|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Федеральное государственное автономное образовательное  учреждение высшего образования | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **Дальневосточный федеральный университет** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **Департамент информационной безопасности** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **О Т Ч Е Т** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| о прохождении производственной практики. Организационно-управленческая практика | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | |  | |  | Руководитель группы практикантов  Студент гр. Б9122-10.03.01 | | | | | | | | | | | | | |
|  | |  |  | | | | | | | | | | Евдокимов И. А. | | | |
|  | | | | | | | | | | |  | |  | (подпись) | | | | | | | | | |  | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Отчет защищен с оценкой | | | | | | | | | | |  | |  | Руководитель практики | | | | | | | | | | |  | | |
|  | | | | | | | | | | |  | |  | Ст. преподаватель департамента информационной безопасности ИМКТ | | | | | | | | | | | | | |
|  | | | |  | Боршевников А. Е. | | | | | |  | |  |  | | | | | | |  | | Боршевников А. Е. | | | | |
| (подпись) | | | |  | (И.О. Фамилия) | | | | | |  | |  | (подпись) | | | | | | |  | | (И.О. Фамилия) | | | | |
| « | 05 | » | августа | | | | | | 2024г. | |  | |  |  | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Регистрационный № | | | | | | | |  | | |  | |  | Практика пройдена в срок | | | | | | | | | | | | | |
| « | 05 | » | августа | | | | | | 2024 г. | |  | |  | с | | « | 24 | | » | июля | | | | | | 2024 г. | |
|  | | | | | | | | | | |  | |  | по | | « | 05 | | » | августа | | | | | | 2024 г. | |
|  | | | | |  | | Е.В. Третьяк | | | |  | |  | на предприятии | | | | | | | | | | | | | |
| (подпись) | | | | |  | | (И.О. Фамилия) | | | |  | |  |
|  | | | | | | | | | | |  | |  | Департамент информационной | | | | | | | | | | | | | |
|  | | | | | | | | | | |  | |  | безопасности ИМиКТ ДВФУ | | | | | | | | | | | | | |
|  | | | | | | | | | | |  | |  |  | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| г. Владивосток | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2024 г. | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| СПИСОК ИСПОЛНИТЕЛЕЙ | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Руководитель группы практикантов, студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Евдокимов И. А | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | | (введение, заключение) | | | | |
| Заместитель руководителя, студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Глущенко Г. И. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | | (компоновка разделов, документация) | | | | |
| Исполнители: | | | | | |  | | | |  | |  | |  |  | | |  | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Вернер В. С. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Нистор И. В. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Чернухина К. С. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Тоболов Д.А. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Мигаль Я. Г. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Мельников А. Н. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Полхонов Р. Б. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |
| Студент | | | | | |  | | | | « | |  | | » |  | | | 2024 г. | | | | Минеев В. А. | | | | |
|  | | | | | | (подпись) | | | | (дата) | | | | | | | | | | | |  | | | | |

**Оглавление**

[Введение 5](#_Toc173151165)

[Задание на практику 5](#_Toc173151166)

[Цель 5](#_Toc173151167)

[Организация работы 5](#_Toc173151168)

[Задачи 5](#_Toc173151169)

[Структура отчёта 7](#_Toc173151170)

[Раздел 1. Методы обнаружения атак, аномалий в сети. Использование методов машинного обучения и моделей в IDS/IPS 8](#_Toc173151171)

[Методы обнаружения изменений, их преимущества и недостатки 8](#_Toc173151172)

[Этапы применения методов машинного обучения в IDS/IPS 11](#_Toc173151173)

[Интеграция методов машинного обучения в системы обнаружения вторжений (IDS/IPS) 13](#_Toc173151174)

[«Обучение без учителя» для выявления аномалий в поведении пользователей и сетевых устройств 14](#_Toc173151175)

[Использование генеративных моделей для обнаружения аномалий в сетевом трафике. 16](#_Toc173151176)

[Применение нейронных сетей для обнаружения вторжений, аномалий и изменений в сетевом трафике 17](#_Toc173151177)

[Сравнение методов машинного обучения для обнаружения аномалий в сетевом трафике 18](#_Toc173151178)

[Раздел 2. Методы классификации атак. Роль глубокого обучения в предотвращении атак. 20](#_Toc173151179)

[Методы классификации с использованием методов машинного обучения 20](#_Toc173151180)

[Роль глубокого обучения в предотвращении кибератак 23](#_Toc173151181)

[Роль больших данных и предварительной обработки в эффективности IDS 24](#_Toc173151182)

[Применение глубокого обучения нейронных сетей для обнаружения сложных атак 25](#_Toc173151183)

[Раздел 3. Анализ работы и стабильность ML-систем в сфере кибербезопасности. Гуманитарные и этические аспекты. 28](#_Toc173151184)

[Анализ уязвимостей и угроз, связанных с применением методов машинного обучения в области кибербезопасности. 28](#_Toc173151185)

[Реальная обстановка ML-систем в сфере информационной безопасности. 33](#_Toc173151186)

[Проблемы безопасности данных и этические аспекты использования больших данных и методов машинного обучения в киберзащите: 34](#_Toc173151187)

[Техники предотвращения вторжений: 37](#_Toc173151188)

[Перспективы развития и будущее направление 42](#_Toc173151189)

[Заключение. Подведение итогов. 45](#_Toc173151190)

[ПРИЛОЖЕНИЕ №1 46](#_Toc173151191)

[Список используемой литературы 51](#_Toc173151192)

# Введение

## Задание на практику

Команда объединилась для прохождения практики и совместного изучения темы «Анализ методов машинного обучения в задаче обнаружения и предотвращения вторжений».

## Цель

Целью выполнения практической работы стало эффективное изучение и получение новой информации по теме, описанной выше. Также была поставлена цель

## Организация работы

После образования команды были распределены роли. Пост лидера, ответственного за выполнение практики взял на себя Евдокимов И. А. В его обязанности было включено определение целей и задач практики, а также распределение работы между членами команды, подведение итогов и связь с руководителем практики. Был также назначен заместитель, занимающийся координацией действий команды, сбором материалов (отчётов) и оформлением основного отчёта практики – Глущенко Г. И.

## Задачи

Основной задачей на практику являлось получение знаний на тему использования методов машинного обучения в обнаружении вторжений. В качестве концепции построения работы было решено разбить тему на подтемы и распределить их между студентами. Таким образом появилось распределение задач, которое можно наблюдать в следующей таблице:

|  |  |  |
| --- | --- | --- |
| **Задача/тема** | **ФИО ответственного** | **Отметка о выполнении (дата)** |
| Распределение задач и тем практики. Подведение итогов. Связь с руководителем практики. | Евдокимов И. А |  |
| Проведение общего организационного сбора, организация работы. | выполнено  (14.04.2024) |
| Проведение вводной лекции для команды. Введение в тему практической работы | выполнено  (26.04.2024) |
| Контроль выполнения и сбор работ (отчётов). Оформление общего (настоящего) отчёта. | Глущенко Г. И. | выполнено  (03.07.2024) |
| Методы обнаружения изменений и вторжений (IDS/IPS) | Чернухина К. С. | выполнено  (09.06.2024) |
| Методы классификации вторжений (IDS/IPS) |
| «Обучение без учителя» для выявления аномалий в поведении пользователей и сетевых устройств | Полхонов Р. Б. | выполнено  (27.05.2024) |
| Интеграция методов машинного обучения в системы обнаружения вторжений (IDS/IPS) |
| Проблемы безопасности и этические аспекты | Нистор И. В. | выполнено  (09.06.2024) |
| Техники предотвращения вторжений |
| Этапы применения методов машинного обучения в IDS/IPS: | Тоболов Д.А. | выполнено  (05.06.2024) |
| Роль машинного обучения в IDS/IPS: |
| Использование генеративных моделей для обнаружения аномалий в сетевом трафике | Мельников А. Н. | выполнено  (05.06.2024) |
| Сравнение методов машинного обучения для обнаружения аномалий в сетевом трафике |
| Анализ уязвимостей и угроз, связанных с применением методов машинного обучения в области кибербезопасности. | Вернер В. С. | выполнено  (27.05.2024) |
| Реальная обстановка ML-систем в сфере информационной безопасности. |
| Применение нейронных сетей в системах обнаружения вторжений | Мигаль Я. Г. | выполнено  (05.06.2024) |
| Роль глубокого обучения в предотвращении кибератак |
| Роль больших данных и предварительной обработки в эффективности IDS | Минеев В. А. | выполнено  (09.06.2024) |
| Перспективы развития и будущее направления |

По мере выполнения командой своей части работы, таблица будет заполняться Глущенко Г. И., как и настоящий отчёт.

## Структура отчёта

Разбиение отчёта произведено в связи с обширностью тем на смежные по смыслу и содержанию темы

Отчёт разбит на 3 основных раздела:

### Раздел 1. Методы обнаружения атак, аномалий в сети. Использование методов машинного обучения и моделей в IDS/IPS

### Раздел 2. Методы классификации атак. Роль глубокого обучения в предотвращении атак.

### Раздел 3. Анализ работы и стабильность ML-систем в сфере кибербезопасности. Гуманитарные и этические аспекты.

Каждый раздел составлен из материалов нескольких студентов, предоставивших свои материалы. Общий список литературы приведён в конце отчёта.

# Раздел 1. Методы обнаружения атак, аномалий в сети. Использование методов машинного обучения и моделей в IDS/IPS

## Методы обнаружения изменений, их преимущества и недостатки

Методы обнаружения изменений (англ. change detection) играют важную роль в обеспечении безопасности информационных систем и сетей. Они направлены на выявление аномальных или вредоносных действий, которые могут указывать на наличие вторжений или попыток взлома. В контексте машинного обучения, задача обнаружения изменений заключается в поиске отклонений от нормального поведения системы, что может свидетельствовать о наличии угрозы.

Существует несколько методов обнаружения изменений, которые применяются в контексте машинного обучения для обнаружения и предотвращения вторжений.

### Методы на основе статистики:

* Контроль Ширтье-Робертса (CUSUM): Этот метод используется для обнаружения небольших изменений в статистике процесса. Он анализирует последовательные данные и выявляет изменения, превышающие заданный порог.
* Контроль Чуи-Пейра (CP): Этот метод также используется для обнаружения небольших изменений, основываясь на статистических тестах и сравнении данных с моделью нормального поведения.

Преимущества:

1. **Относительная простота**: Многие методы обнаружения изменений на основе статистики относительно просты в реализации и понимании. Они могут быть легко внедрены и использованы без глубоких знаний в области машинного обучения.
2. **Эффективность в специфических случаях**: В некоторых сценариях, особенно когда изменения имеют четко выраженные статистические аномалии, методы на основе статистики могут быть очень эффективны.

Недостатки:

1. **Чувствительность к выбору параметров**: Многие методы требуют настройки параметров, таких как пороговые значения или размер окна анализа, что может потребовать экспертных знаний и опыта.
2. **Неэффективность в сложных сценариях**: В реальных сценариях изменения могут быть сложными и неоднородными, что может затруднить обнаружение с помощью методов на основе статистики.
3. **Трудности с адаптацией к динамическим изменениям**: Методы статистики могут иметь трудности с быстрым обнаружением динамических изменений, таких как появление новых видов вторжений или эволюция тактик злоумышленников.

В целом, методы обнаружения изменений на основе статистики полезны в определенных контекстах, но их применение может быть ограничено в более сложных и динамичных сценариях. Комбинация методов на основе статистики с другими подходами, такими как машинное обучение или анализ временных рядов, может помочь улучшить обнаружение вторжений в разнообразных условиях.

### Методы на основе машинного обучения:

* Алгоритмы кластеризации: Неконтролируемые методы кластеризации, такие как k-means, могут использоваться для выявления аномалий в данных. Аномальные точки могут указывать на наличие вторжений.
* Обучение с учителем: Классификационные модели, такие как метод опорных векторов (SVM), случайный лес или нейронные сети, могут быть обучены на основе размеченных данных об аномальном и нормальном поведении для обнаружения вторжений.

Преимущества:

1. **Адаптивность к изменяющимся условиям**: Методы машинного обучения способны обучаться на основе новых данных и адаптироваться к изменяющимся сценариям, что делает их более эффективными в обнаружении сложных и динамических изменений.
2. **Обнаружение неявных шаблонов**: Машинное обучение способно выявлять сложные и неочевидные шаблоны, которые могут указывать на наличие вторжений или изменений в системе.
3. **Уменьшение ложных срабатываний**: Методы машинного обучения могут помочь уменьшить количество ложных срабатываний за счет анализа более широкого спектра признаков и контекста.

Недостатки:

1. Требовательность к данным: Методы машинного обучения требуют большого объема размеченных данных для обучения, что может быть проблематично в случае ограниченного доступа к данным о вторжениях.
2. Сложность интерпретации: Некоторые модели машинного обучения, особенно глубокое обучение, могут быть сложными для интерпретации, что затрудняет понимание причин обнаружения вторжений.
3. Необходимость **постоянного обновления**: Модели машинного обучения требуют постоянного обновления и переобучения, чтобы сохранить свою эффективность в условиях постоянно меняющихся видов вторжений.

Методы машинного обучения предлагают широкий спектр возможностей для обнаружения вторжений, но их применение также сопряжено с рядом вызовов, которые необходимо учитывать при выборе подходящего метода для конкретной задачи обнаружения и предотвращения вторжений.

### Методы на основе анализа потока данных:

* Методы временных рядов: Модели временных рядов, например, авторегрессионные модели (ARIMA) или экспоненциальное сглаживание, могут использоваться для анализа изменений во времени и обнаружения аномалий.
* Методы анализа графов: В контексте сетевой безопасности, анализ графов может использоваться для выявления аномалий в сетевом трафике и обнаружения вторжений.

Преимущества:

1. **Обнаружение новых типов аномалий**: Методы обнаружения изменений на основе анализа потока данных могут выявлять новые типы аномалий или изменений, которые ранее не были известны или не были представлены в обучающих данных.
2. **Реальное время**: Эти методы способны обрабатывать данные в реальном времени, что позволяет оперативно реагировать на возможные угрозы или изменения в системе.
3. **Способность к самообучению**: Некоторые методы анализа потока данных могут быть настроены на автоматическое обучение и адаптацию к изменяющимся условиям, что делает их более гибкими.

Недостатки:

1. **Чувствительность к шуму**: Методы обнаружения изменений на основе анализа потока данных могут быть чувствительны к шуму в данных, что может приводить к ложным срабатываниям.
2. **Сложность определения нормального поведения**: Определение нормального поведения в потоке данных может быть сложной задачей, особенно в случаях, когда данные сильно изменчивы или подвержены сезонным колебаниям.
3. **Необходимость постоянного мониторинга и обновления параметров**: Методы анализа потока данных требуют постоянного мониторинга и обновления параметров, чтобы сохранить свою эффективность в условиях изменяющейся среды.

Методы обнаружения изменений на основе анализа потока данных представляют собой мощный инструмент для обнаружения аномалий и изменений в реальном времени, но их применение требует тщательной настройки и постоянного контроля для минимизации ложных срабатываний и обеспечения высокой точности обнаружения.

Стоит упомянуть, что существуют также подходы, использующие ***гибридные методы***, такие как статистические тесты, машинное обучение и анализ временных рядов, для повышения точности обнаружения и снижения ложных срабатываний.

## Этапы применения методов машинного обучения в IDS/IPS

Современные системы обнаружения и предотвращения вторжений (IDS/IPS) играют ключевую роль в обеспечении информационной безопасности. В последнее время всё большее внимание уделяется применению методов машинного обучения, которые демонстрируют высокую эффективность в обнаружении, идентификации и классификации различных типов атак. Ниже представлены этапы подготовки и применения инструментов, использующих методы машинного обучения:

### Сбор данных

Согласно исследованию, представленному в сборнике научных трудов, важным этапом при использовании методов машинного обучения в системах IDS/IPS является **сбор и подготовка соответствующих данных**. Авторы отмечают, что для эффективного обучения моделей необходимо формирование обширных и репрезентативных наборов данных, включающих как нормальный сетевой трафик, так и записи различных типов атак. Данные должны быть тщательно очищены, структурированы и преобразованы в формат, пригодный для использования алгоритмами машинного обучения.

Процесс сбора данных может включать в себя **мониторинг сетевого трафика**, **интеграцию с журналами событий безопасности**, а также **использование существующих наборов данных**, таких как DARPA, KDD Cup 99 или UNSW-NB15. Важно, чтобы собранные данные были достаточно разнообразны и отражали реальные угрозы, с которыми сталкиваются организации. Это позволит обучить модели, способные эффективно распознавать как известные, так и новые виды атак.

### Выбор и применение подходящих алгоритмов машинного обучения

Исследование также показывает, что для решения задач классификации атак в системах IDS/IPS необходим грамотный **выбор и применение подходящих алгоритмов машинного обучения**. Авторы подчеркивают, что в зависимости от особенностей данных и целей исследования могут использоваться различные подходы, такие как **деревья решений**, **логистическая регрессия**, **искусственные нейронные сети** и другие. Важным является также **настройка гиперпараметров** моделей для достижения оптимальной производительности.

При выборе алгоритмов необходимо учитывать такие факторы, как **сложность и интерпретируемость моделей**, **способность работать с высоко размерными данными**, **устойчивость к шуму и пропущенным значениям**. Например, деревья решений могут обеспечивать хорошую интерпретируемость результатов, в то время как нейронные сети демонстрируют высокую точность на сложных задачах. Применение ансамблевых методов, таких как случайный лес или градиентный бустинг, также может повысить качество классификации.

### Оценка производительности и точности моделей

Заключительным этапом, на который указывают авторы сборника, является тщательная **оценка производительности и точности моделей** машинного обучения, применяемых в IDS/IPS. Это включает в себя использование различных метрик, таких как **точность**, **полнота**, **F1-мера**, для всесторонней характеристики качества классификации. Кроме того, важно проводить перекрестную проверку моделей на независимых тестовых наборах данных, чтобы обеспечить их надежность и способность эффективно работать в реальных условиях.

При оценке моделей следует учитывать не только общие показатели, но и более детальные метрики, такие как **точность распознавания отдельных типов атак**. Это позволит выявить сильные и слабые стороны моделей и при необходимости провести дополнительную оптимизацию. Также важно анализировать ошибки классификации, чтобы понять, где модели могут давать неправильные результаты и принять соответствующие меры.

Представленное в сборнике научных трудов исследование заключает, что использование методов машинного обучения в системах обнаружения и предотвращения вторжений является перспективным направлением, позволяющим повысить эффективность и гибкость таких систем. Применение машинного обучения обеспечивает более точную классификацию атак, в том числе новых и неизвестных угроз, повышает масштабируемость и адаптивность IDS/IPS, а также производится в несколько поддающихся простому нормированию и оценке этапов, таких как:

1. Сбор данных
2. Выбор алгоритмов машинного обучения
3. Оценка производительности моделей

Тщательный подход к сбору и подготовке данных, выбору алгоритмов, а также оценке производительности моделей играет ключевую роль в достижении высоких результатов.

## Интеграция методов машинного обучения в системы обнаружения вторжений (IDS/IPS)

Интеграция методов машинного обучения включает в себя следующие аспекты:

### Разработка гибридных подходов, сочетающих сигнатурный и поведенческий анализ

Сигнатурный анализ - традиционный подход, основанный на сопоставлении сетевого трафика и событий с известными шаблонами атак. Высокая точность выявления распространенных, задокументированных угроз является основным преимуществом, однако возникают проблемы с обнаружением новых, неизвестных типов вторжения.

Поведенческий анализ на основе машинного обучения в основном использует методы без учителя (unsupervised) для выявления аномальных паттернов активности. При анализе трафика метод способен обнаруживать неизвестные, нестандартные типы атак, но в автоматическом режиме способен также ложно срабатывать. Используя подобную систему, стоит тщательно настраивать её работу.

Гибридные подходы в данном случае – некое решение, сочетание сигнатурного и поведенческого анализа для повышения точности и достоверности обнаружения. Практикуется использование ML-моделей для выявления аномалий, а сигнатурного анализа - для их лучшей классификации. Адаптивная настройка весов различных компонентов поможет настроить систему, поддерживать её работоспособность при разной динамике угроз.

Пример: Применение одноклассовой SVM для выявления аномалий, а затем использование экспертных правил для их классификации.

### Вопросы масштабируемости и производительности при использовании ML-моделей в реальном времени:

Для обеспечения масштабируемости для обработки всё больших объемов сетевого трафика и событий безопасности, часто используются распределение вычислений, параллелизация, а также платформы (soft- hardware) для обработки больших данных, такие как Apache Spark, Hadoop.

Для обеспечения производительности следует на постоянной основе модернизировать систему, а также оптимизировать алгоритмы. Особое внимание стоит уделить необходимости быстрого принятия решений в режиме реального времени, оптимизации вычислительной сложности ML-моделей, использованию эффективных алгоритмов, применению техник ускорения вывода, таких как GPU-ускорение, использование ONNX. Также стоит упомянуть, что всегда следует приходить к компромиссу в вопросах соотношения сложности и скорости принятия решений.

### Интеграция с SIEM и другими системами безопасности:

Не стоит забывать, про же существующее ПО. Можно использовать все методы и подходы, ведь при правильной интеграции, они смогут взыимно повысить общую эффективность. Таким образом, уже сегодня происходит передача результатов анализа ML-моделей в SIEM (Security Information and Event Management) для централизованной обработки, взаимодействие с системами предотвращения вторжений (IPS) для автоматического реагирования на выявленные угрозы, а также взаимодействие с такими инструментами, как SOAR (Security orchestration, automation and response), для автоматизации аналитики и реагирования.

## «Обучение без учителя» для выявления аномалий в поведении пользователей и сетевых устройств

В машинном обучении существует понятие, как модели обученные «без учителя», которое особенно эффективны при поиске выбросов, или даже незначительных отклонений, что также может использоваться при обнаружении вторжений. Из существующих на сегодняшний день можно выделить два основных метода группировки объектов, с помощью которых можно идентифицировать выброс: кластеризация и классификация.

### *Методы кластеризации для выявления подозрительных шаблонов активности:*

Алгоритм **K-means кластеризации** разбивает данные на K кластеров, минимизируя сумму квадратов расстояний между объектами и центрами их кластеров. В целом, алгоритм автоматический, за исключением выбора К: стоит эффективно определить число кластеров, для этого существуют специальные алгоритмы - метрики, такие как silhouette score, elbow method и т. д. K-means также весьма чувствителен к наличию аномальных, выбивающихся объектов, которые могут сместить центры кластеров. Кластеризация профилей пользователей, сетевых сессий или других характеристик трафика позволяет выявлять необычные шаблоны активности, указывающие на возможные атаки.

**DBSCAN кластеризация** определяет кластеры как области с высокой плотностью объектов, отделенные друг от друга областями с низкой плотностью. DBSCAN способен идентифицировать аномальные объекты, не относящиеся ни к одному кластеру, что повышает его устойчивость к выбросам. В отличие от K-means, DBSCAN не ограничен гиперсферическими кластерами. Кластеризация профилей активности устройств или пользователей позволяет находить нехарактерные для нормального поведения группы, которые могут указывать на скомпрометированные или аномальные сущности.

### Методы одноклассовой классификации для обнаружения отклонений от нормального профиля:

**OneClassSVM** обучается только на примерах нормального, не аномального поведения. Он строит гиперплоскость, максимально разделяющую нормальные данные от возможных аномалий. Алгоритм не требует размеченных данных для аномалий, что критично в задачах обнаружения вторжений, где новые атаки постоянно появляются. OneClassSVM может применяться для выявления отклонений в сетевом трафике, активности пользователей, конфигурациях устройств от обученной модели нормального профиля.

**Isolation Forest** строит ансамбль деревьев решений, которые разделяют данные таким образом, что аномальные объекты оказываются ближе к корням деревьев. В основе алгоритма лежит предположение, что аномалии более изолированы от основной массы данных, и для их изоляции требуется меньше решений в дереве. Isolation Forest может использоваться для выявления нетипичных паттернов в сетевом трафике, журналах событий, характеристиках устройств.

## Использование генеративных моделей для обнаружения аномалий в сетевом трафике.

**Генеративные модели** — это класс алгоритмов машинного обучения, которые способны генерировать новые данные, похожие на обучающий набор. Они отличаются от дискриминативных моделей, которые предсказывают метки или категории для входных данных. Генеративные модели пытаются понять и воспроизвести распределение исходных данных, что позволяет им создавать новые примеры, которые могут быть неотличимы от реальных данных.

### Основные типы генеративных моделей:

**Generative Adversarial Networks (GANs):** состоит из двух сетей — генератора, который создает данные, и дискриминатора, который оценивает, насколько хорошо они соответствуют реальным данным. Обучение GAN напоминает игру в кошки-мышки, где генератор улучшается в создании данных, а дискриминатор — в их оценке.

**Variational Autoencoders (VAEs):** модели используют архитектуру автокодировщика для создания новых данных. Они преобразуют входные данные в сжатое представление и затем восстанавливают их обратно, при этом обучаясь генерировать данные, близкие к исходным.

### *Применение для обнаружения аномалий в сетевом трафике*

Генеративные модели могут быть использованы для обнаружения аномалий в сетевом трафике, поскольку они способны моделировать нормальное поведение сети и выявлять отклонения от этого нормального поведения. Например, если генеративная модель обучена на данных нормального трафика, она будет генерировать примеры, которые соответствуют этому нормальному трафику. Когда встречается трафик, который значительно отличается от сгенерированных примеров, это может указывать на аномалию.

**GANs** могут быть использованы для генерации примеров нормального трафика, а затем дискриминатор может оценить, является ли новый трафик аномальным. **VAEs** могут использоваться для создания модели нормального трафика и выявления аномалий, когда реальный трафик не соответствует этой модели.

Таким образом, генеративные модели предоставляют мощный инструмент для обнаружения аномалий, способный адаптироваться к новым и изменяющимся паттернам в данных, что делает их ценными для обеспечения безопасности сетей.

## Применение нейронных сетей для обнаружения вторжений, аномалий и изменений в сетевом трафике

Наравне с традиционными методами обнаружения вторжений, неподалёку от методов машинного обучения свою нишу занимают и нейросети. Сегодня нейронные сети стали активно применяться для анализа сетевого трафика, позволяя обнаруживать аномалии и потенциальные угрозы на ранней стадии. Особой популярностью пользуются 2 архитектуры нейронных сетей: автокодировщики и рекуррентные сети

### Архитектуры нейронных сетей

Автокодировщики (Autoencoders, AE) являются одним из наиболее популярных подходов в области обнаружения аномалий. Они состоят из двух основных частей: энкодера, который уменьшает размерность входных данных, и декодера, который восстанавливает исходные данные. В процессе обучения AE стремится минимизировать ошибку между исходными данными и их восстановленными версиями. При этом, нормальное поведение данных приводит к меньшим ошибкам, что позволяет использовать эту величину для обнаружения аномалий. Также автокодировщики могут использоваться для моделирования нормального поведения в сети

Рекуррентные нейронные сети (RNN) хорошо подходят для анализа временных рядов, что делает их полезными для анализа сетевого трафика. Эти сети способны учиться на основе последовательностей данных, что позволяет им учитывать временные зависимости в трафике и обнаруживать аномалии, основываясь на изменениях в паттернах трафика во времени.

### Подходы к обучению

Для обучения нейронных сетей на задаче обнаружения аномалий в сетевом трафике используются различные подходы. Одним из них является использование ансамблей моделей, где каждая модель обучается на отдельном подмножестве данных, а затем результаты их работы объединяются для повышения точности обнаружения аномалий. Также применяются методы регуляризации, такие как dropout и L1/L2 регуляризация, для предотвращения переобучения и улучшения обобщающей способности модели.

### Оптимизация нейронных сетей

Оптимизация нейронных сетей включает в себя выбор правильных гиперпараметров, таких как скорость обучения, количество слоев, размерность скрытых слоев и функции активации. Использование методов, таких как Grid Search и Random Search, помогает найти оптимальные значения этих параметров. Кроме того, современные алгоритмы оптимизации, например Adam и RMSprop, позволяют более эффективно обновлять веса модели в процессе обучения.

Нейронные сети представляют собой мощный инструмент для обнаружения аномалий в сетевом трафике, позволяя идентифицировать потенциальные угрозы на ранней стадии. Автокодировщики и рекуррентные сети являются двумя ключевыми архитектурами, используемыми для этой задачи, благодаря своей способности улавливать сложные зависимости и паттерны в данных. Дальнейшие исследования и разработки в области нейронных сетей могут помочь улучшить эффективность и надежность систем обнаружения вторжений и обеспечить более высокий уровень безопасности для информационных ресурсов.

## Сравнение методов машинного обучения для обнаружения аномалий в сетевом трафике

Отличия методов обнаружения вторжений порождают выбор, который должен совершить каждый специалист информационной безопасности под определённую задачу. В контексте обнаружения аномалий в сетевом трафике, сравнение помогает определить, какой метод лучше всего справляется с выявлением необычного поведения в тех или иных данных.

**Основные категории методов машинного обучения для обнаружения аномалий:**

1. **Классические методы - э**то базовые статистические подходы, такие как методы основанные на расстоянии (k-средних, DBSCAN) или на плотности (LOF). Они эффективны для наборов данных с чётко определёнными группами или кластерами.Классические методы часто используются как первый этап анализа, чтобы быстро выявить потенциальные аномалии.
2. **Методы на основе классификации** используют обученные модели для различения нормального и аномального поведения, например, SVM или деревья решений. Они часто требуют предварительно помеченных данных для обучения, что может быть ограничением, если такие данные недоступны.
3. **Нейронные сети:** Глубокое обучение предоставляет инструменты, такие как свёрточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN), которые могут обрабатывать большие объёмы данных и выявлять сложные паттерны. Нейросети могут автоматически извлекать признаки из данных, что делает их мощным инструментом для обнаружения аномалий, особенно в больших и сложных наборах данных.
4. **Генеративные модели,** такие как GAN и VAE, описанные ранее, которые могут моделировать нормальное распределение данных и выявлять отклонения. Они предлагают новый подход к обнаружению аномалий, позволяя системам обучаться на нормальном трафике и выявлять отклонения без необходимости меток

Выбор метода зависит от множества факторов, включая тип и объём данных, требуемую точность и скорость обработки, а также доступные вычислительные ресурсы. Сравнение методов позволяет специалистам по информационной безопасности выбрать наиболее эффективный инструмент для защиты сетевой инфраструктуры от аномалий и угроз.

# Раздел 2. Методы классификации атак. Роль глубокого обучения в предотвращении атак.

## Методы классификации с использованием методов машинного обучения

Методы классификации в машинном обучении относятся к области обучения с учителем, где модель обучается предсказывать принадлежность объектов к определенным классам на основе обучающих данных. Основная цель методов классификации состоит в том, чтобы построить модель, способную автоматически отнести новые данные к одному из заранее определенных классов.

### *Задачи классификации*

* Обнаружение спама: классификация электронных писем или сообщений на спам и не-спам.
* Медицинская диагностика: классификация медицинских изображений или результатов анализов для определения наличия заболеваний.
* Финансовый мониторинг: классификация финансовых транзакций для выявления мошеннической активности.
* Обнаружение вторжений: классификация сетевого трафика для выявления потенциальных атак и вторжений в компьютерные системы.

Методы классификации могут быть реализованы с использованием различных алгоритмов, таких как логистическая регрессия, метод опорных векторов (SVM), деревья решений, случайный лес, нейронные сети и другие. Каждый из этих методов имеет свои уникальные особенности, преимущества и недостатки, что позволяет выбрать наиболее подходящий метод в зависимости от характеристик данных и требований конкретной задачи.

### Основные методы классификации

* Логистическая регрессия:

Логистическая регрессия используется для бинарной классификации, когда необходимо разделить данные на два класса. В контексте обнаружения вторжений она может быть применена для определения, является ли сетевая активность нормальной или подозрительной. Логистическая регрессия обладает простотой и интерпретируемостью, что позволяет использовать ее для первичного анализа данных и выявления важных признаков.

* Метод опорных векторов (Support Vector Machines, SVM):

SVM является мощным алгоритмом классификации, который строит гиперплоскость в многомерном пространстве, разделяющую данные на различные классы. В задаче обнаружения вторжений SVM может быть использован для выявления аномального поведения в сети. Метод опорных векторов особенно хорошо работает с разделением классов в многомерном пространстве признаков, что делает его эффективным для обнаружения сложных атак.

* Деревья решений и случайный лес (Decision Trees and Random Forest):

Деревья решений разделяют данные на основе серии вопросов о признаках, что делает их понятными для интерпретации. Случайный лес - это ансамбль деревьев решений, который позволяет улучшить качество классификации за счет комбинации нескольких деревьев. Деревья решений и случайный лес позволяют обрабатывать большие объемы данных и выявлять сложные зависимости между признаками.

* Нейронные сети:

Глубокие нейронные сети стали широко применяться в задачах классификации, включая обнаружение и предотвращение вторжений. Они способны автоматически извлекать признаки из данных и строить сложные модели для выявления аномалий. Нейронные сети, в свою очередь, способны выявлять нелинейные закономерности в данных, что делает их мощным инструментом для обнаружения разнообразных видов атак.

* Наивный байесовский классификатор:

Этот метод основан на применении теоремы Байеса и предполагает независимость признаков. Он может быть использован для выявления аномальных паттернов в данных.

* Алгоритмы кластеризации:

В задаче обнаружения вторжений кластеризация может быть полезна для выявления групп подозрительной активности, не требуя заранее определенных классов.

### Преимущества и недостатки самых распространённых методов классификации

Для наглядного представления и удобства использования в будущем была подготовлена таблица с основными преимуществами и недостатками методов классификации:

|  |  |
| --- | --- |
| **Преимущества** | **Недостатки** |
| **Логистическая регрессия** | |
| Простота интерпретации и реализации | Не способна моделировать сложные нелинейные зависимости |
| Эффективность на больших наборах данных | Неэффективна, если данные нелинейно разделимы |
| Малое количество гиперпараметров для настройки |  |
| **Метод опорных векторов (SVM)** | |
| Хорошо работает с высокоразмерными данными | Требует тщательной настройки гиперпараметров |
| Эффективен в пространствах большой размерности | Медленнее при обучении на больших наборах данных |
| Может обрабатывать нелинейные данные с помощью ядер | Неэффективен при работе с большими наборами данных |
| **Деревья решений и случайный лес** | |
| Хорошо интерпретируемы и понятны | Склонны к переобучению, особенно деревья решений |
| Могут работать с категориальными данными без предварительной обработки | Случайный лес может быть сложен для настройки гиперпараметров |
| Эффективны при обработке больших объемов данных |  |
| **Нейронные сети** | |
| Способны моделировать сложные зависимости в данных | Требуют большого объема данных для обучения |
| Могут автоматически извлекать признаки из данных | Сложны в интерпретации |
| Эффективны на больших наборах данных | Могут быть вычислительно затратными |

При рассмотрении методов обнаружения и предотвращения вторжений в компьютерные системы, были выявлены различные подходы, используемые для обработки и анализа данных с целью выявления потенциальных атак: методы классификации, такие как логистическая регрессия, метод опорных векторов (SVM), деревья решений, случайный лес и нейронные сети. Они играют важную роль в обнаружении вторжений. Они позволяют автоматически классифицировать сетевой трафик или системные события на нормальные и аномальные, что способствует выявлению и устранению потенциальных угроз безопасности автоматически.

При анализе методов обнаружения вторжений было выявлено, что комбинация различных методов классификации может повысить точность и надежность системы обнаружения вторжений за счет компенсации недостатков отдельных методов. Можно использовать ансамблирование методов классификации (например, случайный лес или градиентный бустинг) для повышения общей производительности системы обнаружения вторжений.

Рассмотрев основные методы классификации атак, можно сказать, что методы весьма похожи на методы машинного обучения для обнаружения аномальной активности и атак в сети. Отличия присутствуют: в алгоримах, в логике работы методов, однако классификация и обнаружение изменений и выбросов как-никак очень смежные понятия, а следовательно существует возможность объединения этих функций в один полигенный механизм защиты.

## Роль глубокого обучения в предотвращении кибератак

Глубокое обучение — это подразделение машинного обучения, которое использует искусственные нейронные сети для обучения на сложных иерархических представлениях данных. Глубокое обучение имеет широкий спектр применений, включая обработку изображений, распознавание речи, анализ текста и многое другое. В последние годы глубокое обучение стало важным инструментом в борьбе с кибератаками, так как оно способно обнаруживать и анализировать сложные атаки, которые могут обойти традиционные методы защиты. В данном докладе рассмотрим роль глубокого обучения в предотвращении кибератак и какие преимущества оно может предоставить.

Глубокое обучение может быть эффективным средством предотвращения кибератак за счет его способности анализировать сложные и многомерные данные и выявлять необычные или аномальные паттерны, которые могут указывать на активности хакеров или зловредное поведение. Глубокие нейронные сети могут использоваться для обнаружения атак на основе данных с различных источников, включая сетевой трафик, журналы событий, системные журналы и многое другое. Применение глубокого обучения в предотвращении кибератак позволяет создавать системы безопасности, которые могут адаптироваться к новым угрозам и защищать информационные ресурсы от различных видов атак. Преимущества использования глубокого обучения в борьбе с кибератаками включают высокую точность обнаружения, способность обнаруживать скрытые и сложные угрозы и скорость реагирования на потенциальные атаки. Глубокое обучение может обучаться на больших объемах данных и адаптироваться к изменяющимся условиям, что делает его эффективным инструментом для предотвращения и минимизации ущерба от кибератак. Кроме того, глубокое обучение может автоматически анализировать данные и принимать решения в реальном времени, что позволяет оперативно реагировать на угрозы и предотвращать потенциальные атаки.

## Роль больших данных и предварительной обработки в эффективности IDS

В International Data Spaces (IDS), большие данные и предварительная обработка играют важную роль в обеспечении эффективности и безопасности обмена данными. IDS-RAM (Reference Architecture Model) обеспечивает стандартизацию ролей и взаимодействий между ними, включая контракты и политики использования данных. Для эффективного обмена данными, необходимо обеспечить:

* **Trust**: через управление идентификацией и сертификацию пользователей.
* **Security and Data Sovereignty**: через аутентификацию и авторизацию, политики использования данных, защищенный обмен данными и техническую сертификацию.
* **Data Ecosystem**: через описание источников данных, брокераж данных и использование ватермейкеров для метаданных.

Большие данные и предварительная обработка помогают в обеспечении этих аспектов, включая:

* **Data Preprocessing**: для обеспечения качества и точности данных, что является важным для эффективного обмена данными.
* **Data Analytics**: для анализа данных и обнаружения тенденций, что помогает в принятии решений и улучшении бизнес-операций.

Большие данные и предварительная обработка играют ключевую роль в обеспечении эффективности и безопасности обмена данными в рамках International Data Spaces (IDS). Концепция IDS направлена на создание децентрализованной, основанной на доверии экосистемы для обмена данными между организациями, при этом сохраняя контроль над данными и обеспечивая их безопасность. Большие данные и предварительная обработка являются неотъемлемой частью этой концепции. Предварительная обработка данных необходима для обеспечения качества, точности и стандартизации данных, которыми обмениваются участники IDS. Это включает в себя очистку данных от ошибок и шума, преобразование в единый формат, обогащение метаданными и т.д.

Качественные данные являются основой для эффективного обмена и использования в бизнес-процессах. Большие данные позволяют масштабировать обмен данными и извлекать ценные инсайты. Технологии больших данных, такие как распределенные файловые системы, NoSQL базы данных и фреймворки для параллельных вычислений, позволяют обрабатывать огромные объемы разнородных данных в режиме реального времени. Это дает возможность выявлять скрытые закономерности, прогнозировать тенденции и принимать более обоснованные решения на основе данных.

Предварительная обработка и аналитика больших данных тесно интегрированы в архитектуру IDS. Модель референтной архитектуры IDS (IDS-RAM) определяет роли и взаимодействия между участниками экосистемы, включая поставщиков данных, брокеров, аналитиков и потребителей. Она также описывает механизмы обеспечения безопасности, доверия и суверенитета данных, такие как управление идентификацией, аутентификация, авторизация и политики использования данных.

В целом, большие данные и предварительная обработка являются фундаментальными компонентами IDS, обеспечивающими эффективный, безопасный и контролируемый обмен данными между организациями. Они позволяют масштабировать обмен данными, повышать их качество и извлекать ценные инсайты для улучшения бизнес-процессов.

## Применение глубокого обучения нейронных сетей для обнаружения сложных атак

Сложные атаки, такие как DDoS, APT (Advanced Persistent Threats), и другие виды кибератак, становятся все более сложными и трудными для обнаружения. Глубокое обучение, благодаря своей способности к обучению на больших объемах данных и распознаванию сложных шаблонов, становится ключевым инструментом в борьбе с такими угрозами. В этой статье рассматриваются способы применения глубоких нейронных сетей и методы передаточного обучения для обнаружения сложных атак.

Использование глубоких нейронных сетей

### CNN для анализа сетевого трафика

Сверточные нейронные сети (Convolutional Neural Networks, CNN) широко используются для обработки изображений, но их возможности не ограничиваются только этим. CNN могут быть эффективно применены для анализа сетевого трафика, где данные представлены в виде временных рядов или сигналов. CNN способны выявлять сложные паттерны и аномалии в трафике, что делает их полезными для обнаружения атак.

### RNN и LSTM для обработки временных рядов

Рекуррентные нейронные сети (RNN) и их вариации, такие как LSTM (Long Short-Term Memory), идеально подходят для анализа временных рядов, что делает их идеальным выбором для анализа сетевого трафика. Эти сети способны учиться на основе последовательностей данных, что позволяет им учитывать временные зависимости в трафике и обнаруживать аномалии, основываясь на изменениях в паттернах трафика во времени.

### Методы передаточного обучения и настройки предварительно обученных моделей

Передаточное обучение (Transfer Learning) позволяет использовать предварительно обученные модели для задач, отличных от тех, на которых они были первоначально обучены. Это особенно актуально для задач обнаружения вторжений, где доступные данные ограничены. Предварительно обученные модели, такие как VGG16, ResNet, и другие, могут быть адаптированы для анализа сетевого трафика, что существенно сокращает время и ресурсы, необходимые для обучения модели с нуля.

Настройка предварительно обученных моделей включает в себя замену верхних слоев модели на новые, специфичные для задачи обнаружения вторжений, и дальнейшее обучение модели на новом наборе данных. Этот подход позволяет сочетать универсальность предварительно обученных моделей с специфичностью для конкретной задачи, что повышает эффективность обнаружения сложных атак.

Глубокое обучение открывает новые горизонты в обнаружении сложных атак, предоставляя инструменты для анализа и обработки сложных паттернов в сетевом трафике. Использование CNN, RNN, LSTM и методы передаточного обучения позволяют эффективно обнаруживать и реагировать на различные виды кибератак, обеспечивая безопасную работу сетевых систем.

# Раздел 3. Анализ работы и стабильность ML-систем в сфере кибербезопасности. Гуманитарные и этические аспекты.

## Анализ уязвимостей и угроз, связанных с применением методов машинного обучения в области кибербезопасности.

В современном цифровом мире, где технологии машинного обучения становятся все более распространенными и востребованными, вопросы кибербезопасности приобретают особую важность. Применение методов машинного обучения открывает новые возможности для обнаружения и предотвращения киберугроз, однако это с собой несет и ряд уязвимостей. Анализ этих рисков является ключевым аспектом обеспечения безопасности информационных систем и данных. Важной задачей является изучение и анализ уязвимостей, связанных с использованием методов машинного обучения в киберпространстве, а также в выявлении возможных угроз и способов их предотвращения. Понимание вызовов, связанных с применением методов ML, позволит разработать эффективные стратегии защиты и повысить уровень безопасности в цифровом пространстве.

Анализируя доступные источники, а также рассматривая работу систем на базе машинного обучения можно выделить и разобрать следующий список угроз.

### **С***остязательные атаки на модели машинного обучения*

Состязательные атаки представляют собой специальные воздействия на элементы конвейера, направленные на манипуляцию поведением моделей. Целью таких атак является создание обученной машины, где атакующий ищет набор незаметных изменений во входных данных, чтобы модель неправильно классифицировала их. Типичный пример состязательной атаки: изменение некоторых пикселей в изображении перед загрузкой, чтобы система распознавания изображений не могла классифицировать результат. Эти незаметные для человека изменения, называемые состязательными примерами, могут обмануть систему машинного обучения.

Состязательные атаки могут быть направлены на различные типы данных, включая изображения, тексты, аудио-данные и временные ряды. Они могут быть использованы для обмана систем распознавания лиц, спам-фильтров, а также для других целей, где необходимо изменить результаты классификации. Существует несколько методов проведения состязательных атак, таких как FGSM (Fast Gradient Sign Method), DeepFool и JSMA (Jacobian-based Saliency Map Attack). Они различаются по эффективности создания состязательных примеров, количеству возмущений и вычислительным ресурсам. Для защиты от состязательных атак применяются методы состязательного машинного обучения, которые помогают выявлять и устранять аномалии в моделях искусственного интеллекта. Однако злоумышленники также могут использовать эти методы в своих целях, например, для обхода систем безопасности.

### Защита от атак на систему обнаружения вторжений

Для того, чтобы убедиться в устойчивости системы к атакам следует провести тщательные тесты. Вот некоторые из них:

* Тестирование на проникновение:

Попытка обойти систему обнаружения вторжений путем внедрения зловредного кода через различные уязвимости, такие как SQL-инъекции, кросс-сайтовый скриптинг и другие.

Проверка реакции системы обнаружения вторжений на попытки эксплойта уязвимостей, включая известные и новые уязвимости.

* Тестирование на эмуляцию атак:

Имитация различных типов атак, таких как отказ в обслуживании (DDoS), переполнение буфера, а также использование инструментов для эмуляции поведения злоумышленников.

Проверка реакции системы на аномальное поведение, которое может указывать на скрытые атаки.

* Тестирование на обход системы обнаружения вторжений:

Попытка использования различных методов для обхода системы обнаружения вторжений, таких как маскировка трафика, изменение характеристик пакетов и другие техники.

* Тестирование на анализ ложноположительных срабатываний:

Создание сценариев, которые могут привести к ложным срабатываниям системы обнаружения вторжений, например, необычное поведение легитимных пользователей или приложений.

Проверка способности системы отличать нормальное поведение от потенциально вредоносного.

* Тестирование на обновление сигнатур и правил:

Проверка процесса обновления сигнатур и правил системы обнаружения вторжений, включая скорость реакции на новые угрозы и эффективность обновлений.

* Тестирование на мониторинг и журналирование:

Проверка способности системы обнаружения вторжений к мониторингу и журналированию событий, включая анализ логов на предмет необычной активности.

* Тестирование на анализ трафика:

Анализ трафика с использованием различных методов шифрования и маскировки, чтобы проверить способность системы обнаружения вторжений распознавать подобные техники.

Это лишь некоторые из возможных тестовых сценариев для защиты от атак на систему обнаружения вторжений. В реальности проведение тестирования должно быть более детализированным и охватывать различные аспекты безопасности и функциональности системы.

### **Предвзятость и необъяснимость моделей ИИ**

Одной из ключевых угроз, связанных с применением методов машинного обучения, является предвзятость и необъяснимость систем. Предвзятость может возникать из-за особенностей входных данных, на которых обучается модель, или быть искусственно внесенной в процессе разработки. Например, система распознавания лиц может неверно идентифицировать темнокожих пользователей, если в обучающих данных недостаточно представлены лица разных рас.

Проблема усугубляется тем, что современные сложные модели ИИ работают по принципу "черного ящика" - их внутренние механизмы принятия решений становятся непостижимыми для человека. Это затрудняет диагностику предвзятости и исправление ошибок в работе таких систем. Кроме того, предвзятость может быть направленно внесена злоумышленниками для манипуляции поведением модели. Пример: Алгоритм оценки нуждаемости пациентов в медицинской помощи считал чернокожего пациента менее нуждающимся, чем белого, даже если у первого было больше объективных причин для получения помощи. Причина - в данных для обучения алгоритма были заложены расовые предубеждения, существующие в обществе. Но в коде самого алгоритма не было прямого указания отдавать предпочтение белым пациентам. Для противодействия угрозе предвзятости и необъяснимости моделей ИИ необходимо:

* Тщательно анализировать данные для обучения на наличие предвзятости
* Применять методы объяснимого искусственного интеллекта (XAI), позволяющие понять логику принятия решений моделью
* Разрабатывать модели с соблюдением принципов прозрачности и интерпретируемости
* Только комплексный подход к обеспечению безопасности и прозрачности моделей машинного обучения позволит минимизировать угрозы, связанные с их предвзятостью и необъяснимостью.

### Недостаток данных и ложные срабатывания

Недостаточное количество или некачественные обучающие данные могут приводить к тому, что модель будет неправильно классифицировать события и ошибочно срабатывать на безопасные ситуации. Проблема недостатка данных особенно актуальна для редких или новых типов атак, когда в распоряжении разработчиков моделей машинного обучения нет достаточного количества примеров для обучения. Это может привести к тому, что модель будет плохо обобщать и ошибаться при классификации подобных атак в реальных условиях.

Система обнаружения вторжений может пропускать новые типы вредоносных действий, если они не были представлены в обучающих данных. Помимо количества, важное значение имеет качество обучающих данных. Если в них присутствуют ошибки, неточности или шумы, модель машинного обучения будет воспринимать их как нормальные ситуации и ложно классифицировать реальные угрозы. Например, система обнаружения аномалий в сетевом трафике может ошибочно считать вредоносные действия нормальным трафиком, если в обучающих данных присутствовали зашумленные или некорректно размеченные примеры. Для противодействия угрозе ложных срабатываний необходимо:

* Применять методы контроля качества данных и мониторинга работы моделей
* Использовать ансамблевые методы, которые более устойчивы к недостатку данных
* Проводить регулярное тестирование моделей на новых типах данных

### *Время реагирования на угрозу*

Известно, что современные кибератаки очень быстро переходят от эксплуатации уязвимости к развертыванию вредоносных действий. Раньше, прежде чем начать атаку, злоумышленникам приходилось вручную проверять все слабые места и обходными путями выводить из строя системы безопасности – иногда этот процесс мог занимать недели. Однако, с развитием технологий, время, необходимое для развертывания атаки, значительно сократилось. Теперь злоумышленники могут автоматизировать процесс поиска уязвимостей и запуска вредоносного кода. Это ставит под угрозу эффективность традиционных систем обнаружения вторжений и реагирования на инциденты, основанных на правилах. Такие системы часто не успевают обновлять свои сигнатуры и правила быстрее, чем развиваются атаки. Модели машинного обучения, в свою очередь, также сталкиваются с проблемой времени реагирования. Для того, чтобы обучить ее распознавать новые типы атак, требуется время на сбор данных, разметку и переобучение. Кроме того, развертывание обновленной модели в рабочую среду также занимает время. Это создает окно уязвимости, которым могут воспользоваться злоумышленники. Для минимизации данной угрозы необходимо:

* Также использовать ансамблевые методы, более устойчивые к новым типам атак
* Применять методы трансферного обучения для ускорения адаптации моделей к новым данным
* Автоматизировать процесс переобучения и развертывания моделей для сокращения времени реагирования
* Использовать методы обнаружения аномалий для выявления неизвестных угроз

### *Необходимость обеспечения объяснимости системных решений ИИ*

Модели ИИ, особенно основанные на глубоком обучении, часто работают по принципу "черного ящика", когда их внутренняя логика становится непрозрачной для человека. Это создает серьезные проблемы в области кибербезопасности, где требуется объяснение и обоснование решений систем ИИ. Без понимания логики работы таких систем, специалистам сложно выявлять ошибки, манипуляции и атаки, которые могут быть в них встроены. Кроме того, непрозрачность ИИ снижает доверие к таким системам со стороны пользователей и регуляторов.

Для решения этой проблемы необходимо обеспечить объяснимость системных решений. Применяются специальные методы, такие как визуализация признаков, генерация контрфактических примеров и использование интерпретируемых моделей. Эти подходы позволяют "заглянуть внутрь" черного ящика нейронных сетей и понять логику их работы. Обеспечение объяснимости является важным условием для успешного внедрения. Только объяснимые системы ИИ способны завоевать доверие специалистов и обеспечить прозрачность их функционирования. Недостаточная объяснимость решений представляет собой серьезную уязвимость, которую необходимо устранять для обеспечения безопасности информационных систем.

## Реальная обстановка ML-систем в сфере информационной безопасности.

Внедрение систем машинного обучения несет ряд важных преимуществ:

* Алгоритмы способны анализировать большие объемы данных и выявлять сложные, ранее неизвестные атаки, основываясь на поведенческих моделях. Это позволяет быстрее реагировать на инциденты и предотвращать ущерб.
* ML позволяет автоматизировать многие рутинные процессы в сфере кибербезопасности, такие как анализ журналов событий, обнаружение аномалий, классификация вредоносных файлов. Это высвобождает время специалистов для решения более сложных задач.
* Способность выявлять тонкие закономерности и связи в данных, недоступные человеческому анализу. Это позволяет более точно определять риски и принимать обоснованные решения по защите.
* ML-системы могут самообучаться на основе новых данных, постоянно совершенствуя свои модели выявления угроз. Это критически важно в условиях быстро меняющегося ландшафта киберугроз.
* Способность обрабатывать огромные объемы данных и обслуживать тысячи пользователей, сохраняя высокую производительность. Это позволяет внедрять их в крупных распределенных инфраструктурах.

Однако, чтобы систему можно было действительно внедрить и использовать нужно будет решить все ранее описанные проблемы. Именно из-за того, что присутствует необходимость в разрешении или, как минимум, минимизации каждой уязвимости, сегодня сложно найти кейс действительного и полноценного использования ИИ в сфере информационной безопасности, зачастую это будут редкие случаи, применяемые по большей мере для частных автоматизаций тех или иных процессов. Внедрение ML в сферу информационной безопасности открывает новые возможности для повышения эффективности защиты от киберугроз, однако важно помнить, что система не заменяет человеческий анализ, а лишь служит удобным инструментом в руках мастера, позволяя специалистам сосредоточиться на решении наиболее сложных задач.

## Проблемы безопасности данных и этические аспекты использования больших данных и методов машинного обучения в киберзащите:

### Конфиденциальность данных при обработке информации о вторжениях

Для эффективного обучения моделей машинного обучения требуются большие объемы разнообразных данных. Однако сбор и использование персональных данных граждан регулируется строгими законодательными нормами. Законы о конфиденциальности устанавливают ряд требований к обработке персональных данных, включая получение согласия субъекта, ограничение целей использования, обеспечение безопасности и другие. Несоблюдение этих требований может повлечь за собой серьезные штрафы и репутационные потери для организаций. В то же время, для достижения высокой точности моделей машинного обучения часто требуется использовать данные, полученные без явного согласия или в обход ограничений. Условно, система обнаружения мошенничества в финансовых транзакциях может значительно повысить свою эффективность, если будет использовать данные о транзакциях клиентов из разных банков. Однако передача и объединение таких данных между банками может противоречить законам о банковской тайне и защите персональных данных.

В федеральном законе «Об информации, информационных технологиях и защите информации» говорится, что **конфиденциальность информации** – это обязательное для выполнения лицом, получившим доступ к какой-либо информации, требование, не передавать данную информацию третьим лицам без согласия её обладателя. Само же слово «Конфиденциальность» произошло от английского «Конфидэнс», в переводе это означает «доверие». Именно потому, что конфиденциальность информации соблюдается, человек, доверив свои персональные данные государственным органам, может быть уверен, что они не попадут в чужие руки и не будут использованы против него. Так же конфиденциальность информации имеет огромное значение при сохранении государственной и коммерческой тайны.

Конфиденциальность данных при обработке информации о вторжениях — это крайне важный аспект в современном мире, где цифровые технологии играют все более значимую роль в повседневной жизни людей и функционировании организаций. В связи с постоянным развитием информационных технологий и увеличением количества цифровых данных, вопрос защиты конфиденциальности становится все более актуальным.

Особенно важно обеспечить конфиденциальность данных при обработке информации о вторжениях, так как эта информация может содержать чувствительные данные о безопасности, уязвимостях систем, методах защиты и других аспектах, которые могут быть использованы злоумышленниками для нанесения ущерба. Поэтому необходимо принимать все возможные меры для защиты такой информации.

Для обеспечения конфиденциальности данных при обработке информации о вторжениях необходимо соблюдать ряд мер безопасности. В частности, это включает в себя шифрование данных, установку средств защиты информации, контроль доступа к данным, мониторинг и аудит безопасности системы, обучение сотрудников правилам безопасности и многое другое. В любом случае, при возможности стоит избегать использование личных данных или данных, представляющих собой тайну (личную/государственную/коммерческую).

В случае, если использование конфиденциальных данных необходимо, то для минимизации вероятности утечки и бесправного использования конфиденциальных данных необходимо:

* Анализировать соответствие используемых данных требованиям законодательства о конфиденциальности
* Применять методы обезличивания и анонимизации данных для снижения рисков
* Получать явное согласие субъектов на использование их персональных данных в целях обучения моделей
* Ограничивать доступ к персональным данным в процессе разработки и эксплуатации систем

### **Угрозы взлома и нарушения конфиденциальности**

Стоит отметить ещё одну возможную уязвимость ML систем, направленную на кражу конфиденциальных данных. Инверсия модели представляет собой метод атаки на системы машинного обучения, который может привести к восстановлению информации об обучающих данных и нарушению конфиденциальности.

Этот вид атаки основан на обратном инжиниринге обучения модели, позволяя злоумышленнику извлекать подробности об обучающих данных, таких как изображения лиц или текстовые сообщения. Процесс инверсии модели начинается с обучения инверсионной модели, которая затем используется для восстановления информации об обучающих данных, проходивших через модель машинного обучения.

Злоумышленник может начать с выходных данных модели, постепенно раскрывая внутреннюю структуру и веса, чтобы восстановить оригинальные входные данные. Это позволяет злоумышленнику получить доступ к конфиденциальным данным, использованным для обучения модели, что может привести к серьезным нарушениям конфиденциальности и утечкам информации.

### Этические соображения при использовании данных для предотвращения вторжений

Этические соображения играют важную роль в использовании данных для предотвращения вторжений. В современном мире, когда технологии безопасности становятся все более важными, необходимо учитывать ряд этических принципов при сборе, хранении и использовании данных для защиты от кибератак.

1. Приватность и конфиденциальность данных: при использовании данных для предотвращения вторжений необходимо строго соблюдать приватность и конфиденциальность личной информации. Это включает в себя защиту личных данных пользователей, соблюдение законодательства о защите данных и обеспечение безопасности информации.

2. Прозрачность и согласие: важно уведомлять пользователей о том, как их данные используются для предотвращения вторжений. Четкое информирование и получение согласия на использование данных помогает поддерживать доверие и уважение к приватности.

3. Минимизация сбора данных: Сбор только необходимых данных для обеспечения безопасности, а также их хранение и использование в соответствии с целями предотвращения вторжений, помогает сократить риск нарушения приватности.

4. Безопасность данных: Обеспечение безопасности хранения и передачи данных является ключевым аспектом этически правильного использования информации для защиты от вторжений. Защита от утечек, взломов и несанкционированного доступа к данным является обязательной задачей.

5. Использование анонимизированных данных: в некоторых случаях можно использовать анонимизированные данные для обучения систем обнаружения вторжений, что позволяет сохранить конфиденциальность личных данных.

6. Ответственное использование аналитики: при использовании аналитики для предотвращения вторжений необходимо учитывать возможные последствия и не допускать злоупотребления полученной информацией.

7. Соблюдение законодательства: важно соблюдать законы и нормативные акты, регулирующие сбор, хранение и использование данных в целях безопасности.

8. Обучение персонала: Обучение персонала по этическим аспектам работы с данными для предотвращения вторжений помогает обеспечить правильное использование информации и соблюдение принципов конфиденциальности.

Этические соображения в использовании данных для предотвращения вторжений являются ключевыми для поддержания доверия пользователей, соблюдения законодательства и обеспечения безопасности личной информации. Соблюдение этических принципов помогает создать более надежные и устойчивые системы защиты от киберугроз.

## Техники предотвращения вторжений:

### Адаптивные механизмы защиты

Методы, используемые IDS/IPS-системами для обнаружения и предотвращения вторжений, являются таким же важным фактором, как и тип системы.  
Основные методы обнаружения:

* Сигнатурный метод. Системы обнаружения вторжений на основе сигнатур (Signature-Based Intrusion Detection Systems, SIDS) предназначены для выявления шаблонов и сопоставления их с известными признаками вторжений — сигнатурами. SIDS опирается на базу данных предыдущих вторжений. Для эффективной работы SIDS требуется регулярное обновление базы сигнатур, поскольку SIDS может выявлять только те атаки, которые она распознает. От новых угроз, ранее неизвестных такой IDS-системе, SIDS не защитит.
* Обнаружение аномалий. Системы обнаружения вторжений на основе аномалий (Anomaly-Based Intrusion Detection System, AIDS) используют машинное обучение и статистические данные для создания модели «нормального» поведения пользователей и системы. Любое отклонение от этой нормы считается аномалией. Создание этих базовых профилей занимает много времени (также известное как «период обучения»), но даже и после обучения возможны ложные срабатывания, особенно в динамичных средах. Обнаружение аномалий показало себя эффективнее обнаружения на основе сигнатур при рассмотрении новых атак, которых нет в базе данных сигнатур.
* Анализ протокола состояний. Система обнаружения вторжений распознает расхождения протоколов путём сравнения наблюдаемых событий с предварительно построенными профилями общепринятых определений безопасной активности.
* Гибридная система обнаружения вторжений. Большинство IDP-систем предлагают комбинацию нескольких подходов к обнаружению. Просматривая шаблоны и разовые события, гибридная система обнаружения вторжений может обнаруживать новые и известные стратегии вторжений. Единственным недостатком такой IDS является еще больший рост числа срабатываний. Однако, учитывая, что цель IDS — помечать потенциальные вторжения, трудно рассматривать увеличение числа подозрительных активностей с негативной точки зрения.

Методы предотвращения вторжений. К ним относятся:

* Прекращение атаки. Системы предотвращения вторжений устраняют инциденты безопасности до того, как они произойдут. Это делается путем блокировки пользователей или трафика, исходящего с определенного IP-адреса. Данный метод также включает в себя прекращение или сброс сетевого подключения, а также отзыв прав доступа пользователя до выяснения подробностей инцидента.
* Модификация контента атаки. Контент может быть внедрен в систему в различных формах. Один из способов сделать его более безопасным — удалить вредоносные сегменты. Простой пример — блокировка подозрительных вложений в сообщениях электронной почты.
* Изменение среды безопасности. Данный метод включает в себя изменение конфигурации безопасности для предотвращения атак. Примером может служить перенастройка параметров межсетевого экрана по требованию IPS для блокировки определенного IP-адреса или порта.
* Блокирование шаблонов трафика. Система предотвращения вторжений блокирует шаблоны трафика, которые могут быть использованы для атаки, чтобы предотвратить возможность использования уязвимых

Адаптивные механизмы защиты представляют собой подход к обеспечению безопасности информации, который активно реагирует на изменяющиеся угрозы и атаки, а также адаптируется к новым методам вторжения. Эти механизмы разрабатываются для того, чтобы обнаруживать и предотвращать вторжения в реальном времени, минимизируя вред и ущерб, который может быть причинен системе или данным.

Одной из ключевых характеристик адаптивных механизмов защиты является способность обучаться на основе опыта и данных о предыдущих атаках. Это позволяет системе постоянно совершенствоваться, адаптироваться к новым видам угроз и повышать эффективность защиты. Важными компонентами адаптивных механизмов защиты являются:

* Машинное обучение и искусственный интеллект: Алгоритмы машинного обучения используются для анализа больших объемов данных и выявления аномалий, связанных с потенциальными вторжениями. Искусственный интеллект позволяет системе самостоятельно принимать решения на основе полученных данных.
* Анализ поведения: Механизмы защиты могут анализировать типичное поведение пользователей и системы, чтобы выявлять аномальные или подозрительные действия, которые могут свидетельствовать о вторжении.
* Динамическое обновление правил: Адаптивные системы защиты могут автоматически обновлять свои правила и конфигурации на основе новых угроз и изменений в среде.
* Интеграция с угрозами в реальном времени: Механизмы защиты могут использовать информацию о текущих угрозах из различных источников, чтобы быстро реагировать на новые виды атак.
* Ответные меры: помимо предотвращения вторжений, адаптивные механизмы защиты могут также включать функциональность для реагирования на атаки, минимизируя ущерб в случае успешного вторжения.
* Облачные технологии: Использование облачных технологий позволяет создавать гибкие и масштабируемые системы защиты, способные быстро реагировать на изменяющиеся условия и угрозы.
* Разнообразие защитных слоев: Адаптивные механизмы защиты часто используют комбинацию различных методов и технологий для создания многоуровневой защиты, что повышает эффективность предотвращения вторжений.

Использование адаптивных механизмов защиты помогает компаниям и организациям более эффективно защищать свои системы и данные от постоянно меняющихся киберугроз.

### Методы обратной связи и обучения с подкреплением

Методы обратной связи и обучения с подкреплением играют важную роль в различных областях, включая машинное обучение, искусственный интеллект, робототехнику, автоматизацию и даже в кибербезопасности.

Обратная связь — это механизм, при котором информация о выполненном действии возвращается обратно к системе, что позволяет системе корректировать свое поведение на основе этой информации. В контексте предотвращения вторжений обратная связь может использоваться для анализа результатов предыдущих действий и принятия решений о том, какие шаги необходимо предпринять для улучшения безопасности системы.

Обучение с подкреплением — это метод машинного обучения, в котором агент (например, программа или робот) обучается через взаимодействие с окружающей средой. Агент получает награду или наказание за свои действия и использует эту информацию для оптимизации своего поведения. В контексте кибербезопасности обучение с подкреплением может применяться для создания адаптивных систем защиты, способных обучаться на основе новых угроз и атак.

Применение методов обратной связи и обучения с подкреплением в кибербезопасности:

* Обнаружение аномалий: Методы обратной связи могут использоваться для анализа поведения пользователей и системы и выявления аномальных действий, которые могут свидетельствовать о потенциальной угрозе.
* Адаптивные системы защиты: Обучение с подкреплением позволяет создавать адаптивные системы защиты, способные самостоятельно корректировать свои стратегии на основе новых данных о вторжениях.
* Управление доступом: Методы обратной связи могут применяться для анализа доступа пользователей к системе и определения правильных уровней доступа на основе их поведения.
* Оптимизация реакции на угрозы: Обучение с подкреплением позволяет системам защиты быстро адаптироваться к новым видам угроз и разрабатывать эффективные стратегии противодействия.
* Прогнозирование угроз: Методы обратной связи могут использоваться для прогнозирования возможных угроз на основе анализа предыдущих инцидентов и данных.

В целом, методы обратной связи и обучения с подкреплением играют важную роль в развитии инновационных подходов к предотвращению вторжений и повышению уровня безопасности информационных систем. Их применение позволяет создавать более гибкие, адаптивные и эффективные системы защиты, способные быстро реагировать на изменяющиеся условия и угрозы в киберпространстве.

### Применение шифрования для предотвращения вторжений

Применение шифрования играет важную роль в предотвращении вторжений и обеспечении безопасности информации. Шифрование позволяет защитить данные от несанкционированного доступа, искажения и кражи.

Виды шифрований:

* Шифрование данных в покое: Этот метод включает в себя защиту данных в хранилищах, на жестких дисках, в базах данных и в файлах. Применение алгоритмов шифрования позволяет защитить конфиденциальные данные от несанкционированного доступа, так как даже при получении доступа к хранилищу злоумышленнику будет сложно или практически невозможно расшифровать данные без ключа.
* Шифрование данных в движении: Этот метод включает в себя защиту данных при их передаче через сети. Применение шифрования (например, протоколы SSL/TLS) обеспечивает конфиденциальность и целостность передаваемых данных, что помогает предотвратить перехват и изменение информации злоумышленниками.
* Шифрование электронной почты: Шифрование электронной почты помогает защитить конфиденциальную информацию от несанкционированного доступа. При использовании криптографических методов шифрования сообщения могут быть защищены от прослушивания и чтения третьими лицами.
* Шифрование устройств: Защита данных на мобильных устройствах, ноутбуках и других портативных устройствах с помощью шифрования помогает предотвратить утечку конфиденциальной информации в случае утери или кражи устройства.
* Шифрование трафика с помощью виртуальных частных сетей (VPN): VPN-сервисы используют шифрование для защиты интернет-трафика от несанкционированного доступа и подслушивания, что особенно важно при работе с открытыми сетями, такими как общественные Wi-Fi сети.
* Шифрование файлов и директорий: Этот метод позволяет защитить конфиденциальные файлы и директории на уровне операционной системы, предотвращая несанкционированный доступ к данным.
* Использование цифровых подписей: Цифровые подписи используются для аутентификации отправителя и обеспечения целостности данных. Они помогают предотвратить подделку и изменение данных в процессе передачи.

Шифрование является ключевым инструментом в арсенале методов предотвращения вторжений, поскольку оно обеспечивает защиту конфиденциальности, целостности и доступности данных, что является основополагающими аспектами безопасности информации. Важно правильно выбирать и настраивать методы шифрования в зависимости от специфики системы и уровня конфиденциальности информации.

# Перспективы развития и будущее направление

IDS-RAM (Reference Architecture Model) является важным шагом в развитии международных данных и обеспечении безопасности и доверия в обмене данными. В будущем, IDS-RAM может развиваться в следующих направлениях:

* **Увеличение использования искусственного интеллекта**: для улучшения анализа данных и обнаружения тенденций.
* **Увеличение использования блокчейна**: для обеспечения безопасности и прозрачности в обмене данными.
* **Увеличение использования когнитивных технологий**: для улучшения анализа данных и обнаружения тенденций.

IDS-RAM также может быть интегрирован в другие инициативы, такие как Gaia-X, для создания более широкой и открытой платформы для обмена данными. В целом, IDS-RAM играет важную роль в развитии международных данных и обеспечении безопасности и доверия в обмене данными, и его развитие будет продолжаться в будущем.

Концепция International Data Spaces (IDS) и ее модель референтной архитектуры (IDS-RAM) находятся на начальном этапе развития, но уже демонстрируют большой потенциал для трансформации экосистемы обмена данными между организациями. В ближайшие годы можно ожидать дальнейшего развития и расширения IDS в следующих направлениях:

### Расширение экосистемы участников

IDS изначально разрабатывалась как открытая, основанная на стандартах концепция, чтобы привлечь широкий круг участников. В будущем мы увидим все больше организаций, присоединяющихся к IDS, включая крупные корпорации, малые и средние предприятия, государственные учреждения, научно-исследовательские институты и т.д. Это будет способствовать созданию более богатой и разнообразной экосистемы данных.

### Интеграция с другими инициативами

IDS-RAM может быть интегрирована с другими инициативами в области управления данными, такими как Gaia-X, Industrial Data Space, Data Spaces Business Alliance и др. Это позволит создать более широкую, открытую и взаимосвязанную платформу для обмена данными на международном уровне. Интеграция с облачными платформами также расширит возможности IDS.

### Развитие технологий

По мере развития IDS, мы увидим дальнейшее развитие технологий, лежащих в ее основе. Это включает в себя:

* Улучшение методов предварительной обработки и очистки данных с использованием машинного обучения
* Внедрение передовых технологий аналитики больших данных, таких как искусственный интеллект и когнитивные вычисления
* Использование блокчейна и других технологий распределенного реестра для обеспечения безопасности, прозрачности и неизменности транзакций с данными
* Разработка более совершенных механизмов управления идентификацией,

аутентификацией и авторизацией

### Расширение применения

IDS изначально разрабатывалась для промышленного сектора, но в будущем мы увидим расширение ее применения в других областях, таких как здравоохранение, финансы, розничная торговля, транспорт и логистика. Кроссотраслевое использование IDS позволит создать более универсальную и масштабируемую экосистему обмена данными. В целом, перспективы развития IDS выглядят многообещающими. По мере расширения экосистемы участников, интеграции с другими инициативами, развития технологий и применения в различных отраслях, IDS будет играть все более важную роль в трансформации способов обмена данными между организациями, обеспечивая безопасность, доверие и контроль над данными.

# Заключение. Подведение итогов.

По окончании прохождения практики, студенты получили новые знания в области применения методов машинного обучения в задаче обнаружения и предотвращения вторжений, а также рассмотрели сопутствующие темы. Также студенты получили практические навыки работы в команде с определением сроков, ответственных и контролем качества.

Итоговый литературный обзор (настоящий отчёт) может служить методическим материалом, способным первично раскрыть тему, осветить основные моменты темы. Также была подготовлена таблица, в которой сравниваются различные методы машинного обучения в различных задачах (приведены плюсы и минусы каждой модели) (см. Приложение №1).

# ПРИЛОЖЕНИЕ №1

|  |  |  |  |
| --- | --- | --- | --- |
| **Тип машинного обучения** | **Описание** | **Плюсы** | **Минусы** |
| Логистическая регрессия | Это контролируемый алгоритм машинного обучения, используемый для задач классификации, целью которых является предсказание вероятности принадлежности экземпляра к заданному классу или нет. Логистическая регрессия — это статистический алгоритм, который анализирует взаимосвязь между двумя факторами данных. | * Относительная простота алгоритма. * Интерпретируемость. * Хорошие результаты на небольших наборах данных. * Небольшая вероятность переобучения. | * Требуется нормализация признаков. * Работает плохо на сложных задачах. * Линейность. * Низкая точность. |
| Метод опорных векторов | Это алгоритм машинного обучения, применяемый для задач линейной и нелинейной классификации, регрессии и обнаружения аномальных данных. С помощью этого метода можно классифицировать текст, изображения, обнаружить спам, идентифицировать почерк, анализировать экспрессии генов, распознавать лица, делать прогнозы и так далее. | * Работа с большими пространствами признаков. * Работа с полу структурированными и неструктурированными данными. * Возможность использовать концепцию фокуса ядра для решения любой сложной задачи. | * Сложность реализации, если количество классов больше двух. * Требование много времени для обучения и чувствительность к шуму. * Непростой выбор хорошей функции ядра и необходимость проведения большого количества тестов. |
| Дерево решений | Это тип алгоритма контролируемого обучения, который обычно используется в машинном обучении для моделирования и прогнозирования результатов на основе входных данных. Это древовидная структура, в которой каждый внутренний узел проверяет атрибут, каждая ветвь соответствует значению атрибута, а каждый конечный узел представляет окончательное решение или прогноз. | Простота в понимании и интерпретации.  Возможность фиксировать нелинейные взаимосвязи между объектами и целевой переменной.  Нечувствительность к масштабу объектов.  Обработка как числовых, так и категориальных данных.  Устойчивость к выбросам в данных. | Подверженность переоснащению.  Высокая дисперсия.  Нестабильность.  Сложности с поддержкой пропусков в данных.  Модель умеет только интерполировать, но не экстраполировать. |
| Случайный лес | Это модель, которая представляет собой набор деревьев решений. | Высокая точность.  Устойчивость к переобучению.  Обработка пропущенных данных.  Интерпретируемость. | Высокие вычислительные затраты.  Сложность интерпретации.  Проблемы с высокоразмерными данными. |
| Нейронные сети | В машинном обучении нейронная сеть представляет собой модель, вдохновленную структурой и функциями биологических нейронных сетей в мозге животных. | Способность обрабатывать большие объёмы данных.  Создание и обучение не требует знания специфических физических или математических закономерностей.  Обработка информации в режиме реального времени.  Обучаемость на существующих данных. | Работа только с данными, которые были использованы при их обучении.  Создание и обучение занимают много времени и требуют значительных ресурсов, особенно при работе с крупными объёмами данных.  Результаты недостаточно точные и могут оказаться чрезмерно сложными для анализа.  Нейронные сети не могут объяснить свои решения  Коррекция ошибок может быть трудной процедурой. |
| Наивный байесовский классификатор | Это вероятностный алгоритм, используемый в машинном обучении для задач классификации. Классификатор основан на теореме Байеса (или формуле Байеса), которая позволяет определить вероятность гипотезы при имеющихся доказательствах. Вероятностный классификатор — простой алгоритм, который тем не менее эффективен в различных ситуациях. | простота в реализации и интерпретации  практически не требуется настройка параметров  высокая скорость работы и точность прогнозов во многих ситуациях  имеет относительно хорошую устойчивость к шуму и выбросам | в случае нарушения предположения о независимости признаков, точность прогнозов может значительно снизиться  может отдавать предпочтение к классам с бОльшим количеством образцов в случае несбалансированных данных  если переменная имеет категорию, которая не наблюдалась в обучающем наборе данных, то модель присвоит ей нулевую вероятность |
| Алгоритмы кластеризации | Это метод машинного обучения, который используется для группировки объектов в кластеры, так чтобы объекты внутри одного кластера были более похожи друг на друга, чем на объекты из других кластеров. Этот метод широко применяется в анализе данных, маркетинге, биоинформатике и других областях. Кластеризация помогает выявить скрытые структуры в данных и сделать их более понятными. | Простота. Многие алгоритмы кластеризации просты в реализации и использовании.  Гибкость. Кластеризация может применяться к различным типам данных и задач.  Выявление скрытых структур. Кластеризация помогает обнаружить скрытые структуры и паттерны в данных. | Чувствительность к параметрам.  Чувствительность к шуму.  Сложность интерпретации.  Ограничения алгоритмов. Некоторые алгоритмы кластеризации могут быть неэффективны при обработке больших объёмов данных или не учитывать определённые зависимости. |

# Список используемой литературы

* Применение нейронных сетей для обнаружения аномалий в сетевом трафике – Электрон. дан. – Режим доступа: <https://cyberleninka.ru/article/n/obnaruzhenie-anomaliy-trafika-s-ispolzovaniem-neyronnoy-seti-dlya-obespecheniya-zaschity-informatsii>
* Применение глубокого обучения нейронных сетей для обнаружения сложных атак - <https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-dlya-obnaruzheniya-setevyh-atak>
* Проблемы безопасности и этнические аспекты: [https://documentation.rsuh.ru/](https://documentation.rsuh.ru/jour/article/viewFile/36/37)
* Техники предотвращения вторжений: [https://ideco.ru/](https://ideco.ru/sistemy_obnaruzhenia_vtorzheniy_ids_ips)
* Применение методов машинного обучения в задаче обнаружения и предотвращения вторжений: [https://cyberleninka.ru/](https://cyberleninka.ru/article/n/primenenie-metodov-mashinnogo-obucheniya-v-sdn-v-zadachah-obnaruzheniya-vtorzheniy)
* Бабаш, А.В. Информационная безопасность. Лабораторный практикум: учеб. пособие / А.В. Бабаш, Е.К. Баранова, Ю.Н. Мельников. - М.: КНОРУС, 2013. - 432 с. [Электронный ресурс]. - URL: https://znanium.com/catalog/product/415026 <https://www.researchgate.net/publication/376376776_Metody_masinnogo_obucenia_dla_obnaruzenia_vtorzenij_v_setah_interneta_vesej>.
* Боровков, А.И. Математическая статистика: учебник / А.И. Боровков. - М.: Наука, 2010. - 688 с. [Электронный ресурс]. - URL: https://e.lanbook.com/book/2099
* Васильев, В.И. Интеллектуальные системы защиты информации / В.И. Васильев, Д.В. Поляков. - М.: Машиностроение, 2012. - 224 с. [Электронный ресурс]. - URL: <https://e.lanbook.com/book/5762>
* Вишняков, Я.Д. Общая теория рисков: учеб. пособие / Я.Д. Вишняков, Н.Н. Радаев. - 2-е изд., испр. - М.: Академия, 2008. - 368 с. [Электронный ресурс]. - URL: https://znanium.com/catalog/product/161541
* Исследование атак на модели машинного обучения в сетях 5G на основе генеративно-состязательных сетей / Д. И. Парфенов, И. П. Болодурина, Л. В. Легашев [и др.] // Прикаспийский журнал: управление и высокие технологии. – 2023. – № 1(61). – С. 89-96. – EDN VNNVJT.
* Намиот, Д. Е. О работе AI Red Team / Д. Е. Намиот, Е. В. Зубарева // International Journal of Open Information Technologies. – 2023. – Т. 11, № 10. – С. 130-139. – EDN NRSLMA.
* Осман, С. Ш. О. Перспективы искусственного интеллекта в системах кибербезопасности / С. Ш. О. Осман // Наукосфера. – 2023. – № 9-1. – С. 213-217. – EDN DRYSZN.
* Сычев, Д. И. Искусственный интеллект и кибербезопасность: будущие тенденции и вызовы / Д. И. Сычев // Международный журнал информационных технологий и энергоэффективности. – 2023. – Т. 8, № 5-2(31). – С. 9-14. – EDN FBZSBS.
* Ламонина, Л. В. К вопросу о применении искусственного интеллекта в обеспечении информационной безопасности / Л. В. Ламонина, О. Б. Смирнова // Электронный научно-методический журнал Омского ГАУ. – 2023. – № 3(34). – EDN MARYTD.
* Generative Adversarial Networks – Автор: Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio [Электронный ресурс] – Электрон. дан. – URL-ссылка: <https://arxiv.org/abs/1406.2661>
* Auto-Encoding Variational Bayes – Автор: Diederik P. Kingma, Max Welling [Электронный ресурс] – Электрон. дан. – URL-ссылка: <https://arxiv.org/abs/1312.6114>
* "Deep Learning" (2016) – Автор: Ian Goodfellow, Yoshua Bengio, Aaron Courville – Издательство: MIT Press
* Клишин Д. А., Чечулин А. А. «АНАЛИЗ СТАНДАРТОВ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ», 2020.
* A.V. Krutskikh, E.S. Zinovieva «International Information Security: Russia’s Approaches», 2021.
* [V. V. Arutyunov](https://link.springer.com/article/10.3103/S0147688217020071#auth-V__V_-Arutyunov-Aff1), «Clustering of information-security standarts», 2017.
* 3. Малыгин, А.Ю. Применение методов машинного обучения в системах обнаружения вторжений / А.Ю. Малыгин, Д.А. Петров // Вопросы кибербезопасности. - 2017. - № 5 (24). - С. 2-10. [Электронный ресурс]. - URL: <https://www.elibrary.ru/item.asp?id=32286186>
* 4. Рассоха, С.В. Интеллектуальные компоненты в системах обнаружения вторжений / С.В. Рассоха, С.Б. Бельченко // Вопросы кибербезопасности. - 2020. - № 2 (36). - С. 14-22. [Электронный ресурс]. - URL: <https://www.elibrary.ru/item.asp?id=42877790>
* 5. Васильев, В.И. Интеллектуальные системы защиты информации / В.И. Васильев, Д.В. Поляков. - М.: Машиностроение, 2012. - 224 с. [Электронный ресурс]. - URL: <https://e.lanbook.com/book/5762>
* 6. Елисеев, А.В. Интеллектуальные информационные системы: учебник / А.В. Елисеев. - М.: ИНФРА-М, 2018. - 232 с. [Электронный ресурс]. - URL: <https://znanium.com/catalog/product/954293>
* Методы машинного обучения для обнаружения вторжений в сетях интернета вещей [Электронный ресурс] – Электрон. дан. – Режим доступа: <https://www.researchgate.net/publication/376376776_Metody_masinnogo_obucenia_dla_obnaruzenia_vtorzenij_v_setah_interneta_vesej>.
* Применение методов машинного обучения в sdn в задачах обнаружения вторжений [Электронный ресурс] – Электрон. дан. – Режим доступа: <https://cyberleninka.ru/article/n/primenenie-metodov-mashinnogo-obucheniya-v-sdn-v-zadachah-obnaruzheniya-vtorzheniy>.
* Комплексная система защиты информации предприятия [Электронный ресурс] – Электрон. дан. – Режим доступа: <https://vsbdo.com/zashita_informazii>.
* Li, Yingjing, et al. "Deep learning for network security." IEEE Access 7 (2019): 49263-49278.
* Scaiano, Austin, et al. "Cyber Intrusion Detection: A Machine Learning Approach." arXiv preprint arXiv:2102.03933 (2021).
* Bhattacharya, Jyoti, and Soman Kp. "A Survey of Deep Learning Techniques for Cyber Security." Journal of King Saud University-Computer and Information Sciences (2020).