Тестовое задание для стажера на позицию «Исследователь» ЦНИПР по направлению «Машинное обучение в сфере информационной безопасности»

**Задание №1**

*Вводная информация:*

Поведенческий анализ компьютерных систем и сетей с использованием машинного обучения представляет большой интерес для сферы информационной безопасности, так как позволяет обнаруживать те признаки компьютерных атак, которые невозможно обнаружить сигнатурными и эвристическими методами.

Целью выполнения данного задания является демонстрация базовых знаний и навыков в области сетевых технологий, способов оцифровки данных предметной области, использования нейронных сетей (НС), представления результатов исследований.

Предоставляемый для выполнения задания материал включает: 1 файл pcap для обучения НС, 3 файла pcap для тестирования (поиска аномалий).

Предлагается, но остается на усмотрение исполнителя использовать для решения задачи язык программирования Python и библиотеки numpy, Keras, matplotlib.

*Требуется:*

1. Извлечь из файлов pcap информацию о TCP сессиях.
2. Оцифровать существенные параметры сессий, такие как: IP-адреса, транспортный протокол, серверный порт и т.д. Список используемых параметров сессий остается на усмотрение исполнителя.
3. Построить обучающую и 3 тестовые выборки в виде числовых матриц с фиксированным количеством столбцов. Часть обучающей выборки можно взять для валидации.
4. Создать и обучить НС типа автокодировщик.
5. Обнаружить аномальные TCP сессии в тестовых выборках с использованием мгновенной ошибки реконструкции.

**Присылаемое решение должно включать:**

1. Описание алгоритма решения задачи. Использование блок-схем и прочей визуализации будет преимуществом.
2. Код, содержащий необходимые комментарии.
3. Полученные результаты в графическом и табличном виде.

**Максимальное время на выполнение задания:** 2 недели.