності та безпеки мережі. Просуваючи автоматизовані методи моніторингу та аналізу, дослідження відповідає нагальній потребі в надійних системах кібербезпеки у великих офісних середовищах та за їх межами.

1.3. Мета та завдання дослідження

Метою цього дослідження є розробка та впровадження ефективного підходу до аналізу мережевого обладнання для посилення кібербезпеки в офісних середовищах. Зі збільшенням складності кіберзагроз та зростанням залежності від взаємопов'язаних систем в сучасних офісах, забезпечення захисту мережевої інфраструктури є критично важливим пріоритетом. Це дослідження має на меті забезпечити систематичний та автоматизований метод моніторингу, аналізу та класифікації даних, що генеруються мережевими пристроями, для виявлення вразливостей, визначення потенційних загроз та своєчасного реагування на ризики кібербезпеки. Кінцевою метою є посилення надійності та безпеки офісних мереж при збереженні операційної стабільності.
Для досягнення окресленої мети дослідження керується наступними завданнями:
Розробити структурований підхід до моніторингу та аналізу даних, що генеруються мережевим обладнанням, включаючи маршрутизатори, комутатори та датчики.
Включити сучасні аналітичні методи, такі як класифікація, кластеризація та виявлення аномалій для обробки даних.
Вивчіть існуючу мережеву інфраструктуру, щоб виявити типові вразливості та загрози, характерні для офісних середовищ.
Підвищення автоматизації та точності
Інтегрувати автоматизовані рішення для безперервного моніторингу та аналізу мережевої активності.
Забезпечте точність і надійність виявлення потенційних інцидентів кібербезпеки за допомогою вдосконалених алгоритмів.
Розробка та оцінка системи виявлення загроз
Розробити систему, яка використовує аналіз даних для класифікації та реагування на мережеві інциденти.
Перевірте ефективність системи у виявленні, класифікації та зменшенні загроз безпеці в офісних мережах.
Виконуючи ці завдання, дослідження зробить внесок у ширшу сферу кібербезпеки, вирішуючи критичну проблему захисту мережевого обладнання в офісних середовищах, забезпечуючи як захист даних, так і безперервність бізнесу.

1.3 Методи дослідження

У цьому дослідженні використовується поєднання передових аналітичних методів і методів візуалізації для забезпечення всебічного аналізу та виявлення загроз у мережевих даних. Обрані методи дозволяють ефективно обробляти та інтерпретувати великі масиви даних, що генеруються мережевими датчиками, надаючи практичні висновки для посилення кібербезпеки в офісних середовищах. У дослідженні були використані наступні методи:

1) Метод класифікації використовує алгоритми машинного навчання, такі як 1-Rule та Naive Bayes, для розподілу мережевих подій за попередньо визначеними категоріями безпеки. Аналізуючи параметри подій і відносячи їх до певних класів, цей метод сприяє швидкому виявленню інцидентів, які потребують негайної уваги. Він дає змогу автоматизовано та ефективно обробляти події зі схожими характеристиками, скорочуючи час, необхідний для ручної перевірки, та покращуючи реагування на потенційні загрози.

2) Для виявлення закономірностей і взаємозв'язків між мережевими подіями в дослідженні застосовується алгоритм Apriori. Цей метод визначає часті набори подій і встановлює асоціації, допомагаючи зрозуміти, які події відбуваються разом або послідовно. Асоціативний аналіз особливо цінний для побудови моделей прогнозування на основі історичної поведінки мережі, що дозволяє організаціям передбачати потенційні загрози, розуміючи повторювані комбінації подій.

3) Кластерний аналіз, в основі якого лежить метод K-середніх, використовується для групування подій зі схожими атрибутами в окремі кластери. Цей метод має важливе значення для виявлення аномалій і виявлення ненормальних патернів у мережевому трафіку. Організовуючи дані в значущі кластери, дослідження дає уявлення про сегментацію подій, що дозволяє глибше зрозуміти поведінку мережі та виявити потенційні ризики або вразливості.

4) Для покращення інтерпретації та прийняття рішень дослідження інтегрує методи візуалізації даних, включаючи графіки та діаграми. Ці візуальні інструменти використовуються для представлення результатів кластеризації та асоціативного аналізу, виділяючи ключові закономірності та аномалії в чіткому і зрозумілому форматі. Візуалізація є критично важливим компонентом для оцінки ефективності аналітичних моделей і полегшення комунікації висновків із зацікавленими сторонами.

Завдяки інтеграції цих методів дослідження забезпечує надійний і системний підхід до аналізу мережі. Ця комплексна методологія дозволяє виявляти аномалії в режимі реального часу, ефективно класифікувати події та розробляти прогнозні моделі, що значно підвищує безпеку та відмовостійкість мережевих інфраструктур офісів.

1.4 Наукова новизна і практична значущість роботи

Наукова новизна цього дослідження полягає в розробці комплексного підходу до обробки та аналізу великих потоків мережевих даних для виявлення аномалій та запобігання потенційним загрозам в режимі реального часу. Унікальним аспектом цього дослідження є інтеграція декількох методів машинного навчання - класифікації, асоціативного аналізу та кластеризації - які рідко застосовуються разом у контексті мережевої безпеки. Таке поєднання не лише дозволяє виявляти аномалії, але й розкриває закономірності та взаємозв'язки між мережевими подіями, створюючи надійну систему для безперервного моніторингу поведінки мережі.

Інноваційно, дослідження використовує алгоритм 1-Rule для базової класифікації подій та Naive Bayes для більш точного аналізу загроз. Асоціативний аналіз ще більше посилює його, виявляючи приховані зв'язки між подіями, що дозволяє ідентифікувати моделі кіберзагроз на основі історичних даних. Ці знання сприяють створенню прогностичних моделей, які можуть запобігати інцидентам у сфері безпеки. Кластеризація, зокрема за допомогою методу K-середніх, сегментує дані на групи зі схожими характеристиками, що дозволяє швидко ідентифікувати кластери, які можуть вказувати на приховані загрози або аномалії. Такий багатогранний підхід забезпечує ефективний моніторинг і надає можливість масштабування для адаптації системи до різних середовищ і масштабів даних.

Практичне значення даного дослідження полягає в розробці ефективної автоматизованої системи моніторингу та аналізу мережевих даних. Ця система може бути легко інтегрована в існуючі інфраструктури безпеки. Використання алгоритмів класифікації та кластеризації покращує час реагування на інциденти безпеки, одночасно спрощуючи процес моніторингу для аналітиків. Запропоновані моделі добре адаптуються до реальних умов, особливо в мережах з високим трафіком, де необхідна швидка обробка великих обсягів даних.

Реалізація асоціативного аналізу дозволяє виявляти нетрадиційні, але критичні взаємозв'язки між подіями, що дає змогу співробітникам служби безпеки приймати обґрунтовані рішення в умовах обмеженого часу. Практичне застосування цього

дослідження зменшує навантаження на фахівців з кібербезпеки завдяки автоматизованому аналізу подій та класифікації аномалій. Це мінімізує помилкові спрацьовування систем безпеки, тим самим підвищуючи їхню точність та ефективність. Крім того, ці методи можуть бути інтегровані в сучасні системи підтримки прийняття рішень для моніторингу та управління мережевою безпекою, значно підвищуючи надійність і стабільність інформаційних інфраструктур.

Застосовуючи ці алгоритми до великомасштабних інформаційних систем, що працюють під високим навантаженням і в динамічних мережевих умовах, це дослідження є важливим кроком на шляху до забезпечення кібербезпеки в цифровому середовищі. Отримані результати можуть підвищити безпеку існуючих систем, а також слугувати основою для нових високоефективних рішень у сфері кібербезпеки.

2 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

2.1. Огляд сучасних технологій для аналізу даних

У сучасному світі аналіз даних відіграє ключову роль у різних секторах, включаючи кібербезпеку, охорону здоров'я, фінансові послуги, бізнес-операції та багато інших. Постійний розвиток технологій збору та обробки даних дозволив організаціям обробляти величезні обсяги даних у режимі реального часу, що дає змогу приймати обґрунтовані рішення та швидко реагувати на виклики. Нижче наведено огляд сучасних технологій та інструментів аналізу даних, які зазвичай використовуються в інформаційній безпеці та моніторингу мереж.

1) Машинне навчання та алгоритми класифікації

Машинне навчання дозволяє системам автоматично навчатися на основі даних і прогнозувати результати шляхом виявлення закономірностей. Алгоритми класифікації, такі як наївний Байєс, 1-правило, дерева рішень, випадковий ліс і машини опорних векторів (SVM), широко використовуються для класифікації подій і виявлення аномалій у потоках даних в режимі реального часу. Наприклад, випадковий ліс і наївний Байєс особливо ефективні для класифікації подій мережевого трафіку, допомагаючи ефективно виявляти потенційні кіберзагрози.

2) Глибоке навчання

Глибоке навчання, підвид машинного навчання, використовує багатошарові нейронні мережі для аналізу складних структур даних. Здатність розпізнавати складні закономірності та взаємозв'язки робить його цінним у таких сферах, як обробка

зображень, обробка природної мови та кібербезпека. У мережевій безпеці глибоке навчання використовується для виявлення незвичайних патернів у трафіку та підвищення точності моделей прогнозування загроз, особливо в динамічному та непередбачуваному середовищі.

3) Технології асоціативного аналізу

Асоціативний аналіз фокусується на виявленні взаємозв'язків між елементами даних. Алгоритм Apriori, один з найвідоміших у цій категорії, визначає часті шаблони подій, допомагаючи виявляти повторювані взаємодії. У кібербезпеці ця технологія використовується для встановлення кореляцій між мережевими подіями, що може забезпечити раннє попередження про потенційні інциденти безпеки.

4) Кластерний аналіз

Методи кластеризації групують схожі точки даних у кластери на основі спільних характеристик. Такі алгоритми, як K-середні, DBSCAN та ієрархічна кластеризація, широко застосовуються для сегментації наборів даних, що полегшує виявлення аномальних кластерів. В інформаційній безпеці кластеризація допомагає виявляти групи подій, які відхиляються від нормальної поведінки, що вказує на можливу зловмисну активність.

5) Великі дані та обробка в реальному часі

Технології великих даних, такі як Hadoop і Apache Spark, дозволяють зберігати і аналізувати величезні масиви даних на розподілених серверах. Ці інструменти чудово справляються з обробкою великих обсягів даних у режимі реального часу, що робить їх незамінними для додатків кібербезпеки. Вони підтримують швидке виявлення аномалій і загроз, аналізуючи високопродуктивні мережеві потоки даних.

6) Візуалізація даних

Такі інструменти, як Tableau, Power BI і Matplotlib, полегшують представлення аналітичних результатів у наочній і зрозумілій формі. Візуалізація розширює можливості для швидкого виявлення тенденцій, закономірностей і відхилень. У контексті мережевої безпеки це дозволяє фахівцям ефективно контролювати стан системи та оперативно реагувати на аномалії або загрози.

Сучасні технології аналізу даних надають безпрецедентні можливості для посилення безпеки інформаційних систем. Інтеграція машинного навчання, глибокого навчання, асоціативного аналізу та кластеризації дозволяє організаціям

виявляти та прогнозувати загрози на основі історичних даних. Платформи Big Data забезпечують ефективну роботу з даними в режимі реального часу, а інструменти візуалізації роблять результати доступними і придатними до дії. Ці технології складають основу автоматизованих систем моніторингу та реагування на загрози, що мають вирішальне значення для захисту сучасних динамічних і взаємопов'язаних мережевих середовищ.

2.2 Методи і алгоритми для кластерного аналізу та пошуку асоціативних правил

Кластерний аналіз, або кластеризація, — це метод групування об'єктів на основі їх схожості. У контексті аналізу даних та інформаційної безпеки, кластерний аналіз дозволяє структурувати великі обсяги даних для виділення сегментів або груп подій, що мають спільні характеристики, а також для виявлення аномалій. Важливою особливістю кластерного аналізу є те, що він не потребує заздалегідь визначених міток для даних, дозволяючи автоматично виявляти структури в інформації. Розглянемо основні алгоритми кластеризації детальніше.

Алгоритм K-means є одним із найбільш поширених методів кластерного аналізу. Суть алгоритму полягає у розподілі даних на KKK кластерів на основі відстані до центрів (центроїдів) кластерів. Першим кроком є ініціалізація KKK центроїдів, після чого кожен об'єкт даних приєднується до найближчого центроїду. Після цього центроїди переміщуються до середнього положення об'єктів у своєму кластері. Процес повторюється до тих пір, поки центроїди не перестануть змінюватися, що свідчить про стабілізацію кластерів.

Основні характеристики алгоритму K-means:

- Потребує заздалегідь визначеної кількості кластерів (KKK).
- Чутливий до вибору початкових центроїдів, що може вплинути на кінцевий розподіл даних.
- Ефективний для великих обсягів даних, але обмежений у здатності кластеризувати складні структури, такі як кластери з різними розмірами або густотою.
- Швидкий для роботи з великими масивами даних завдяки лінійній часовій складності.

Застосування K-means у сфері інформаційної безпеки може бути ефективним для виділення груп аномальних дій або об'єктів, що мають схожі характеристики, таких як поведінка користувачів або мережеві події. Це дозволяє, наприклад, ідентифікувати групи підозрілих користувачів або сегментувати трафік для глибшого аналізу.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN є алгоритмом кластеризації на основі густини, що дозволяє ефективно знаходити кластери довільної форми та виявляти аномалії як окремі точки або невеликі групи точок. На відміну від K-means, DBSCAN не потребує визначення кількості кластерів заздалегідь. Алгоритм розпізнає кластери як області з високою густотою точок, а області з низькою густотою класифікує як шум або аномалії.

Основні етапи роботи DBSCAN:

- Алгоритм починається з вибору випадкової точки і визначення, чи має вона достатню кількість сусідів (згідно з параметрами радіуса (є\epsilonє) та мінімальної кількості точок).
- Якщо точка відповідає вимогам, вона стає ядром кластера, а сусідні точки приєднуються до нього.
- Процес повторюється для кожної точки, поки не сформуються кластери.

DBSCAN є корисним для виявлення аномалій у мережевих даних, оскільки дозволяє ідентифікувати ізольовані або рідко зустрічаються події як потенційно підозрілі. Цей алгоритм особливо підходить для аналізу даних, де очікується наявність нерівномірно розподілених кластерів, наприклад, під час моніторингу мережевого трафіку.

Ієрархічна кластеризація

Ієрархічна кластеризація будує дендрограму — структуру, що представляє ієрархічні зв’язки між об’єктами даних. Вона підходить для завдань, де необхідно отримати багато рівнів кластеризації. Основні види ієрархічної кластеризації:

- Агломеративна: починається з кожного об'єкта як окремого кластера, які поступово об'єднуються у більші кластери.
- Дивізивна: починається з усіх об'єктів у єдиному кластері, який поступово розбивається на менші кластери.

Ієрархічна кластеризація дозволяє виділяти групи різних рівнів у даних, що корисно для багатошарового аналізу. У сфері інформаційної безпеки її можна

застосовувати для багаторівневої класифікації загроз або ієрархічного групування різних типів подій для детальнішого аналізу.

Mean-Shift є методом, що ідентифікує області з високою густотою шляхом зсуву центроїдів до найбільш густих точок даних. Mean-Shift не потребує попереднього знання кількості кластерів, оскільки він автоматично знаходить області з високою густотою, де утворюються кластери.

Mean-Shift підходить для обробки складних даних, зокрема у завданнях з великими обсягами трафіку. У безпекових сценаріях він дозволяє ідентифікувати групи подій з різною щільністю, що може бути корисним для виявлення атак з різними рівнями інтенсивності або пошуку аномальних поведінкових патернів у мережевих середовищах.

Пошук асоціативних правил є методом, що використовується для виявлення закономірностей у даних. Він дозволяє виявити часті набори подій та правила, що описують зв'язки між ними, наприклад, "якщо подія A відбувається, то подія B також ймовірно відбудеться". У мережевій безпеці асоціативні правила можуть допомогти ідентифікувати типові моделі загроз або сценарії атак.

Алгоритм Apriori є одним із найперших і найпоширеніших методів асоціативного аналізу. Він працює за принципом пошуку частих наборів елементів у даних, які з’являються разом з певною частотою. Apriori використовує підтримку та довіру для визначення найважливіших правил.

Основні кроки роботи алгоритму:

1. Знаходження частих елементів з мінімальною підтримкою.
2. Побудова більших частих наборів шляхом комбінації елементів, які вже задовольняють мінімальну підтримку.
3. Формування асоціативних правил на основі знайдених частих наборів.

Apriori добре підходить для пошуку закономірностей у великих наборах даних і дозволяє визначати типові взаємозв'язки між подіями у системі безпеки.

FP-Growth (Frequent Pattern Growth)

FP-Growth є вдосконаленою версією алгоритму Apriori, який дозволяє скоротити кількість операцій шляхом використання дерева частих патернів (FP-дерево). FP-Growth забезпечує швидший пошук частих наборів без необхідності зберігання всіх можливих комбінацій.

FP-Growth підходить для великих наборів даних, де стандартний Apriori може виявитися неефективним через кількість комбінацій. У сфері безпеки FP-Growth дозволяє виявляти зв’язки між подіями у великих масивах даних, наприклад, при аналізі історичних даних про атаки.

ECLAT (Equivalence Class Transformation)

ECLAT використовує вертикальний підхід до пошуку частих наборів, який зберігає інформацію про транзакції для кожного елемента. Завдяки цьому він може швидко знаходити перетини між списками транзакцій, де трапляються певні елементи, і визначати часті набори без необхідності багаторазового сканування даних, як у випадку з Apriori.

ECLAT особливо ефективний для роботи з великими та вертикально структурованими даними, де кожен елемент або подія представлена як окрема колонка, що містить індекси транзакцій. Цей алгоритм знаходить часті набори швидше завдяки обчисленню перетинів списків транзакцій. У мережевій безпеці він підходить для аналізу подій, де важливо визначати часті набори транзакцій, наприклад, під час виявлення закономірностей у взаємодіях між різними типами атак.

AIS (Artificial Immune System)

Алгоритм AIS базується на принципах біологічних імунних систем і використовує аналогії з процесами взаємодії антигенів та антитіл для пошуку закономірностей у даних. AIS дозволяє динамічно адаптуватися до нових умов, виявляючи нові патерни та відхилення, що можуть вказувати на потенційні загрози. Цей підхід особливо корисний для виявлення аномалій, оскільки моделі, побудовані за допомогою AIS, можуть виявляти нові, раніше невідомі загрози, подібно до того, як імунна система реагує на нові патогени.

Основні етапи AIS включають:

1. Створення початкового набору антитіл, які реагують на певні патерни у даних.
2. Визначення подібності між вхідними даними та існуючими антитілами.
3. Збільшення частоти появи антитіл, які успішно виявляють аномалії, та адаптація системи до нових патернів.

AIS є надзвичайно гнучким методом для аналізу даних у мережевій безпеці, оскільки він дозволяє створювати адаптивні моделі, які оновлюються в реальному

часі на основі нових даних. Це робить його ідеальним для динамічних середовищ, де характер загроз постійно змінюється.

Порівняння алгоритмів кластеризації та асоціативного аналізу

Кожен з розглянутих алгоритмів кластерного аналізу та асоціативного аналізу має свої переваги та недоліки, які визначають їхню придатність для конкретних завдань у сфері інформаційної безпеки.

Розглянуті алгоритми кластеризації та асоціативного аналізу пропонують потужний інструментарій для аналізу даних у системах безпеки. Використання кластерного аналізу дозволяє структуровано групувати події за схожими ознаками, що є ефективним для виявлення аномалій та ідентифікації подій, які можуть становити загрозу. Асоціативний аналіз допомагає встановлювати закономірності між подіями, що дозволяє передбачати можливі загрози та створювати прогностичні моделі.

Таким чином, ефективне поєднання кластерного аналізу та асоціативного аналізу в межах єдиної системи моніторингу здатне значно підвищити безпеку мережевих систем. Це дозволяє швидко реагувати на інциденти, мінімізувати кількість хибних спрацьовувань та забезпечити надійний контроль за інфраструктурою в умовах постійно зростаючих кіберзагроз.

2.2 Огляд існуючих рішень в аналізі даних для оптимізації процесів

У сучасному цифровому ландшафті рішення для аналізу даних стали незамінними інструментами для організацій, які прагнуть оптимізувати свої процеси. Ці рішення допомагають підприємствам і галузям у різних сферах, включаючи кібербезпеку, обробку великих даних, візуалізацію даних і машинне навчання, отримувати дієві ідеї, підвищувати ефективність і приймати обґрунтовані рішення. У цьому огляді розглядаються існуючі рішення для аналізу даних у цих критично важливих сферах, висвітлюються їхні можливості та сфери застосування.

Рішення для аналізу даних для кібербезпеки

Зі зростанням кількості кіберзагроз, рішення для кібербезпеки, що використовують передову аналітику даних, набувають все більшого значення. Ці інструменти дозволяють організаціям виявляти, запобігати та реагувати на інциденти безпеки, аналізуючи величезні обсяги мережевого трафіку, журналів та поведінки користувачів.

Алгоритм	Переваги	Недоліки	Застосування у безпеці
K-means	Швидкий, ефективний для великих даних; простота реалізації	Потребує визначення кількості кластерів; не працює з шумом	Кластеризація типових подій
DBSCAN	Виявляє кластери довільної форми, працює з шумом	Вибір параметрів радіуса та щільності може бути складним	Виявлення аномалій та сегментування даних
Ієрархічна	Створює багаторівневу кластеризацію; не потребує параметрів К	Повільний для великих даних, неефективний для складних структур	Групування подій для глибинного аналізу
Mean-Shift	Автоматично визначає кількість кластерів; не чутливий до шуму	Висока обчислювальна складність	Кластеризація подій з різною щільністю
Apriori	Легкий для розуміння, добре працює для малих даних	Неефективний для великих наборів даних	Виявлення зв’язків між подіями
FP-Growth	Ефективний для великих даних завдяки FP-дереву	Складний для реалізації	Пошук зв'язків у великих обсягах даних
ECLAT	Швидкий для вертикально організованих даних	Не підходить для горизонтально структурованих наборів	Асоціація між подіями

Провідні рішення:

Splunk - це популярна платформа для аналізу даних, яка надає інформацію про IT-середовище в режимі реального часу. Вона дозволяє організаціям збирати, відстежувати та аналізувати машинні дані з різних

AIS	Адаптивний, виявляє нові загрози	Складний для налаштування; потребує великої обчислювальної потужності	Виявлення нових, непередбачуваних загроз

джерел. Splunk широко використовується для аналізу журналів, виявлення вторгнень та управління загрозами в кібербезпеці.

IBM QRadar:це рішення для управління інформацією та подіями безпеки (SIEM), яке використовує розширену аналітику для виявлення аномалій і кореляції подій у мережах. Це дозволяє організаціям проактивно виявляти та пом'якшувати потенційні порушення безпеки.

Cortex XDR:Розроблений Palo Alto Networks, Cortex XDR надає розширені можливості виявлення та реагування. Він інтегрує дані з різних джерел, включаючи кінцеві точки, мережі та хмарні середовища, щоб виявляти складні загрози та оптимізувати зусилля з реагування.

Darktrace:використовує штучний інтелект і машинне навчання для автономного виявлення та реагування на кіберзагрози. Він відстежує мережеву активність у режимі реального часу, виявляючи аномалії, які можуть свідчити про внутрішні загрози або зовнішні атаки.

Переваги:

Моніторинг та аналіз загроз у режимі реального часу.

Автоматизоване виявлення загроз та реагування на них.

Комплексне розуміння вразливостей мережі.

Рішення для обробки великих даних (Big Data)

Рішення для обробки великих даних призначені для ефективної обробки великих обсягів даних. Ці інструменти необхідні для організацій, які генерують величезні обсяги даних, дозволяючи їм отримувати значущу інформацію та оптимізувати процеси.

Провідні рішення:

Apache Hadoop:це фреймворк з відкритим вихідним кодом для розподіленого зберігання та обробки великих наборів даних. Він використовує модель програмування MapReduce і широко застосовується для пакетної обробки в аналітиці великих даних.

Apache Spark:це швидкий і універсальний движок для обробки великих даних. Він підтримує обробку даних в пам'яті, що робить його значно швидшим за Hadoop для ітеративних задач, таких як машинне навчання та аналітика в реальному часі.

Google BigQuery:це повністю кероване, безсерверне рішення для сховища даних, що надається Google Cloud. Воно дозволяє організаціям за лічені секунди виконувати SQL-запити до величезних наборів даних, забезпечуючи аналіз у реальному часі.

Amazon EMR:це хмарна платформа великих даних, яка обробляє великі обсяги даних за допомогою Hadoop, Spark та інших фреймворків. Використовується для аналізу логів, трансформації даних і машинного навчання.

Переваги:

Масштабованість для роботи з терабайтами та петабайтами даних.

Швидша обробка даних завдяки паралельним обчисленням.

Економічно ефективні рішення для зберігання та аналізу даних.

Рішення для візуалізації даних

Інструменти візуалізації даних відіграють важливу роль у спрощенні складних даних та представленні їх у зрозумілому форматі. Ці інструменти допомагають особам, які приймають рішення, виявляти тенденції, закономірності та відхилення в наборах даних, що призводить до прийняття більш обґрунтованих рішень.

Провідні рішення:

Tableau: одна з найпоширеніших платформ для візуалізації даних. Вона надає інтерактивні дашборди, діаграми та графіки, що полегшують вивчення та розуміння

даних. Tableau підтримує інтеграцію з широким спектром джерел даних, включаючи бази даних, електронні таблиці та хмарні сервіси.

Power BI: це інструмент бізнес-аналітики, який пропонує потужні можливості візуалізації. Він дозволяє користувачам створювати та ділитися інтерактивними звітами та інформаційними панелями, а також підтримує потокову передачу даних у реальному часі.

D3.js: це бібліотека JavaScript, яка дозволяє розробникам створювати власні динамічні візуалізації даних для веб-додатків. Вона забезпечує гнучкість і контроль над способом представлення даних, що робить її ідеальною для унікальних потреб у візуалізації.

Qlik Sense: це платформа самообслуговування для аналізу даних, яка надає аналітику на основі штучного інтелекту та розширені можливості візуалізації. Вона дозволяє користувачам створювати кастомізовані дашборди та проводити глибоке дослідження даних.

Переваги:

Сприяє кращому розумінню складних даних.

Покращує комунікацію завдяки візуальному сторітелінгу.

Надає особам, які приймають рішення, дієві інсайти.

Рішення для машинного навчання та штучного інтелекту

Рішення для машинного навчання (ML) та штучного інтелекту (AI) трансформують галузі, автоматизуючи процеси, прогнозуючи результати та виявляючи закономірності, які традиційні методи можуть не помітити. Ці інструменти легко інтегруються з робочими процесами аналізу даних для підвищення ефективності та інтелекту.

Провідні рішення:

TensorFlow: розроблений Google, - це фреймворк з відкритим вихідним кодом для побудови та розгортання моделей машинного навчання. Він підтримує глибоке навчання, нейронні мережі та складні завдання штучного інтелекту, що робить його універсальним інструментом для предиктивної аналітики.

PyTorch: розроблений Facebook, є ще одним популярним фреймворком для машинного навчання та глибокого навчання. Він відомий своїм динамічним графіком обчислень, який дозволяє розробникам будувати моделі більш інтуїтивно.

H2O.ai: надає корпоративні рішення з відкритим вихідним кодом для машинного навчання та штучного інтелекту. Він включає інструменти для автоматичного машинного навчання (AutoML) і підтримує різні алгоритми класифікації, регресії та кластеризації.

Amazon SageMaker: це повністю керований сервіс машинного навчання від AWS, який дозволяє розробникам створювати, навчати та розгортати моделі ML. Він спрощує робочий процес ML та інтегрується з іншими сервісами AWS.

Google Cloud AI: пропонує попередньо навчені моделі для таких поширених завдань, як розпізнавання зображень, мовний переклад і перетворення мови в текст. Він також надає інструменти для індивідуального навчання моделей.

Переваги:

Автоматизує складні завдання та прогнозування.

Підвищує ефективність і швидкість прийняття рішень.

Скорочує час і зусилля, необхідні для розробки моделей.

2.3 Постановка дослідження

У сучасну епоху цифрової трансформації організації все більше покладаються на взаємопов'язані мережеві інфраструктури для забезпечення безперебійної комунікації, управління операціями та зберігання конфіденційної інформації.

Однак ця залежність принесла з собою сплеск загроз кібербезпеці - від витоків даних і атак шкідливих програм до сучасних постійних загроз (APT) і програм-вимагачів. Мережеве обладнання - маршрутизатори, комутатори, брандмауери та точки доступу - є основою цих інфраструктур і критичною точкою вразливості. Кібератаки часто використовують неконтрольовані або погано налаштовані мережеві пристрої, що призводить до катастрофічних наслідків, включаючи втрату даних, фінансові збитки та шкоду репутації.

Ключовий виклик полягає в забезпеченні безперервного моніторингу та захисту мережевого обладнання при одночасному управлінні зростаючим обсягом і складністю даних, що генеруються цими пристроями. Сучасні мережі генерують величезні обсяги трафіку, журналів і даних про конфігурацію в режимі реального часу, які вимагають передових аналітичних методів для виявлення потенційних загроз і аномалій. Традиційні методи моніторингу та аналізу мереж часто виявляються недостатніми, оскільки вони покладаються на реактивні підходи та ручне втручання, що робить їх неадекватними для виявлення та пом'якшення складних кіберзагроз, що виникають в режимі реального часу.

Це дослідження спрямоване на нагальну потребу в автоматизованій, інтелектуальній системі аналізу мережевого обладнання для посилення кібербезпеки в офісних середовищах. Існуючі рішення, такі як системи виявлення вторгнень (IDS) та інструменти управління інформацією та подіями безпеки (SIEM), часто зосереджені на аналізі мережевого трафіку на рівні додатків або користувачів. Однак вони часто не враховують більш глибоке розуміння конфігурацій, продуктивності та поведінки самого мережевого обладнання. Відсутність аналізу в режимі реального часу та проактивного виявлення загроз на рівні обладнання створює значні прогалини в системі безпеки організації.

Проблема дослідження ще більше ускладнюється зростаючою складністю сучасних мережевих середовищ. Впровадження пристроїв Інтернету речей (IoT), віддаленої роботи та хмарних обчислень розширило сферу атак, зробивши традиційні заходи безпеки, орієнтовані на периметр, недостатніми. Це зумовлює необхідність розробки передових методів, таких як класифікація на основі машинного навчання, асоціативний аналіз та кластеризація, для обробки великих потоків даних, що генеруються мережевими пристроями, виявлення аномалій та виявлення закономірностей, що вказують на потенційні загрози.

Таким чином, проблема дослідження полягає в необхідності розробки та впровадження системи, здатної:

Моніторинг та аналіз даних мережевого обладнання в режимі реального часу.

Виявлення вразливостей та потенційних загроз з високою точністю.

Надавати дієві висновки для підвищення безпеки та стабільності мережевих інфраструктур офісів.

Дослідження спрямоване на подолання розриву між традиційними методами забезпечення безпеки та зростаючими вимогами сучасної кібербезпеки шляхом впровадження комплексного підходу до аналізу мережевого обладнання, заснованого на даних. Це не лише посилює стійкість офісних середовищ до кібератак, але й сприяє розширенню сфери проактивних рішень з кібербезпеки.

3 ТЕОРЕТИЧНІ ОСНОВИ МЕТОДІВ АНАЛІЗУ ДАНИХ

3.1 Метод 1-Rule та його застосування

Метод 1-Rule, або OneR, - це простий алгоритм класифікації, який широко використовується в інтелектуальному аналізі даних і машинному навчанні. Його основна функція полягає у створенні простого правила на основі одного атрибута, яке найкраще прогнозує цільову змінну. Ця простота забезпечує швидку реалізацію, легкість інтерпретації та ефективне використання обчислювальних ресурсів. У контексті кібербезпеки цей метод має практичне застосування для моніторингу, виявлення аномалій та прийняття рішень у реальному часі.

По суті, метод 1-Rule оцінює всі атрибути в наборі даних, щоб визначити той, який найкраще прогнозує цільову змінну. Для кожної ознаки він створює категорії або біни, пов'язує кожну з них з імовірним цільовим значенням і обчислює частоту помилкової класифікації. Вибирається атрибут з найнижчим рівнем помилок, і для класифікації будується єдине правило. В результаті виходить модель, яка є інтуїтивно зрозумілою і легкою в обчислювальному плані.

У сфері кібербезпеки метод 1-Rule особливо корисний для виявлення патернів або поведінки, які тісно пов'язані з інцидентами безпеки. Наприклад, він може класифікувати мережевий трафік як нормальний або зловмисний на основі таких атрибутів, як розмір пакета або тривалість з'єднання. У цьому контексті метод може швидко виявляти аномалії або загрози, застосовуючи прості правила, такі як «Якщо розмір пакета > 500 байт, то класифікувати як зловмисний».Однією з ключових переваг методу 1-Rule в кібербезпеці є його інтерпретованість. Команди безпеки часто потребують прозорих моделей, які дають змогу діяти. Єдине правило, створене за допомогою цього методу, легко зрозуміти та інтегрувати в існуючі робочі процеси моніторингу. Крім того, його легка природа забезпечує ефективну роботу в середовищах, де швидке прийняття рішень є критично важливим, наприклад, в системах виявлення вторгнень або моніторингу кінцевих точок.

Незважаючи на свої переваги, метод 1-Rule не позбавлений обмежень. Він покладається на один атрибут і може не враховувати складні взаємодії між кількома ознаками, які часто мають вирішальне значення в умовах складних кіберзагроз. Крім того, статичні правила можуть потребувати регулярного оновлення, щоб адаптуватися до нових моделей атак. Однак ці проблеми можна вирішити, використовуючи метод як базовий інструмент або інтегруючи його з більш досконалими моделями для комплексного виявлення загроз.Таким чином, метод 1-Rule забезпечує ефективний і доступний підхід до вирішення проблем кібербезпеки. Його здатність швидко класифікувати події та виявляти аномалії робить його цінним інструментом для покращення моніторингу в режимі реального часу та прийняття рішень. Хоча він може не охоплювати всю складність сучасних загроз, його простота і швидкість гарантують, що він залишається практичним вибором для багатьох застосувань у сфері кібербезпеки.

3.2Пошук асоціативних правил у великих масивах даних

Процес пошуку асоціативних правил у великих масивах даних є фундаментальним завданням інтелектуального аналізу даних. Він передбачає виявлення значущих взаємозв'язків і закономірностей між елементами або змінними у великих, складних наборах даних. Ці закономірності часто виражаються у вигляді правил «якщо-тоді», які показують, як певні елементи зустрічаються або взаємодіють. Наприклад, правило може бути таким: «Якщо покупець купує хліб, то він, швидше за все, купить і масло». Цей підхід широко застосовується в таких сферах, як аналіз ринкового кошика, виявлення шахрайства, охорона здоров'я та мережева безпека.

Процес починається з виявлення частих наборів елементів - груп елементів, які часто зустрічаються разом у даних. Для цього зазвичай використовують такі алгоритми, як Apriori, FP-Growth та ECLAT. Після того, як визначено часті набори елементів, генеруються асоціативні правила шляхом оцінки взаємозв'язків між елементами в цих наборах. Для кількісної оцінки сили та надійності цих правил використовуються такі показники, як підтримка, впевненість і підйом:

Підтримка вимірює, як часто набір елементів з'являється в наборі даних.

Впевненість оцінює ймовірність того, що наслідок відбудеться, враховуючи попередні події.

Lift визначає силу правила порівняно з випадковістю, причому значення, більші за 1, вказують на значущі асоціації.

Одним з найпоширеніших алгоритмів пошуку асоціативних правил є алгоритм Apriori. Цей алгоритм працює шляхом ітеративного розширення наборів елементів і відсікання тих, які не відповідають мінімальному порогу підтримки. Хоча він ефективний, він може бути обчислювально дорогим,

особливо для великих наборів даних з великою кількістю елементів. Алгоритм FP-Growth пропонує ефективнішу альтернативу, стискаючи набір даних у деревоподібну структуру, що зменшує потребу у вичерпній генерації кандидатів.

У контексті великих наборів даних видобуток асоціативних правил стикається з унікальними проблемами. Розмір і складність даних можуть призвести до проблем з масштабуванням, коли алгоритми намагаються обробити величезну кількість транзакцій або змінних. Крім того, розрідженість даних може ускладнити виявлення значущих закономірностей без переважної кількості правил з низькою підтримкою. Сучасні рішення часто включають фреймворки паралельних і розподілених обчислень, такі як Hadoop і Apache Spark, щоб вирішити ці проблеми. Ці технології дозволяють видобувати асоціативні правила в розподілених системах, що робить можливим аналіз терабайт даних.

Інструменти візуалізації відіграють важливу роль в інтерпретації результатів видобутку правил асоціацій. Такі методи, як діаграми розсіювання та гістограми, допомагають проілюструвати взаємозв'язок між підтримкою, довірою та підйомом, полегшуючи аналітикам визначення пріоритетності правил для прийняття рішень.

Таким чином, пошук асоціативних правил у великих масивах даних є потужним методом для виявлення прихованих закономірностей і взаємозв'язків. Незважаючи на такі проблеми, як масштабованість і розрідженість даних, розвиток алгоритмів і розподілених обчислень зробив цей процес більш ефективним. Завдяки кращому розумінню та прогнозуванню поведінки, зумовленої даними, видобуток асоціативних правил залишається наріжним каменем аналізу даних у різних галузях, надаючи цінну інформацію та покращуючи процеси прийняття рішень.

Пошук асоціативних правил у великих масивах даних

Процес пошуку асоціативних правил у великих масивах даних є фундаментальним завданням інтелектуального аналізу даних. Він передбачає виявлення значущих взаємозв'язків і закономірностей між елементами або змінними у великих, складних наборах даних. Ці закономірності часто виражаються у вигляді правил «якщо-тоді», які показують, як певні елементи зустрічаються або взаємодіють. Наприклад, правило може бути таким: «Якщо покупець купує хліб, то він, швидше за все, купить і масло». Цей підхід широко застосовується в таких сферах, як аналіз ринкового кошика, виявлення шахрайства, охорона здоров'я та мережева безпека.

Процес починається з виявлення частих наборів елементів - груп елементів, які часто зустрічаються разом у даних. Для цього зазвичай використовують такі алгоритми, як Apriori, FP-Growth та ECLAT. Після того,

як визначено часті набори елементів, генеруються асоціативні правила шляхом оцінки взаємозв'язків між елементами в цих наборах. Для кількісної оцінки сили та надійності цих правил використовуються такі показники, як підтримка, впевненість і підйом:

Підтримка вимірює, як часто набір елементів з'являється в наборі даних.

Впевненість оцінює ймовірність того, що наслідок відбудеться, враховуючи попередні події.

Lift визначає силу правила порівняно з випадковістю, причому значення, більші за 1, вказують на значущі асоціації.

Одним з найпоширеніших алгоритмів пошуку асоціативних правил є алгоритм Apriori. Цей алгоритм працює шляхом ітеративного розширення наборів елементів і відсікання тих, які не відповідають мінімальному порогу підтримки. Хоча він ефективний, він може бути обчислювально дорогим, особливо для великих наборів даних з великою кількістю елементів. Алгоритм FP-Growth пропонує ефективнішу альтернативу, стискаючи набір даних у деревоподібну структуру, що зменшує потребу у вичерпній генерації кандидатів.

У контексті великих наборів даних видобуток асоціативних правил стикається з унікальними проблемами. Розмір і складність даних можуть призвести до проблем з масштабуванням, коли алгоритми намагаються обробити величезну кількість транзакцій або змінних. Крім того, розрідженість даних може ускладнити виявлення значущих закономірностей без переважної кількості правил з низькою підтримкою. Сучасні рішення часто включають фреймворки паралельних і розподілених обчислень, такі як Hadoop і Apache Spark, щоб вирішити ці проблеми. Ці технології дозволяють видобувати асоціативні правила в розподілених системах, що робить можливим аналіз терабайт даних.

Інструменти візуалізації відіграють важливу роль в інтерпретації результатів видобутку правил асоціацій. Такі методи, як діаграми розсіювання та гістограми, допомагають проілюструвати взаємозв'язок між підтримкою, довірою та підйомом, полегшуючи аналітикам визначення пріоритетності правил для прийняття рішень.

Таким чином, пошук асоціативних правил у великих масивах даних є потужним методом для виявлення прихованих закономірностей і взаємозв'язків. Незважаючи на такі проблеми, як масштабованість і розрідженість даних, розвиток алгоритмів і розподілених обчислень зробив цей процес більш ефективним. Завдяки кращому розумінню та прогнозуванню поведінки, зумовленої даними, видобуток асоціативних правил залишається

наріжним каменем аналізу даних у різних галузях, надаючи цінну інформацію та покращуючи процеси прийняття рішень.

3.3 Кластерний аналіз для сегментації даних

Кластерний аналіз - це метод інтелектуального аналізу даних, який використовується для групування об'єктів у кластери на основі їхньої схожості. Цей підхід необхідний для сегментації даних, виявлення прихованих закономірностей і спрощення складних наборів даних. Він особливо цінний у сфері кібербезпеки, де швидкий і ефективний аналіз великих обсягів даних має вирішальне значення для виявлення загроз і підвищення безпеки системи.

Кластерний аналіз вимірює схожість або відстань між точками даних та організовує їх в окремі групи. Ці кластери часто виявляють закономірності і взаємозв'язки, які можуть бути не відразу помітні. Використовуються різні алгоритми, такі як K-Means, DBSCAN та ієрархічна кластеризація, кожен з яких підходить для різних типів даних і цілей сегментації. Наприклад, K-Means розбиває дані на заздалегідь визначену кількість кластерів, мінімізуючи дисперсію всередині груп, тоді як DBSCAN ідентифікує кластери на основі щільності даних, що робить його стійким до викидів.

У кібербезпеці кластерний аналіз особливо корисний для виявлення аномалій і сегментації мережевого трафіку. Кластеризуючи мережеву активність, команди безпеки можуть розрізняти нормальні та підозрілі патерни. Наприклад, мережевий трафік може природним чином об'єднуватися в групи, такі як електронна пошта, передача файлів або потокове відео. Будь-яка активність, що не вписується в ці кластери, може бути позначена як аномальна, наприклад, раптовий сплеск трафіку з несподіваного джерела або незвичний розмір пакетів.

Цей метод також застосовується для виявлення інсайдерських загроз і скомпрометованих акаунтів. Поведінка користувачів, наприклад, час входу в систему, частота доступу та використання ресурсів, може бути згрупована для встановлення базових шаблонів. Відхилення в цих кластерах можуть вказувати на несанкціоновану активність або потенційні порушення безпеки. Аналогічно, в моніторингу кінцевих точок кластеризація може виявити пристрої, що демонструють аномальне використання процесора або мережеву поведінку. Кластерний аналіз приносить користь кібербезпеці, надаючи спосіб аналізу даних без необхідності використання маркованих наборів даних, що робить його придатним для середовищ, де часто виникають нові та неочікувані загрози. Його здатність швидко обробляти великі масиви даних добре узгоджується з потребами сучасних систем кібербезпеки, де обсяг і швидкість даних можуть перевершити традиційні методи. Крім того, його гнучкість дозволяє обробляти як структуровані, так і неструктуровані дані, уможливлюючи ширше застосування в різних сферах кібербезпеки.

Незважаючи на його переваги, виклики залишаються. Дані високої розмірності, такі як мережеві журнали, можуть ускладнювати кластеризацію, а шум або викиди можуть спотворювати результати. Такі алгоритми, як DBSCAN, вирішують деякі з цих проблем, обробляючи шум окремо, але пошук правильних параметрів або адаптація моделей до еволюції загроз часто вимагає додаткових зусиль.

Кластерний аналіз продовжує розвиватися, а досягнення в галузі штучного інтелекту і розподілених обчислень дозволяють створювати більш складні і масштабовані рішення. Його інтеграція в системи, що працюють в режимі реального часу, відкриває потенціал для проактивного виявлення загроз і посиленого моніторингу системи.Використовуючи ці можливості, фахівці з кібербезпеки можуть краще сегментувати і аналізувати дані, що в кінцевому підсумку посилює захист від динамічного і складного ландшафту загроз.

РОЗРОБКА ТА ОПТИМІЗАЦІЯ МОДЕЛЕЙ ДЛЯ АНАЛІЗУ ДАНИХ

4.1 Створення бази даних

Початково було створено топологію системи для розуміння як саме створити базу даних

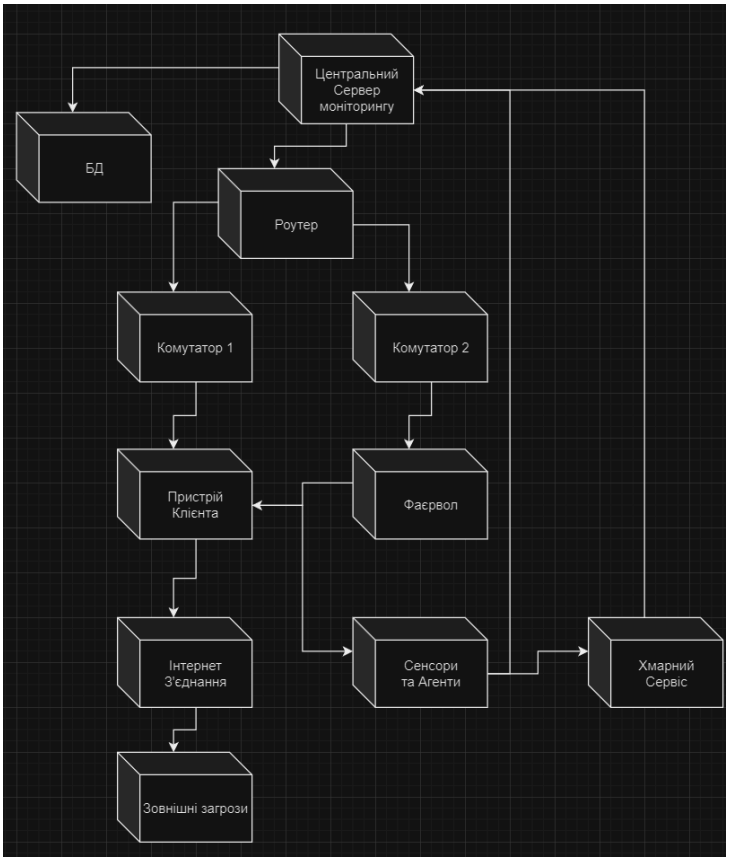

Рис. 1 Топологія системи

Метою створення бази даних було створення структурованого централізованого сховища даних, здатного забезпечити збір, зберігання та аналіз даних для кібербезпеки в офісних мережах. База даних підтримує такі завдання, як виявлення аномалій, моніторинг продуктивності та звітування про інциденти, організовуючи дані з різних мережевих пристроїв в уніфікованому форматі. Ці структуровані дані слугують основою для процесів інтелектуального аналізу даних та розширеної аналітики.

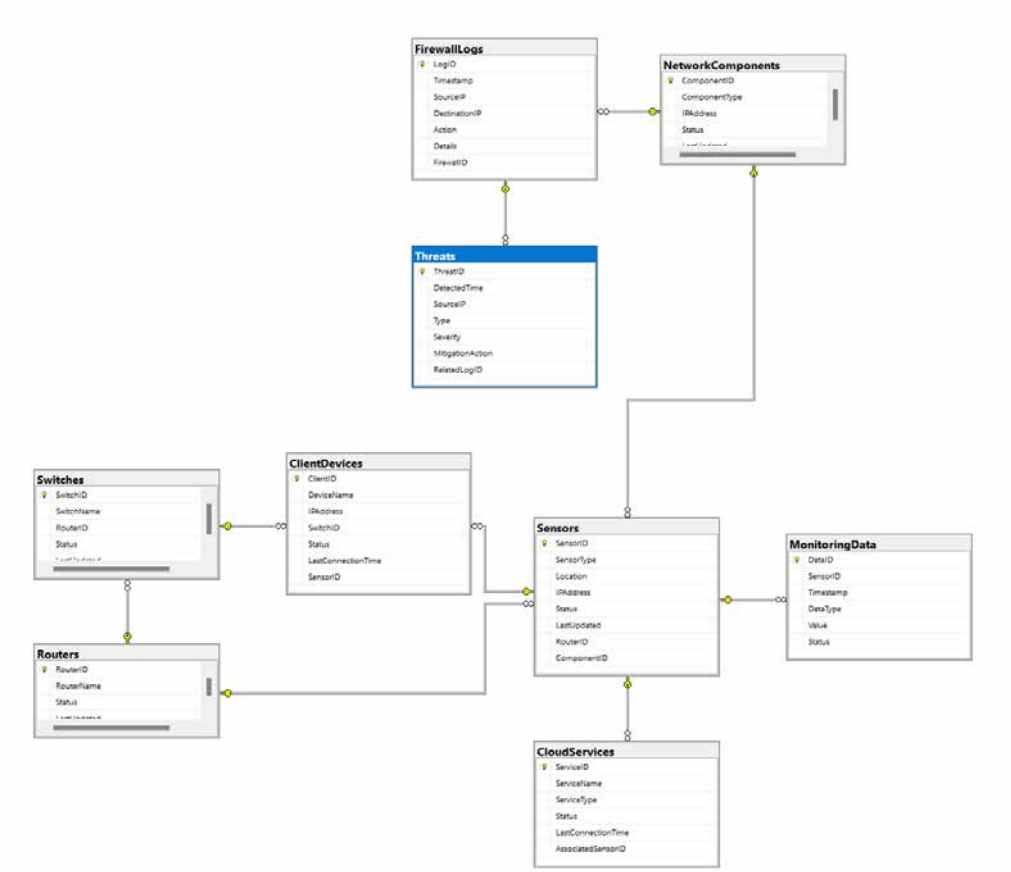

Рис. 2 Оперативна база даних

Сфера застосування бази даних

База даних була розроблена для агрегування даних з різних джерел, таких як мережеві пристрої (маршрутизатори, комутатори, брандмауери) та журнали систем кібербезпеки. Вона підтримує інтеграцію з аналітичними інструментами та інструментами звітності для отримання корисної інформації. Архітектура бази даних забезпечує масштабованість, надійність та ефективні запити.

Основні функції включають

Централізоване зберігання даних про всі події, пов'язані з мережею.

Підтримка аналізу даних у реальному часі та історичних даних.

Нормалізація даних для зменшення надмірності та покращення узгодженості.

Процес розробки бази даних

1 Аналіз вимог

Перший крок включав визначення ключових джерел даних та вимог користувачів, які будуть взаємодіяти з базою даних. Основна увага була зосереджена на розумінні:

Типи необхідних даних (наприклад, журнали пристроїв, активність користувачів, виявлені загрози).

Взаємозв'язків між сутностями (наприклад, пристрій генерує журнали, журнали містять часові мітки і типи подій).

Варіанти використання для запитів і звітності, такі як моніторинг у реальному часі, перегляд історії подій і аналіз тенденцій.

2 Моделювання даних

Для візуалізації логічної структури бази даних була розроблена модель «сутність-зв'язок» (ER). Ця модель допомогла визначити взаємозв'язки між об'єктами даних і забезпечити охоплення всіх ключових аспектів системи моніторингу мережі та кібербезпеки.

3 Розробка схеми

Використовуючи ER-модель як орієнтир, схема була розроблена для забезпечення

Логічну узгодженість між об'єктами даних.

Нормалізацію для усунення дублікатів даних та покращення цілісності.

Оптимізацію для частих запитів шляхом індексування ключових атрибутів.

Схему було реалізовано на MySQL, обраній за її надійність, масштабованість та підтримку складних запитів. Це включало створення таблиць, визначення обмежень та реалізацію первинних і зовнішніх ключів для підтримки посилальної цілісності.

4 Інтеграція даних

Дані з мережевих пристроїв були інтегровані в базу даних за допомогою ручного імпорту. Були розроблені автоматизовані процеси:

Регулярного оновлення стану пристроїв та журналів.

Перетворення необроблених даних у стандартизований формат, придатний для аналізу.

Обробка помилок і перевірка для забезпечення якості даних.

5 Тестування та оптимізація

Для перевірки було проведено комплексне тестування:

Точність та узгодженість даних.

Продуктивність запитів під різними навантаженнями.

Сумісність з аналітичними інструментами для інтелектуального аналізу та візуалізації даних.

Методи оптимізації включали індексування, налаштування запитів і розбиття на розділи для великих наборів даних, щоб гарантувати, що база даних може обробляти великі обсяги даних у реальному часі без зниження продуктивності.

6 Безпека даних та резервне копіювання

Враховуючи чутливість даних з кібербезпеки, було впроваджено численні заходи безпеки:

Аутентифікація користувачів та контроль доступу на основі ролей (RBAC) для обмеження доступу.

Шифрування даних для захисту конфіденційної інформації.

Регулярне резервне копіювання для запобігання втрати даних.

7 Документація та навчання користувачів

Було підготовлено детальний документ, що включає

Схеми для пояснення структури бази даних.

Шаблони запитів для поширених випадків використання.

Вказівки щодо інтеграції нових джерел даних та управління оновленнями.

8 Майбутній розвиток

База даних структурована таким чином, щоб підтримувати майбутні вдосконалення, такі як:

1)Інтеграція з алгоритмами інтелектуального аналізу даних для виявлення аномалій у реальному часі.

2)Розробка куба даних для OLAP (Online Analytical Processing) для забезпечення багатовимірного аналізу.

3)Розширення джерел даних по мірі додавання нових мережевих пристроїв до інфраструктури.

Ця база даних формує основу системи кібербезпеки, організовуючи, зберігаючи та дозволяючи аналізувати критичні мережеві дані. Її дизайн забезпечує масштабованість, надійність та інтеграцію з аналітичними інструментами, надаючи надійне рішення для моніторингу та захисту офісних мереж.

4.2 Вибір методу та обґрунтування вибору для кожного з підходів

В0438бір відповідних методів для аналізу даних має вирішальне значення для досягнення точних, дієвих та інтерпретованих результатів. У цьому дослідженні я обрав комбінацію методу 1-го правила, наївного алгоритму Байєса, асоціативного видобування правил та кластерного аналізу. Кожен підхід був обраний на основі його сильних сторін та відповідності конкретним завданням, що дозволило провести всебічний аналіз набору даних для класифікації, виявлення взаємозв'язків та ідентифікації закономірностей.

Метод 1-Rule (OneR) було обрано через його простоту та ефективність у задачах класифікації. Цей метод оцінює всі доступні атрибути в наборі даних, вибирає один з них з найбільшою прогностичною силою і будує просте правило для класифікації точок даних. Він особливо добре підходить для сценаріїв, де важлива інтерпретованість, оскільки модель створює прості правила «якщо-тоді», які легко зрозуміти і застосувати.

1-Rule було застосовано для класифікації станів датчиків, таких як «Нормальний» або «Тривога», на основі таких атрибутів, як температура. Цей метод було обрано, оскільки набір даних містив атрибути з високою прогностичною значущістю, що робить класифікатор з одним правилом практичним вибором. Його обчислювальна ефективність також робить його ідеальним для систем моніторингу в реальному часі, де необхідні швидкі рішення. Метод 1-го правила забезпечує базову модель, яку можна легко

донести до нетехнічних зацікавлених сторін, забезпечуючи прозорість у процесах прийняття рішень.

Наївний алгоритм Байєса був обраний за його надійну ймовірнісну основу та ефективність в обробці як категоріальних, так і числових даних. Він особливо ефективний для задач класифікації, що включають декілька атрибутів, навіть у випадках, коли ознаки не є повністю незалежними. Алгоритм є обчислювально легким, що робить його придатним для обробки великих наборів даних або роботи в умовах обмежених ресурсів.

Він використовувався для класифікації даних з датчиків на різні категорії шляхом аналізу декількох ознак, таких як температура, мережевий трафік і рух. Цей метод був обраний через його здатність швидко моделювати складні взаємозв'язки між ознаками та прогнозувати результати з високою точністю. Крім того, Naive Bayes добре працює з незбалансованими наборами даних, що є поширеною проблемою при виявленні аномалій, коли нормальні події значно переважають над аномальними. Його імовірнісні результати дозволяють приймати рішення на основі довіри, що ще більше підвищує його корисність в додатках кібербезпеки.

Рішення про включення видобутку асоціативних правил ґрунтувалося на його здатності виявляти приховані зв'язки між об'єктами або подіями у великих наборах даних. Визначаючи набори елементів, що часто зустрічаються, і генеруючи правила, цей метод дає цінну інформацію про те, як різні змінні пов'язані між собою. Такі показники, як підтримка, впевненість і підйом, гарантують, що виявлені правила є значущими і такими, що піддаються інтерпретації.

Видобуток асоціативних правил використовувався для виявлення взаємозв'язків між подіями датчиків, такими як «високий мережевий трафік» і «стрибки температури», які часто передують критичним станам. Алгоритм Apriori був обраний за його ефективність у створенні правил асоціацій на основі транзакційних наборів даних. Цей підхід допоміг виявити дієві закономірності, які могли б стати основою для превентивних заходів та оптимізації продуктивності системи.

Кластерний аналіз було обрано для сегментації даних на значущі групи на основі схожості. На відміну від методів класифікації, які вимагають маркованих даних, кластеризація є неконтрольованою технікою, яка допомагає виявити приховані закономірності та структури в даних. Вона особливо цінна для виявлення аномалій і групування поведінки, яка відхиляється від очікуваних шаблонів.

У контексті цього дослідження кластерний аналіз був використаний для сегментації даних мережевого трафіку і виявлення незвичайних патернів, які

можуть свідчити про потенційні загрози безпеці. Наприклад, звичайні дії користувачів, такі як передача файлів або перегляд веб-сторінок, можуть бути згруповані в окремі кластери, тоді як відхилення можуть свідчити про зловмисні дії. Такі алгоритми, як K-Means і DBSCAN, були використані завдяки їхній ефективності та здатності обробляти шум. Цей метод мав вирішальне значення для розуміння базової структури даних і забезпечення основи для виявлення аномалій.

Кажучи про інтеграцію методів

Метод 1-Rule забезпечив просту модель, що легко інтерпретується, для швидкої класифікації.

Наївний Байєс уможливив ефективну багатоатрибутивну класифікацію з імовірнісними висновками.

Видобування асоціативних правил дозволило виявити складні взаємозв'язки та взаємозалежності в даних.

Кластерний аналіз полегшив сегментацію та виявлення аномалій, виявивши приховані закономірності.

Разом ці методи запропонували цілісний підхід до аналізу набору даних, забезпечивши всебічну

4.3 Створення моделі для методу 1-Rule

Огляд методу 1-го правила

Метод 1-го правила (OneR) - це простий, але ефективний алгоритм класифікації, який широко використовується в аналізі даних. Він будує модель з одним правилом, вибираючи атрибут, який забезпечує найбільш значну прогностичну силу для цільової змінної. На відміну від складних багатоатрибутивних підходів, OneR фокусується на створенні зрозумілих для людини правил, що робить його дуже зручним для інтерпретації. Ця простота дозволяє йому бути еталоном для оцінки більш складних моделей або слугувати полегшеним рішенням у сценаріях, де простота і прозорість є пріоритетними.

Алгоритм OneR працює, оцінюючи кожен атрибут у наборі даних і визначаючи його прогностичну ефективність за допомогою заздалегідь визначеної метрики, найчастіше - частоти помилкових класифікацій. Після того, як визначено найкращий атрибут, алгоритм генерує правила на основі його унікальних значень для прогнозування цільового класу. Ці правила потім застосовуються для класифікації нових точок даних.

Застосування методу 1-го правила в аналізі сенсорних даних

У представленому дослідженні та прикладі використання метод 1-Rule був застосований для класифікації стану датчиків (нормальний або тривожний)

на основі показань температури. Це застосування особливо актуальне для моніторингу мережі та кібербезпеки, де виявлення аномалій в режимі реального часу може запобігти збоям або порушенням системи.

Набір даних складається з трьох ключових атрибутів:

Ідентифікатор датчика: ідентифікація джерела даних.

Температура: Основний прогностичний атрибут.

Статус: Цільова змінна, що вказує на те, чи перебуває датчик у «нормальному» або «тривожному» стані.

Приклад набору даних:

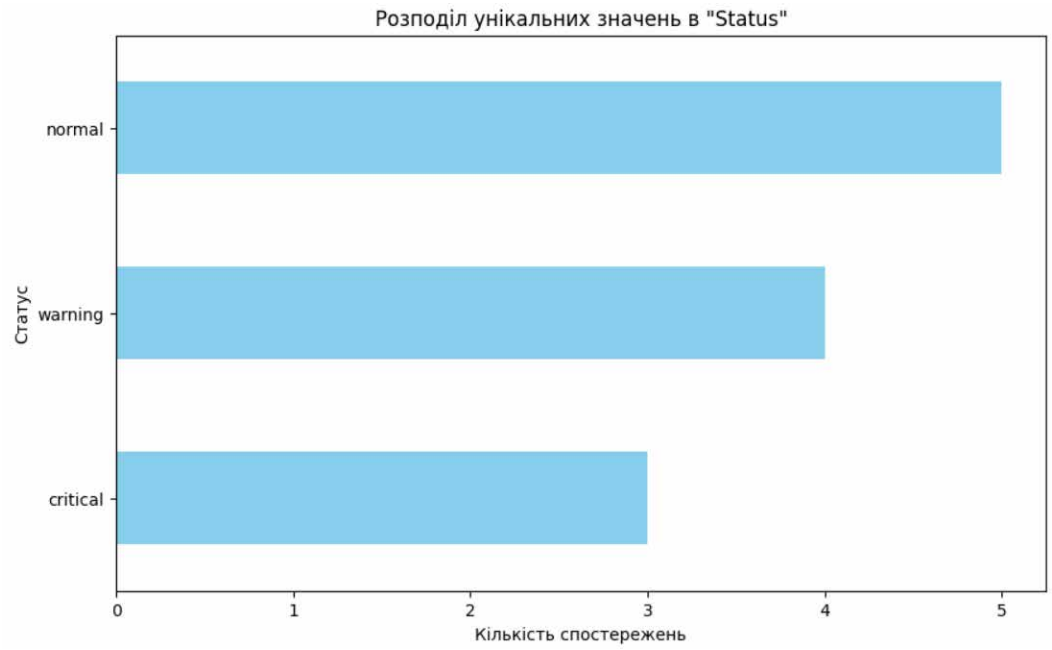

Рис. 3 Приклад моделі OneR

Модель OneR була навчена на цьому наборі даних, де вона оцінювала прогностичну здатність атрибуту «Температура». Алгоритм створив простий набір правил, наприклад

`Якщо Температура ≤ 28, то Статус = Нормальний`

Якщо температура > 28, то статус = Тривога

Ці правила були отримані шляхом вивчення того, як різні температурні діапазони відповідають цільовому статусу, і вибору діапазону з найнижчим рівнем помилок.

Використав логістичну регресію для порівняння разом з OneR.

Розділив дані на навчальну та тестову вибірки за допомогою train_test_split.

Оцінили точність моделі та порівняли прогнози з фактичними значеннями.

Алгоритм OneR успішно створив просту, зрозумілу модель класифікації. Правило, засноване на температурі, ефективно передбачило стан датчика, виділивши наступні переваги:

Простота: Підхід з одним правилом зробив модель легкою для інтерпретації та застосування.

Ефективність: Модель вимагала мінімальних обчислювальних ресурсів, що робить її придатною для застосування в режимі реального часу.

Точність: Незважаючи на простоту, модель забезпечила надійні прогнози для даного набору даних.

Для порівняння, модель логістичної регресії запропонувала дещо вищу точність, але ціною меншої інтерпретованості.

Практичні висновки

Виявлення аномалій: Виявляючи аномальні показники датчиків, система може позначити потенційні проблеми, які потребують негайної уваги.

Кібербезпека: Застосовуючись до аналізу мережевого трафіку, метод може класифікувати події як доброякісні або підозрілі на основі конкретних атрибутів, таких як розмір пакетів або частота запитів.

Обмеження і майбутні напрямки

Хоча метод 1-Rule вирізняється простотою та інтерпретованістю, він має певні обмеження:

Залежність від одного атрибута: Метод може не враховувати важливі взаємодії між кількома змінними.

Ефективність на складних наборах даних: Для наборів даних зі складними взаємозв'язками можуть знадобитися більш досконалі методи, такі як дерева рішень або ансамблеві моделі.

Майбутні дослідження можуть бути зосереджені на цьому:

Гібридні підходи: Поєднання OneR з іншими методами для підвищення продуктивності.

Динамічна генерація правил: Адаптація правил у реальному часі до мінливих умов у мережі або середовищі.

Метод 1-Rule є практичним і ефективним рішенням для побудови простих моделей класифікації. Його застосування в аналізі сенсорних даних ілюструє його здатність надавати дієві висновки в системах моніторингу в реальному часі. Незважаючи на свої обмеження, OneR залишається цінним інструментом для сценаріїв, де інтерпретованість і швидкість є критично важливими, прокладаючи шлях до прогресу в системах прийняття рішень на основі даних.

4.4 Реалізація пошуку асоціативних правил

Реалізація правил видобування асоціацій була зосереджена на виявленні взаємозв'язків і закономірностей у наборі даних про події з датчиків. Цей процес мав на меті виявити часті набори елементів - комбінації подій, які часто трапляються разом - і вивести дієві правила асоціацій для покращення управління системою та виявлення аномалій.

Першим кроком у процесі була підготовка набору даних, який містив записи показань датчиків. До нього входили ключові атрибути:

SensorID: ідентифікатор датчика, який генерує дані.

Подія: Опис типу виявленої події (наприклад, трафік, температура, рух або безпека).

Значення: Представляє записану метрику або стан.

Набір даних було перетворено в транзакційний формат, придатний для видобування асоціативних правил. Це перетворення згрупувало дані за SensorID і перетворило типи подій у двійкові значення. Події кожного датчика були представлені у вигляді набору, де 1 вказувала на появу події, а 0 - на її відсутність. Цей крок був необхідний для застосування таких алгоритмів, як Apriori, які покладаються на транзакційні дані.

Алгоритм Apriori було використано для виявлення частих наборів елементів з набору транзакційних даних. Алгоритм ітеративно аналізував дані, щоб знайти комбінації подій, які відповідали заданому мінімальному порогу підтримки (наприклад, 20%). Потім на основі цих наборів створювалися правила асоціацій. Наприклад, якщо «Рух» і «Попередження про небезпеку» часто зустрічалися на всіх датчиках, то вони утворювали часті набори елементів.

Після того, як були визначені часті набори елементів, були створені правила асоціації для дослідження зв'язків між подіями. Ці правила оцінювалися за допомогою таких метрик, як

Підтримка: Частка транзакцій, що містять набір елементів.

Впевненість: Ймовірність того, що наступна подія відбудеться, враховуючи попередню.

Підйом: Сила правила в порівнянні з випадковістю.

Наприклад, правило на кшталт «Якщо трафік високий, то спрацьовує Попередження про безпеку» було отримано, якщо спільна поява цих подій показала значні значення підтримки, впевненості та підйому.

Для більш ефективної інтерпретації результатів були створені візуалізації. Гістограма проілюструвала 10 найпоширеніших наборів елементів, показавши їхні значення підтримки та виділивши найпоширеніші комбінації подій. Діаграма розсіювання була використана для представлення правил асоціацій, відображення значень підтримки та достовірності, при цьому розмір бульбашок змінювався для відображення метрики підйому. Ці візуалізації допомогли з першого погляду визначити найважливіші закономірності та правилаРеалізація забезпечила розуміння взаємозв'язків між подіями датчиків, що дало змогу краще керувати системою та виявляти аномалії. Наприклад, часта асоціація «високого трафіку» з «стрибками температури» вказувала на потенційну проблему перегріву, що спонукало вжити превентивних заходів для зменшення ризиків. Аналогічно, правила, що виділяють незвичні комбінації, такі як «Високий трафік» і «Заборона руху», можуть вказувати на підозрілу активність, що вимагає подальшого розслідування.

Розуміючи ці асоціації, система може бути налаштована на проактивне реагування. Автоматизовані дії, такі як генерування сповіщень або запуск механізмів безпеки, були впроваджені для вирішення критично важливих проблем. Наприклад, якщо певні події постійно призводили до критичного стану системи, активувалися заздалегідь визначені дії, щоб запобігти проблемам до їх ескалації.Ця реалізація продемонструвала практичну цінність видобування асоціативних правил для виявлення закономірностей і залежностей у даних датчиків. Це дозволило підвищити ефективність моніторингу системи, покращити виявлення аномалій та прийняття обґрунтованих рішень. Однак, серед викликів були обробка зашумлених даних і забезпечення обчислювальної ефективності алгоритму Apriori, особливо з великими наборами даних. Ретельне налаштування параметрів, наприклад, встановлення відповідних порогів підтримки та довірчої ймовірності, було необхідним, щоб збалансувати релевантність правил з обчислювальними вимогами.

Пошук асоціативних правил за допомогою алгоритму Apriori надав практичні висновки, які покращили моніторинг та управління системою. Завдяки виявленню прихованих взаємозв'язків між подіями, реалізація

уможливила проактивне реагування на потенційні проблеми на основі даних. Цей процес не лише покращив загальну продуктивність системи, але й продемонстрував потужність видобування правил асоціацій у реальних додатках. У майбутній роботі можна дослідити інтеграцію потоків даних у реальному часі та оптимізацію обчислювальної продуктивності для ефективної обробки ще більших наборів даних.

4.5 Кластеризація даних для виявлення прихованих патернів

Кластеризація - це потужний метод виявлення прихованих закономірностей у даних шляхом групування схожих точок даних у кластери. Він особливо корисний у неконтрольованому навчанні, де немає заздалегідь визначених міток. У цьому дослідженні кластеризація була застосована до даних датчиків для аналізу таких показників, як температура, трафік і стан системи. Мета полягала в тому, щоб визначити значущі групи, які могли б допомогти в моніторингу системи та виявленні аномалій.

Набір даних складався з таких атрибутів, як температура, трафік і стан датчиків, що забезпечило основу для групування схожої поведінки. Перед кластеризацією дані були масштабовані, щоб переконатися, що всі характеристики в рівній мірі сприяли аналізу. Цей крок був важливим, оскільки температура і трафік працюють у різних масштабах, що може вплинути на результати кластеризації.Кластеризація за методом K-середніх була обрана завдяки своїй ефективності та простоті реалізації. Цей алгоритм розбиває дані на кластери, мінімізуючи дисперсію всередині кожної групи. Було обрано три кластери, що відображають різні робочі характеристики датчиків. Алгоритм відніс точки даних до найближчого центроїда та ітеративно скоригував центроїди, щоб досягти стабільного групування.

Кластеризація виявила три ключові групи датчиків. Одна група представляла нормальний режим роботи, що характеризується помірною температурою та рівнем трафіку. Інша включала датчики з високим трафіком, але нормальною температурою, що, ймовірно, відображає високу мережеву активність. Третя група включала датчики з екстремальними показниками, що часто корелювали з попереджувальними або критичними станами. Ці кластери надали практичну інформацію, наприклад, визначили датчики, які потребують ретельнішої перевірки через потенційні аномалії.

Діаграма розсіювання візуалізувала результати кластериазації, використовуючи кольори для представлення призначень кластерів. Це допомогло проілюструвати межі та взаємозв'язки між кластерами, що полегшило інтерпретацію групувань. Наприклад, датчики з екстремальними значеннями чітко виділялися, що дозволило швидко ідентифікувати проблеми.

Кластеризація значно покращила моніторинг системи, згрупувавши датчики зі схожою поведінкою. Такий підхід спростив ідентифікацію аномалій, наприклад, датчиків, що повідомляють про незвично високий трафік або температуру. Ці аномалії часто вказували на потенційні апаратні збої, екологічні проблеми або загрози кібербезпеці, що дозволяло вчасно втрутитися.Метод виявився ефективним і дієвим. Він виявив закономірності, які могли бути неочевидними при індивідуальному аналізі, і сприяв цілеспрямованому моніторингу. Хоча існували такі проблеми, як

шум і визначення оптимальної кількості кластерів, переваги виявлення прихованих закономірностей переважили ці обмеження. Загалом, кластеризація додала цінний рівень розуміння до управління та оптимізації системи.

5. АНАЛІЗ ТА ІНТЕРПРЕТАЦІЯ РЕЗУЛЬТАТІВ

5.1. Оцінка точності моделі для методу 1-Rule

Метод OneR був застосований до навчального набору даних для створення єдиного правила. У процесі навчання було проаналізовано взаємозв'язок між температурними діапазонами та станами датчиків, щоб визначити найбільш вірогідну класифікацію для кожного діапазону. Згенероване правило було протестовано на окремому тестовому наборі даних, щоб оцінити його прогностичну ефективність.

Ефективність і точність моделі

Модель OneR досягла 50% точності під час тестування, як показано в наданих прогнозах:

Прогнозовані статуси: ['Попередження', 'Попередження']

Фактичні статуси: ['Попередження', 'Нормально']

Це вказує на те, що хоча модель правильно класифікувала один з тестових прикладів, вона невірно класифікувала другий. Точність у 50% підкреслює простоту методу OneR та його обмеження при роботі зі складними наборами даних з декількома цільовими категоріями.

На точність моделі OneR вплинуло кілька факторів:

Залежність від одного атрибута: Залежність моделі від одного атрибуту, «Температура», обмежила її здатність враховувати інші потенційні фактори, що впливають на стан датчика. Наприклад, мережевий трафік або дані про рух також можуть відігравати важливу роль у визначенні поведінки датчика.

Розподіл даних: Гістограма, що показує розподіл унікальних значень в атрибуті «Стан», виявила дисбаланс, причому «Нормальний» є найчастішою категорією. Такий дисбаланс може зміщувати модель у бік прогнозування більшості класів, знижуючи її продуктивність для класів меншості, таких як «Критичний».

Температурні діапазони: Правило, згенероване моделлю, може не повністю враховувати нюанси взаємозв'язку між температурою та станом, особливо для діапазонів, що перекриваються, де температура біля межі може належати до кількох класів.

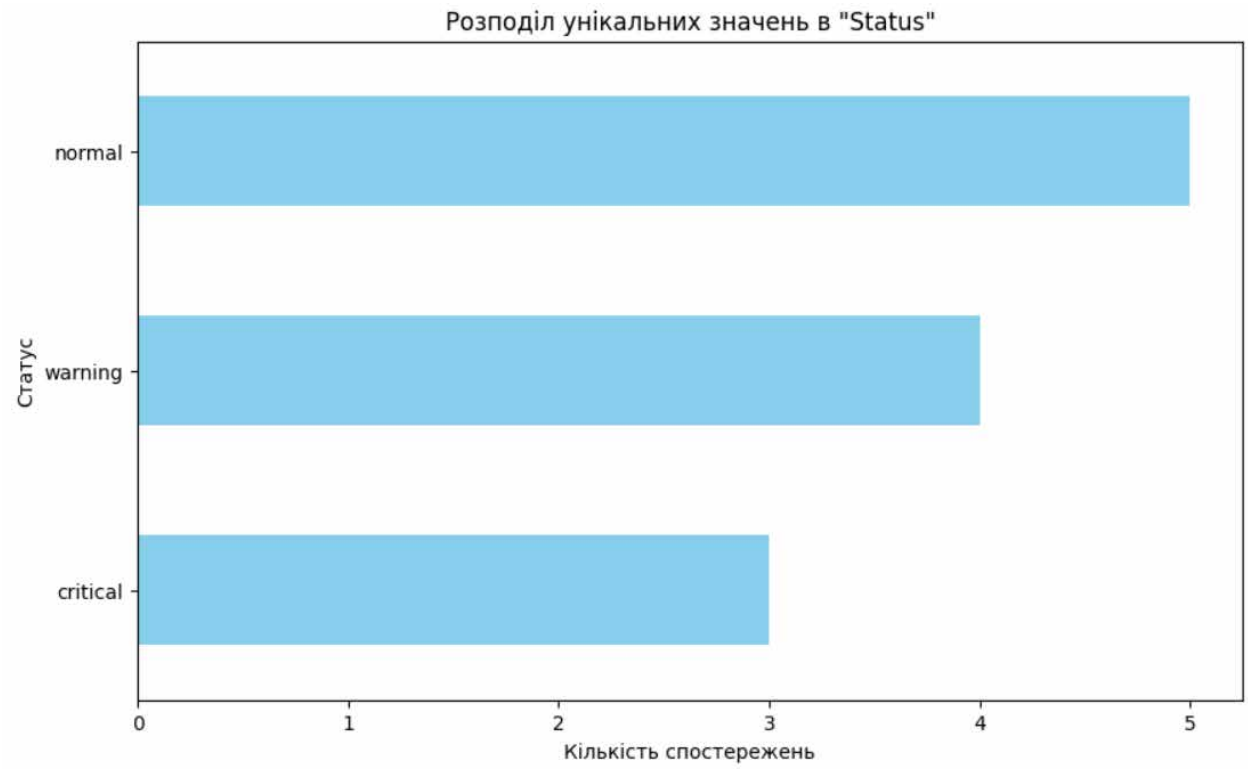

Рис. 4 Результат моделі OneR

Висновки з візуалізації

Гістограма, що візуалізує розподіл категорій «Статус», підкреслює нерівномірність представлення класів. Цей дисбаланс, ймовірно, спричинив обмежену точність моделі, оскільки вона намагалася розрізнити менш часті статуси, такі як «Критичний» і «Попередження».

Хоча метод OneR є ефективним для швидкої та зрозумілої класифікації, його результати в цьому дослідженні підкреслюють його обмеження:Він найкраще підходить для наборів даних, де один атрибут сильно прогнозує цільову змінну.Його простота робить його ідеальним для базового моделювання або випадків, коли обчислювальна ефективність є критично важливою.Для складних наборів даних з багатьма факторами впливу кращу продуктивність можуть забезпечити більш просунуті методи, такі як наївний Байєс або дерева рішень.

```
Точность модели: 0.50
Прогнозы модели:
['warning' 'warning']
Реальные значения:
['warning' 'normal']
```

Рис. 5

Модель OneR забезпечила просте, зрозуміле правило для класифікації станів датчиків, але під час тестування досягла лише помірної точності (50%). Цей результат підкреслює важливість розгляду альтернативних моделей для наборів даних зі складними патернами або незбалансованими класами. Хоча метод OneR ефективний

як відправна точка, він вимагає ретельного застосування для забезпечення надійних прогнозів.

5.2. Виявлені асоціативні правила та їх інтерпретація

Процес видобування асоціативних правил виявив кілька значущих закономірностей у наборі даних. Ці правила визначають взаємозв'язки між різними атрибутами, такими як «Рух», «Трафік», «Температура» та «Безпека». Аналізуючи ці взаємозв'язки, було отримано корисну інформацію для моніторингу системи та прийняття рішень.

Окремі елементи, такі як «Трафік», «Переміщення», «Температура» та «Безпека», мають високі значення підтримки, що вказує на їх часту появу в наборі даних.

Комбіновані набори елементів, такі як (Рух, Трафік) та (Безпека, Температура), також мали значну підтримку, що підкреслює залежність між цими атрибутами.

Найкращий набір елементів (Трафік) має підтримку 1.0, що означає, що він зустрічається у всіх транзакціях, що робить його домінуючим фактором у наборі даних.

Правила асоціації, отримані з цих наборів елементів, виявляють зв'язки типу «якщо-тоді» між подіями. Приклади включають

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Movement)	(Traffic)	0.666667	1.000000	0.666667	1.000000	1.0	0.0	inf	0.0
1	(Traffic)	(Movement)	1.000000	0.666667	0.666667	0.666667	1.0	0.0	1.0	0.0
2	(Security)	(Traffic)	0.666667	1.000000	0.666667	1.000000	1.0	0.0	inf	0.0
3	(Traffic)	(Security)	1.000000	0.666667	0.666667	0.666667	1.0	0.0	1.0	0.0
4	(Temperature)	(Traffic)	0.666667	1.000000	0.666667	1.000000	1.0	0.0	inf	0.0
5	(Traffic)	(Temperature)	1.000000	0.666667	0.666667	0.666667	1.0	0.0	1.0	0.0
6	(Movement, Security)	(Traffic)	0.333333	1.000000	0.333333	1.000000	1.0	0.0	inf	0.0
7	(Traffic)	(Movement, Security)	1.000000	0.333333	0.333333	0.333333	1.0	0.0	1.0	0.0
8	(Movement, Temperature)	(Traffic)	0.333333	1.000000	0.333333	1.000000	1.0	0.0	inf	0.0
9	(Traffic)	(Movement, Temperature)	1.000000	0.333333	0.333333	0.333333	1.0	0.0	1.0	0.0
10	(Security, Temperature)	(Traffic)	0.333333	1.000000	0.333333	1.000000	1.0	0.0	inf	0.0
11	(Traffic)	(Security, Temperature)	1.000000	0.333333	0.333333	0.333333	1.0	0.0	1.0	0.0

Рис.6 Результат алгоритму асоціативного аналізу

Якщо рух, то трафік (Підтримка: 0.6667, Впевненість: 1.0, Підйом: 1.0)

Якщо безпека, то трафік (підтримка: 0.6667, довіра: 1.0, підйом: 1.0)

Якщо Температура, то Трафік (Підтримка: 0.6667, Довіра: 1.0, Підйом: 1.0)

Якщо трафік, то рух і температура (підтримка: 0.3333, довіра: 1.0, підйом: 1.0)

	support	itemsets
0	0.666667	(Movement)
1	0.666667	(Security)
2	0.666667	(Temperature)
3	1.000000	(Traffic)
4	0.333333	(Movement, Security)
5	0.333333	(Movement, Temperature)
6	0.666667	(Movement, Traffic)
7	0.333333	(Security, Temperature)
8	0.666667	(Security, Traffic)
9	0.666667	(Temperature, Traffic)
10	0.333333	(Movement, Security, Traffic)
11	0.333333	(Movement, Temperature, Traffic)
12	0.333333	(Security, Temperature, Traffic)

Рис. 7 Результатом алгоритму частих наборів елементів

Ці правила припускають, що певні комбінації подій тісно пов'язані між собою і можуть сигналізувати про певну поведінку системи.

Інтерпретація метрик

Підтримка вимірює, як часто набір елементів або правило з'являється в наборі даних. Правила з високою підтримкою, такі як (Трафік) або (Рух → Трафік), вказують на закономірності, що постійно повторюються.

Достовірність відображає ймовірність того, що наслідок відбудеться, враховуючи попередню подію. Правила з достовірністю 1.0, такі як (Безпека → Трафік), є надійними предикторами в наборі даних.

Підйомний коефіцієнт оцінює силу правила поза випадковістю. Хоча в цьому аналізі всі правила мали підйомний коефіцієнт 1.0, це означає, що ймовірність настання наслідку не вища, ніж очікується за базовою ймовірністю.

Аналіз діаграми розсіювання

Діаграми розсіювання підтримки та довірчої ймовірності візуально виділяють найбільш значущі правила. Правила з більш високими значеннями підтримки і достовірності, такі як

(Трафік → Безпека, Температура), виділяються як критичні патерни, що заслуговують на увагу під час моніторингу системи.

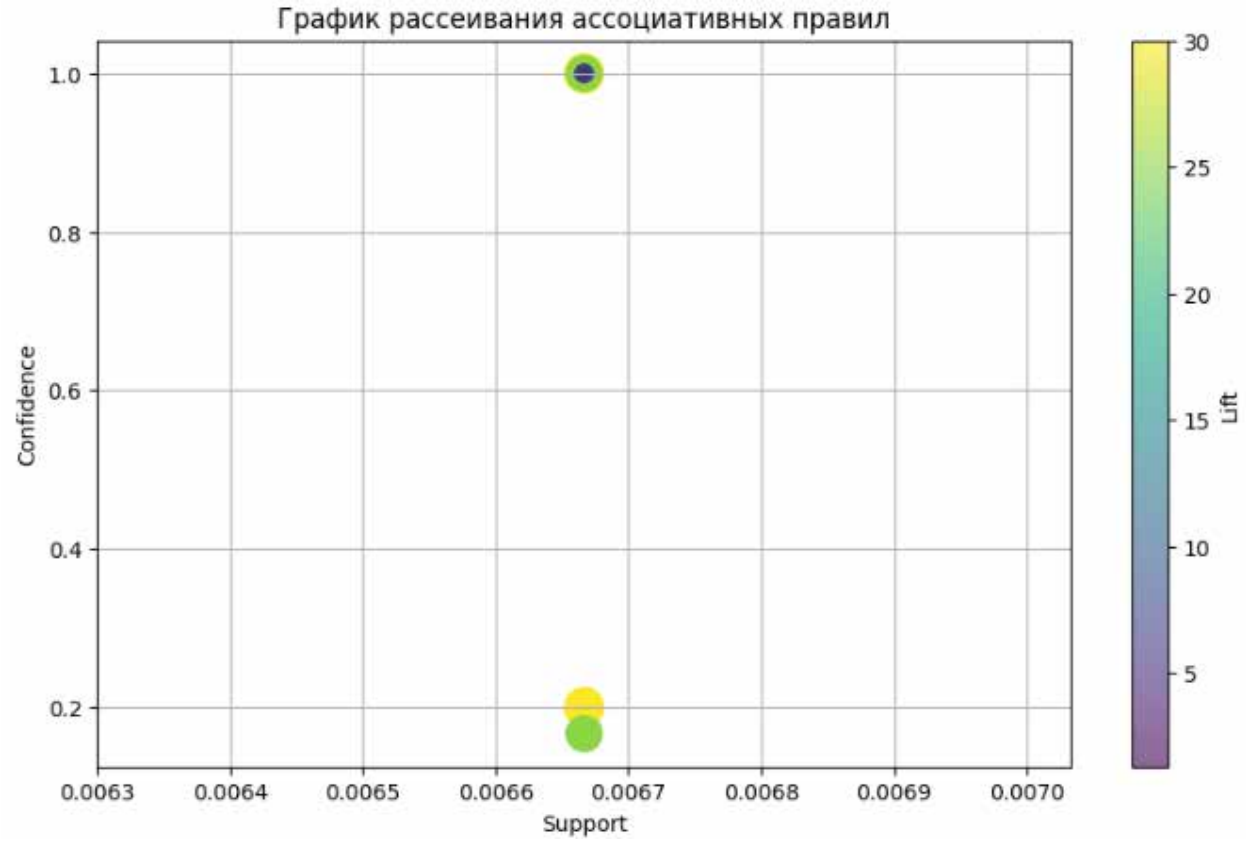

Рис. 8 Графік розсіювання (scatter plot) асоціативних правил

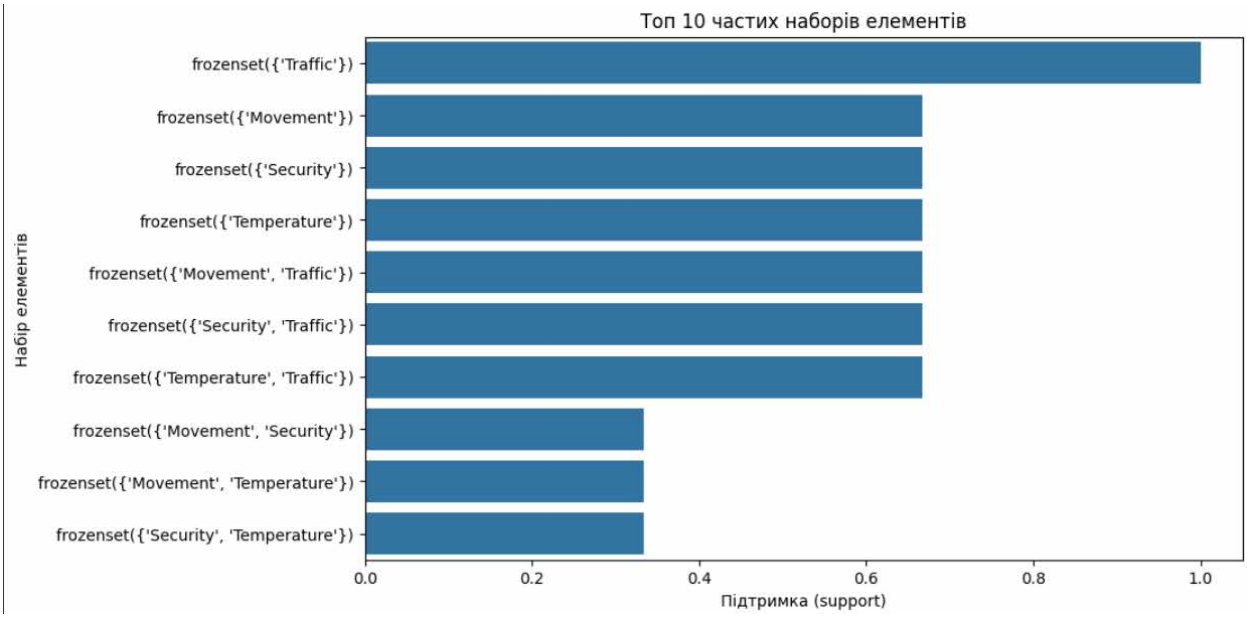

Рис. 9 **Топ-10 частих наборів елементів** у вигляді горизонтальної гістограми

Домінування трафіку в частих наборах елементів і правилах вказує на його важливість у впливі на поведінку системи. Він може слугувати основним індикатором для виявлення аномалій або прогнозування критичних станів системи.

Залежності від руху та температури: Сильні взаємозв'язки між рухом і трафіком свідчать про те, що сплески руху часто корелюють зі змінами в активності мережі. Аналогічно, коливання температури пов'язані з трафіком, що може сигналізувати про ризики перегріву.

Вплив на безпеку: Правила на кшталт (Безпека → Трафік) підкреслюють потенційні зв'язки між попередженнями безпеки і структурою трафіку, допомагаючи в упереджувальних заходах для виявлення загроз.

Визначені правила асоціацій дають чітке розуміння того, як події взаємодіють в системі. Використовуючи ці знання, системні адміністратори можуть проактивно відстежувати і вирішувати критичні патерни, підвищуючи операційну ефективність і знижуючи ризики. Подальше вивчення правил з вищими значеннями підйому в більших наборах даних може виявити ще більш значущі взаємозв'язки.

5.3 . Характеристика кластерів та їхній вплив на процеси в предметній області

Кластерний аналіз поділяє дані на групи на основі спільних характеристик, що дає змогу виявити закономірності та інтерпретувати їхню значущість у предметній області. У цьому дослідженні датчики були згруповані в три кластери на основі їхніх атрибутів «Температура» і «Трафік», причому кожен кластер представляє окремі поведінкові патерни. Ці групування дають практичну інформацію для моніторингу системи, оптимізації продуктивності та виявлення аномалій.

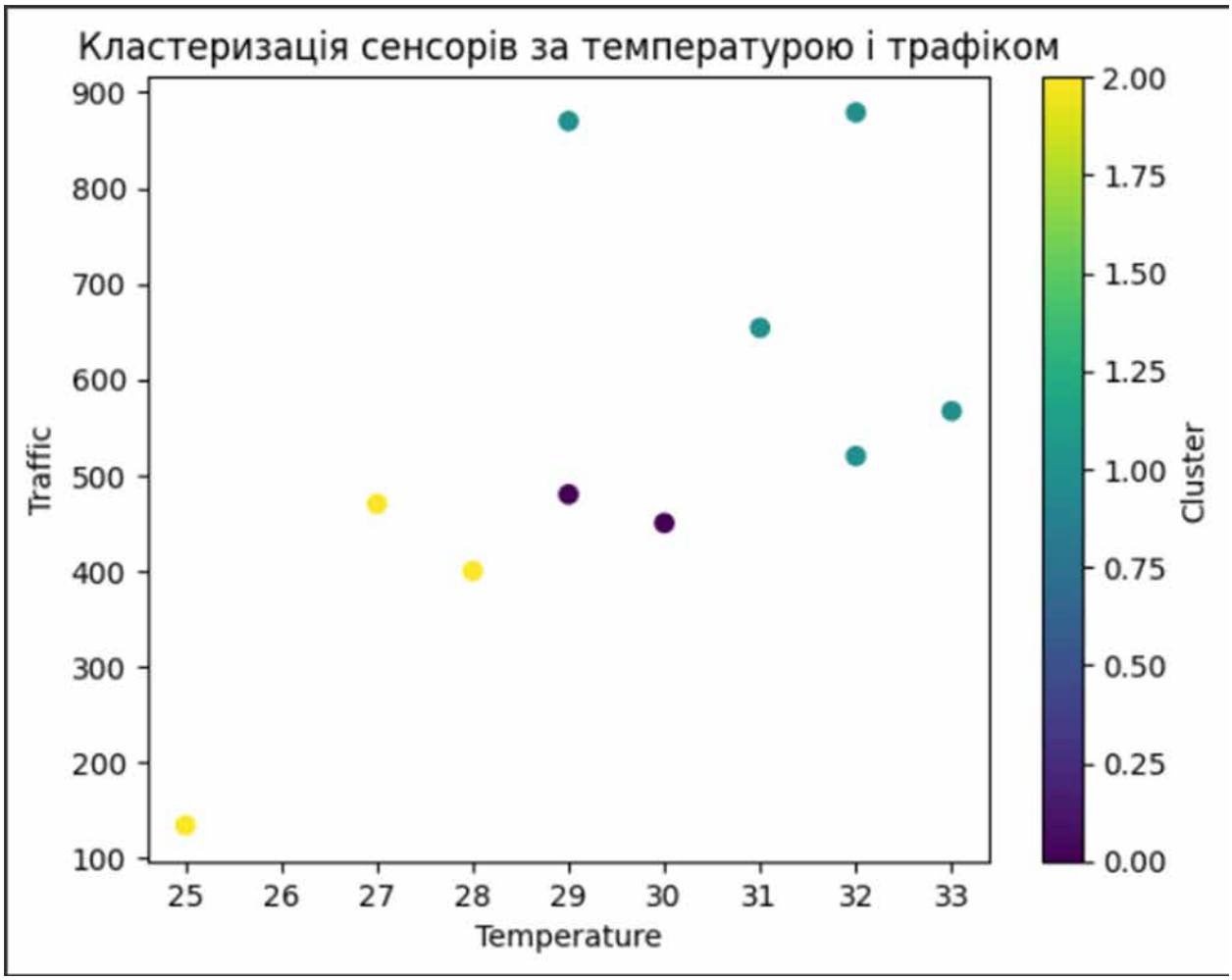

Рис. 10 Кластеризація сенсорів за температурою і трафіком

Характеристики кластерів

Кластер 0:

Середня температура: ~29°C

Середній трафік: ~450 одиниць

Характеристики: До цього кластеру входять переважно датчики з помірним трафіком і близькими до оптимальних показниками температури. Датчики в цій групі, як правило, стабільні, що свідчить про типову продуктивність системи без значних аномалій.

Вплив: Відображає базовий або нормальний робочий стан системи. Моніторинг цієї групи забезпечує стабільність і може слугувати орієнтиром для виявлення відхилень.

Кластер 1:

Середня температура: ~31-33°C

Середній трафік: >650 одиниць

Характеристики: Датчики в цьому кластері демонструють високий рівень трафіку і підвищену температуру. Ця група, ймовірно, вказує на ділянки з інтенсивним використанням мережі, що може бути пов'язано з підвищеним робочим навантаженням або умовами навколишнього середовища, які впливають на продуктивність.

Вплив: Високі рівні трафіку і температури вказують на необхідність ретельнішого моніторингу, щоб запобігти потенційному перегріванню або вузьким місцям у продуктивності. Ця група також може вказувати на вищий ризик перевантаження системи.

Кластер 2:

Середня температура: ~25-27°C

Середній трафік: <200 одиниць

Характеристики: Цей кластер включає датчики з низьким трафіком і нижчими за середні показниками температури. Ці датчики недовикористовуються, можливо, вказуючи на ділянки з мінімальною мережевою активністю або надлишковість системи.

Вплив: Надає можливість оптимізувати розподіл ресурсів. Недостатньо завантажені датчики можуть бути перерозподілені або відкалібровані для підвищення ефективності мережі.

Кластерні інсайти та їхній вплив на процеси

	SensorID	Temperature	Traffic	Status	Cluster
0	1	28	400	0	2
1	2	32	520	1	1
2	3	30	450	0	0
3	4	29	480	1	0
4	5	27	470	0	2
5	6	29	870	1	1
6	7	25	134	0	2
7	8	31	654	0	1
8	9	32	879	1	1
9	10	33	567	1	1

Рис.11 Інформативна таблиця про дані сенсорів

Кластери виявляють закономірності в поведінці датчиків, що полегшує виявлення аномалій. Наприклад, датчики, що значно відхиляються від середнього

показника по кластеру (наприклад, датчик кластера 0 з незвично високим трафіком), можуть бути позначені для подальшого дослідження.

Виявлення недовантажених датчиків у Кластері 2 дає змогу краще управляти ресурсами. Ці датчики можна перерозподілити для виконання завдань у зонах з вищими вимогами, зменшуючи загальну неефективність.

Відстежуючи типову поведінку кластерів, аномалії, такі як раптові стрибки трафіку або температури, можна виявити в режимі реального часу, що підвищує стійкість системи до несподіваних проблем.

Профілактичні заходи:

Датчики кластера 1, які показують високі температури, можуть потребувати профілактичного охолодження або перенаправлення трафіку, щоб зменшити ризики виходу з ладу обладнання або погіршення продуктивності.

Візуалізація кластерів

Діаграма розсіювання температури в залежності від трафіку, розфарбована відповідно до приналежності до кластеру, чітко розділяє датчики на три окремі групи. Ця візуалізація допомагає швидко виявити відхилення або потенційні ризики в кожному кластері. Наприклад:

Датчики кластера 1 сконцентровані на більш високих рівнях температури і трафіку, що підкреслює їх критично важливу роль в процесах системи.

Датчики кластера 2 з низьким трафіком можна оптимізувати або контролювати рідше.

Кластерний аналіз надав цінну інформацію про поведінку датчиків та їхню роль у системі. Кожен кластер представляє певний робочий стан, від нормальної продуктивності до сценаріїв з високим попитом. Використовуючи ці знання, організації можуть підвищити надійність системи, оптимізувати використання ресурсів і проактивно реагувати на потенційні ризики. Такий підхід до кластеризації забезпечує ефективний моніторинг та управління, що значно підвищує стабільність та продуктивність мережевої інфраструктури.

ВИСНОВКИ

У цій бакалаврської кваліфакіційної роботі розглядається різноманітність мережевого обладнання та способи керування ним. Проаналізовано ринок існуючих систем контролю та обліку мережевого обладнання. Звертається увага на функціональні та нефункціональні вимоги системи. Розроблено низку архітектурних рішень із порівнянням переваг і недоліків кожного.

У бакалаврської кваліфакіційної роботі описано проектування автоматизованої системи контролю та обліку мережевого обладнання та використовуваних засобів. Було можливим реалізувати програмний засіб для зміни даних обладнання віддалено та інтерфейс користувача як частину автоматизованої системи. Це спростить і скоротить витрати часу на налаштування мережевого обладнання.

Цей програмний продукт сприятиме автоматизації роботи з мережевим обладнанням, без використання інтерфейсу командного рядка, який, до того ж, у кожного виробника різний, дозволить користувачам взаємодіяти з обладнанням без вивчення роботи обладнання та його функцій. детально. Все це призведе до полегшення взаємодії технічних фахівців і менш обізнаних користувачів з мережевим обладнанням. А можливість перегляду та збереження інформації про пристрої дозволить зручніше керувати їх обігом, допоможе зручніше та якісніше побудувати мережеву архітектуру з наявних пристроїв або спростить складання списку для їх покупок.

Список Джерел

1. Аналіз мережевого обладнання для побудови мережі Інтернет-провайдера [Електронний ресурс] / С.С. Волошко, А.Ю. Фролов - Режим доступу до ресурсу: http://reposit.nupp.edu.ua/xmlui/bitstream/handle /PoltNTU/4575/%d0%a1%d1%82%d0%b0%d1%82%d1%82%d1%8f%20%d0%9d%d0%86%d0%a1%d0%a2%20%d0%a4%d1%80%d0%be%d0%bb%d0%be%d0%b2.pdf?послідовність =1&isAllowed=y.

2. Огляд та оновлення рішень Juniper Networks для маршрутизації, комутації та безпеки [Електронний ресурс] / Мук. – 2016. – Режим доступу до ресурсу: https://habr.com/ru/company/muk/blog/280338/.

3. Мережеве обладнання Cisco [Електронний ресурс] / Аквилона – Режим доступу до ресурсу: http://www.akvilona.ru/serv/cisco.htm.

4. Протоколи та інтерфейси керування дротовими мережами доступу [Електронний ресурс] / Studbooks – Режим доступу до ресурсу:https://studbooks.net/2179548/informatika/protokoly_upravleniya_setyu_dostupa .

5. Аудит мережевої безпеки [Електронний ресурс] / https://eska.global/blog/setevoj-monitoring-protokoly-luchshie-praktiki-instrumenty-2020

6. Функції систем управління [Електронний ресурс] / iptcp – Режим доступу до ресурсу: http://iptcp.net/funktsii-sistem-upravleniya.html.

7. Вартові мережі. Утиліти з відкритим кодом для керування, моніторингу та резервного копіювання налаштувань мережевого обладнання [Електронний ресурс] / Мартін Пранкевич. – 2015. – Режим доступу до ресурсу: https://xakep.ru/2015/08/10/network-management/.

8. Ansible vs Salt (SaltStack) vs StackStorm [Електронний ресурс] / Ігор Олемський. – 2017. – Режим доступу до ресурсу:https://habr.com/ru/company/southbridge/blog/330594/.

9. 10 кращих програм для інвентаризації мережі 2020 [Електронний ресурс] / softinventive – Режим доступу до ресурсу: https://www.softinventive.ru/best-network-inventory-tools/.

10. Архітектура інформаційних систем [Електронний ресурс] / it-claim – Режим доступу до ресурсу:http://it-claim.ru/Education/Course/ISDevelopment/Lecture_3.pdf.

11. Порівняння сучасних СУБД [Електронний ресурс] / Володимир Драч. – 2017. – Режим доступу до ресурсу:https://drach.pro/blog/hi-tech/item/145-db-comparison.

12. Ранжування систем управління базами даних за популярністю [Електронний ресурс] / db-engines. – 2020. – Режим доступу до ресурсу: https://db-engines.com/en/ranking.

13. Загальне та приватне про веб-сервери [Електронний ресурс] / Олексій Кошелев - Режим доступу до ресурсу: https://compress.ru/article.aspx?id=11744.

14. APACHE VS NGINX – ПОРІВНЯННЯ ТА ПЕРЕВАГИ [Електронний ресурс] / MERION NETWORKS. – 2020. – Режим доступу до ресурсу:https://wiki.merionet.ru/servernye-resheniya/34/apache-vs-nginx-sravnenie-i-преимущества/.

15. Моніторинг мережі. Протоколи, найкращі практики, інструменти 2021"ХЇ[Електронний ресурс] / - Режим доступу до ресурсу: https://lemon.school/blog/kiberbezpeka-aktualni-zagrozy-ta-metody-zahystu

16. Аудит мережевої безпеки. [Електронний ресурс] / - Режим доступу до ресурсу: https://netwave.ua/service/audyt-it-system-ta-infrastruktury/analiz-zahyshhenosti-merezhi/