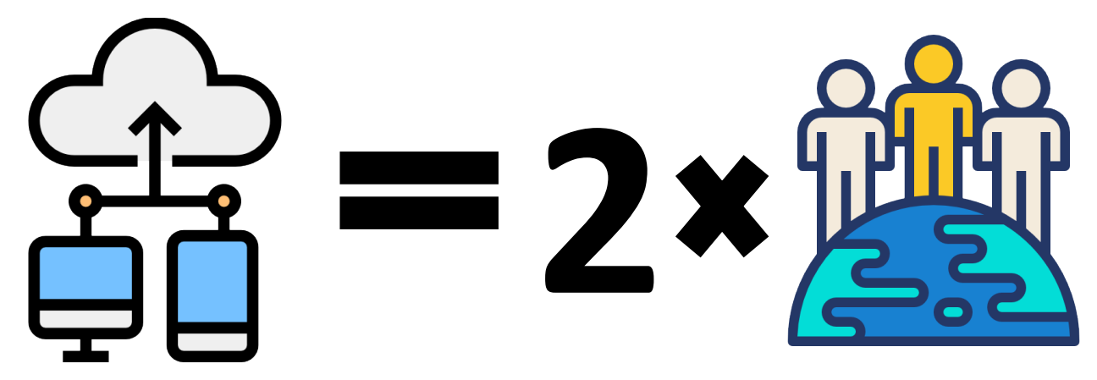
Clustering Connections and Hosts using a Hybrid Sequential Approach

Azqa Nadeem[1] **Mark Patrick Roeling[1,2,Φ]** Sicco Verwer[1]
[1] Cyber Security Group, Delft University of Technology, The Netherlands
[2] University of Oxford., United Kingdom
[Φ]m.p.roeling@tudelft.nl

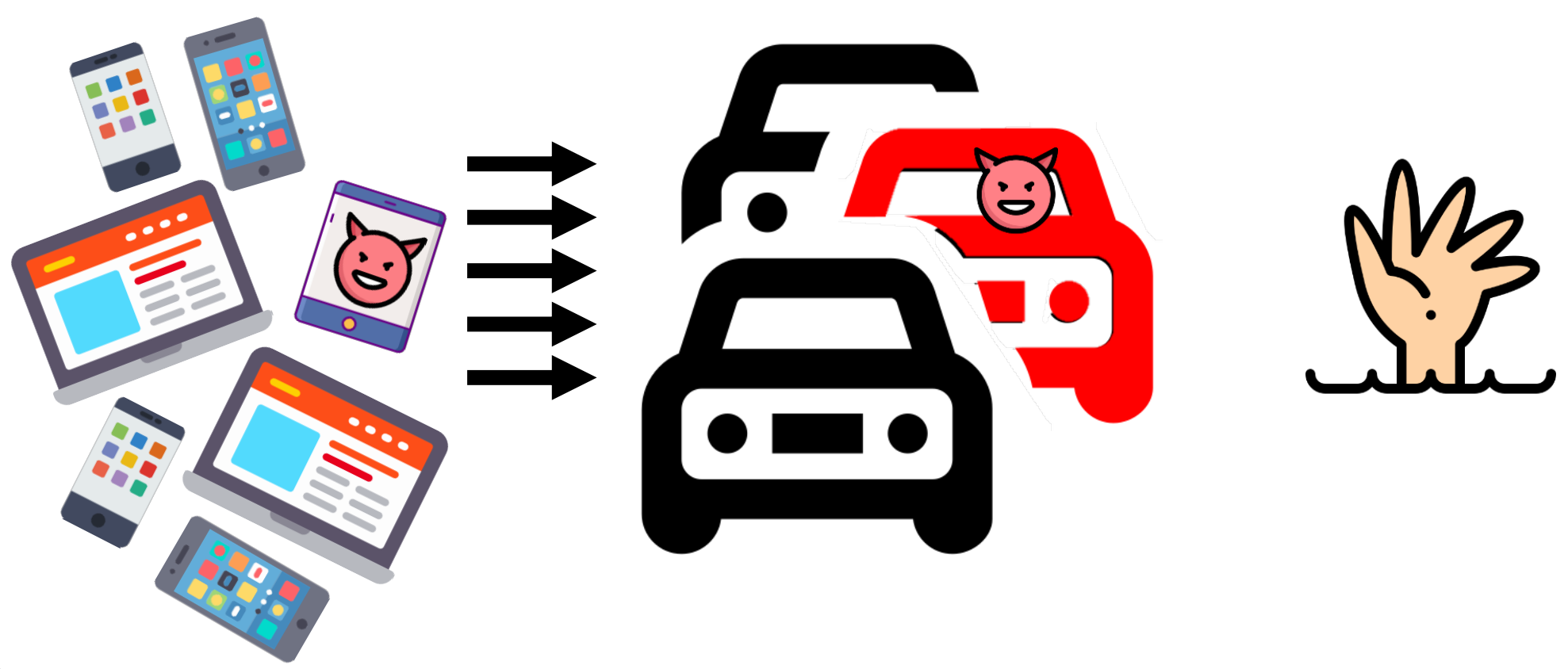
Problem: Host classification is difficult due to high traffic volumes

Too much traffic generation



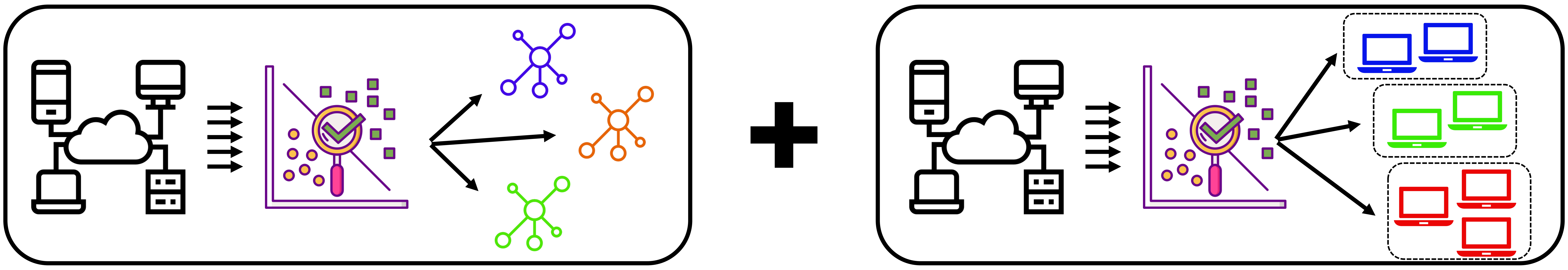
Prediction: 75bn devices to be connected by 2025*

Difficult to detect infected hosts

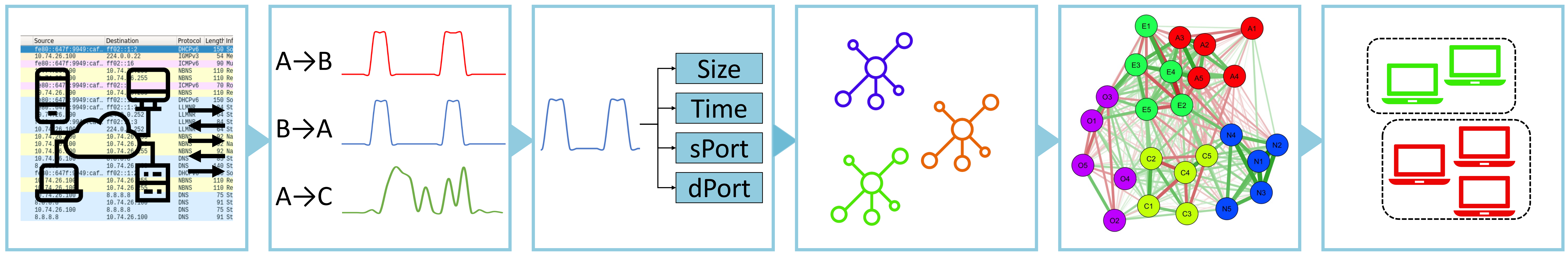


Multi-layered efficient host classification

Key idea: *Host classification via connection classification*



Proposed Framework

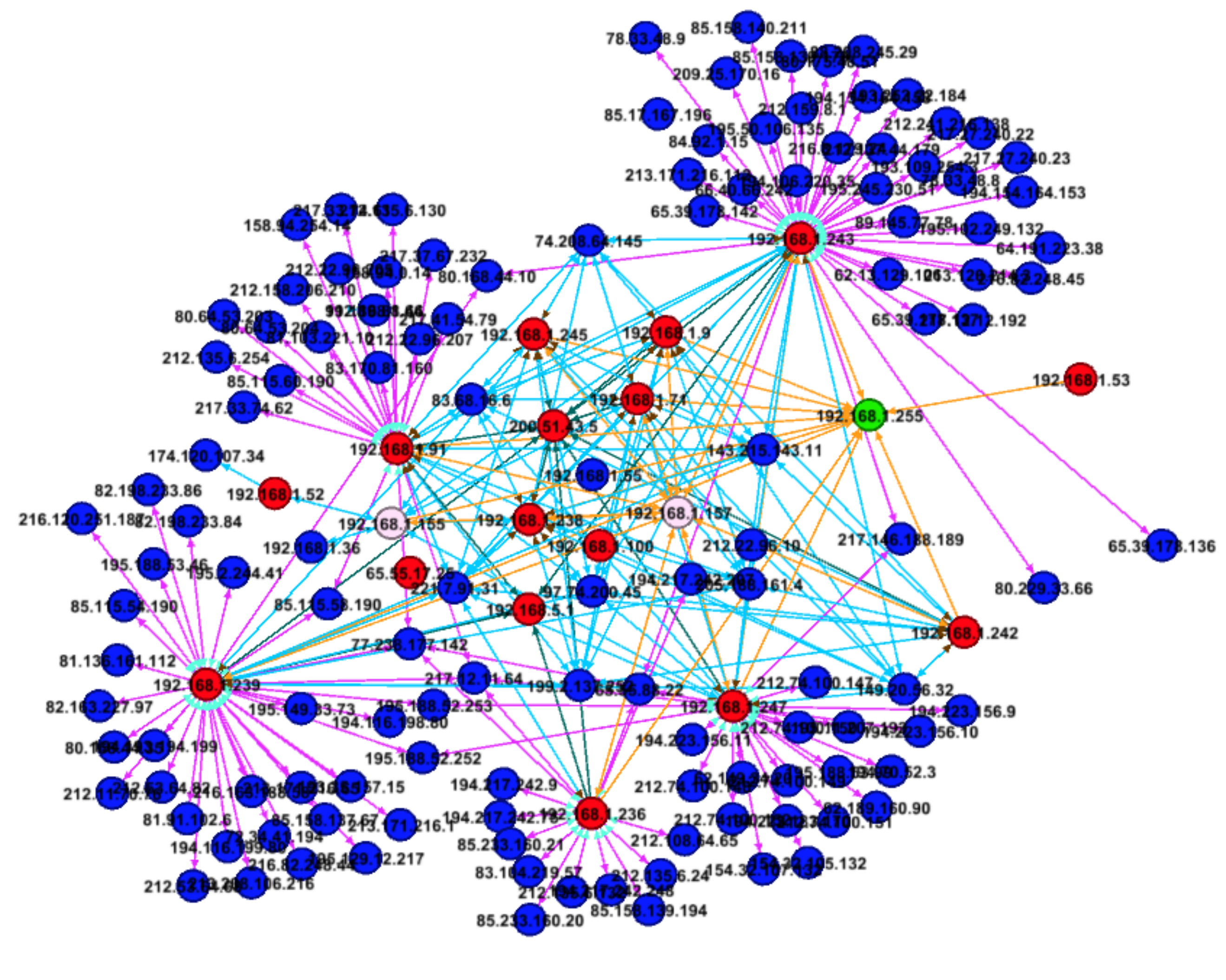


Netflows → **Connection generation** → **Feature extraction** → **HDBScan clustering** → **SMB clustering** → **Host identification**

Initial Results

- Misclassification of the nodes was lower in the combined method compared to the separate methods.
- Conjointly, 88.5% of nodes were labelled correctly.

This study provides a rationale for using clustered connections as input for host-classification in this context, by sequential modeling of two powerful clustering methods, without the need for labelled data.



Icons made by Flat icons, Eucalyp, Smashicons, Freepik, Ultimatearm, Becris, Skyclick from www.flaticon.com | *Statistics taken from [https://safeatlast.co/blog/iot-statistics/]