|  |  |
| --- | --- |
| **Level** | bachelor  master |
| **Project type** | diploma |
| **Bachelor** | Data Science and Artificial Intelligence |

Metadata

|  |  |
| --- | --- |
| **Author** | Luca Lucchina |
| **Title** | Clustering and classification of cyberattacks |
| **Abstract in Italian** | Gli Cyber-attacks sono diventati molto più sofisticati, causando interruzioni dei sistemi in tutti i settori. Questo progetto mirava a classificare grandi insiemi di attacchi non etichettati utilizzando le caratteristiche del payload. I metodi non supervisionati hanno fallito a causa dei confini deboli dei cluster, motivando un approccio semi-supervisionato con etichette esterne, anche se l'incertezza è rimasta. Integrando payload, tempistica, geografia, ASN e porte, abbiamo creato un metodo di pseudo-etichettatura per isolare in modo affidabile gli attacchi e consentire avvisi automatici. |
| **Abstract in English** | Cyberattacks have grown highly sophisticated, disrupting systems across sectors. This project aimed to classify large sets of unlabelled attacks using payload features. Unsupervised methods failed due to weak cluster boundaries, motivating a semi-supervised approach with external labels, though uncertainty remained. By integrating payload, timing, geography, ASNs, and ports, we built a pseudo-labelling method to reliably isolate attacks and enable automated alerts. |
| **Supervisors** | Angelo Consoli |
| **Submission date** | 22/08/2025 |
| **Company involved** | Yes No |
| **Confidentiality requirements and other comments** | Dataset IP and coordinates will be kept hidden as the honeypot system is still in use and could compromise the series of servers |