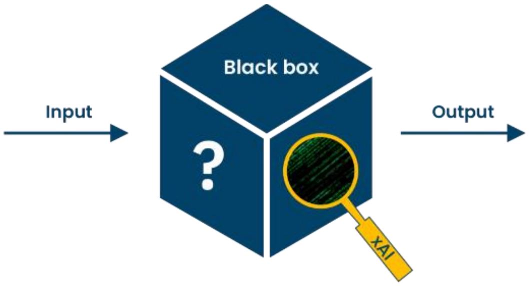
Project 2: Explainable Artificial Intelligence and Cybersecurity



Abstract:

Artificial intelligence models for cybersecurity behave like “black-boxs”. These models do not provide any explanation how a certain prediction was produced which decreases the confidence in their utility.
Explainable artificial intelligence (XAI) was introduced to help users, security experts and organizations in understanding the results of deployed models. This project aims at exploring explainable artificial intelligence and cybersecurity.

Keywords: Explainable AI, Data Driven Security, Cybersecurity.

Tasks:

1. Provide a brief study of XAI tools used in the following paper:
    - Mane, Shraddha & Rao, Dattaraj. (2021). Explaining Network Intrusion Detection System Using Explainable AI Framework.

    Note: You can use AI tools to explore the paper:

    2024 Twelve Best free AI tools for Academic Research || Latest AI tools || AI for researchers.

    https://www.youtube.com/watch?v=qtlUwwtvuEg

    But you must formulate any idea with your own words and thoughts.

2. Implement XAI using LIME and SHAP to explain the predictions of an AI based IDS.

Deliverables:

Report in .pdf format to be uploaded on Moodle. The report includes an introduction, commented code, results, conclusion and references.

Groups: 2 students.

Deadline:

Presentations in the final lab session on 15th February 2024.

Grading:

- 80% Presenting (**in English**) the commented code while showing understanding of the different functionalities and results-You can prepare slides, but this is not compulsory-.
- 10% report (Content and quality).

- 10% Originality (i.e. new tool, new graphs, new explications, another dataset than nsl-kdd…).

References:

Mane, Shraddha & Rao, Dattaraj. (2021). Explaining Network Intrusion Detection System Using Explainable AI Framework.