**Документация проекта: Анализ киберугроз**

**1. Введение**

С развитием технологий и увеличением объемов интернет-трафика кибератаки становятся все более сложными и массовыми. Этот проект посвящен анализу сетевых данных для выявления угроз и аномалий, таких как DDoS-атаки, фишинг или сканирование портов. Используя датасет "Cyber Security Attacks", мы изучаем связь между типами атак и параметрами сетевого трафика, чтобы предложить подходы к усилению киберзащиты.

**2. Цели**

1. Изучить поведение различных типов атак и их связь с сетевыми протоколами, размерами пакетов и источниками трафика.
2. Выявить аномалии в поведении отправителей и получателей трафика.
3. Предложить методы визуализации для более эффективного понимания сетевых угроз.

**3. Описание данных**

Датасет состоит из следующих столбцов:

* **Protocol:** Тип используемого интернет-протокола (например, TCP, UDP).
* **Packet Size:** Размер сетевых пакетов.
* **Source IP Address / Destination IP Address:** IP-адреса источника и получателя.
* **Sender ID / Receiver ID:** Уникальные идентификаторы отправителей и получателей.
* **Target Variable:** Категории атак, такие как Phishing, DDoS, SQL Injection и другие.

Объем данных позволяет анализировать различные аспекты сетевого взаимодействия, включая распределение трафика и аномальные паттерны.

**4. Реализованные запросы**

1. **Как часто разные типы атак используют определённые интернет-протоколы?**  
   Цель: Понять распределение протоколов между различными атаками.  
   **Результат:** Выявлено, что большинство атак использует TCP, а UDP встречается реже, но связан с DDoS-атаками.
2. **Типы атак и их связь с размерами пакетов.**  
   Цель: Определить, какие атаки предпочитают определенные размеры пакетов.  
   **Результат:** Обнаружено, что фишинг и SQL Injection чаще всего используют малые размеры пакетов.
3. **Какие типы интернет-трафика связаны с конкретными атаками?**  
   Цель: Анализ распределения типов трафика (FTP, HTTP и др.) по категориям атак.  
   **Результат:** HTTP-трафик доминирует при фишинговых атаках, а FTP часто используется в Man-in-the-Middle атаках.
4. **Идентификация аномальных паттернов трафика.**  
   Цель: Найти IP-адреса, генерирующие необычно большое количество трафика.  
   **Результат:** Обнаружены IP-адреса, с которых осуществляется сканирование сети и DDoS.
5. **Определение самых активных отправителей.**  
   Цель: Подсчитать количество запросов от каждого отправителя для выявления подозрительных.  
   **Результат:** Обнаружены Sender ID, совершившие наибольшее количество запросов.

**5. Методы анализа**

Для анализа данных использовались следующие библиотеки и инструменты:

* **Python (pandas, numpy):** Для обработки данных и расчета статистик.
* **Matplotlib и Seaborn:** Для визуализации полученных результатов.
* **NetworkX:** Для построения графовых связей между IP-адресами.
* **Percentile Analysis:** Для определения порогов аномального поведения.

**6. Выводы**

Проект позволил:

1. Определить ключевые характеристики интернет-трафика, связанные с атаками.
2. Выявить аномальные IP-адреса и связи между отправителями и получателями.
3. Понять предпочтения атакующих по протоколам и размерам пакетов.

Эти результаты могут быть полезны для создания систем раннего обнаружения атак и повышения осведомленности пользователей о киберугрозах.