Network Attack Outlier/Anomaly Detection

In this dataset, there are attacks that you will need to detect. This data is network data from physical hosts. Can you find which hosts are anomalous/ outliers?

Questions:

1. Data exploration- what have you learned?
2. Which algorithms group are suitable for this task and why?
3. Please create a report that will explain how you solved the problem?
   a. What is the approach you tried? Why them?
   b. How do you know the algorithm is good?

File descriptions

- Each record has 4 fields (features) that are described in the "Data fields" section. Your task if you accept to take it is to find the attacks in the data.

Data fields

- record ID - The unique identifier for each connection record.
- duration_ This feature denotes the number of seconds (rounded) of the connection. For example, a connection for 0.17s or 0.3s would be indicated with a “0” in this field.
- src_bytes This field represents the number of data bytes transferred from the source to the destination (i.e., the number of out-going bytes from the host).
- dst_bytes This feature represents the number of data bytes transferred from the destination to the source (i.e., the number of bytes received by the host).

What to submit

- CSV with:
  - record ID - The unique identifier for each connection record.
  - is_anomaly?_ This binary field indicates your detection result: 0 denotes the transmission is normal, 1 indicates anomalous.
- Summary Repor + answered questions (no more than two pages).

Good Luck