Network Attack Outlier/Anomaly Detection

Submission: in pairs
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?

It is a must to submit all requested parts failure to do that will result in a 0.

Questions:

1. Data exploration- what have you learned? **[20%]**
2. Which algorithms group is suitable for this task and why?**[5%]**
3. Please create a report that will explain how you solved the problem.[**5%]**
    a. What is the approach you tried? Why them?
    b. How do you know the algorithm is good?
4. What is the accuracy and recall of the algorithm the team developed?
    Show the confusion matrix. Will be ranked based on the results of the entire class [**40**%]
5. User anomaly detection UI based on Docker implementation.

File descriptions

- Each record has four fields (features) that are described in the "Data fields" section.

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 outgoing 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 normal transmission, and one indicates anomalous.
- Jupyter Notebook with:
    - Working code with comments
    - Data exploration

    - Summary of the results.
    - Answers to the questions and how and why the team selected to solve this problem in this way.
- Docker application with UI- [**30**%]
    - will be presented in class. Jupyter notebook will contain a link to a GitHub project. The project can be installed fully from git.
    - The docker will contain a UI application that a user can add an incident and get a prediction if this incident is an anomaly.

Good Luck