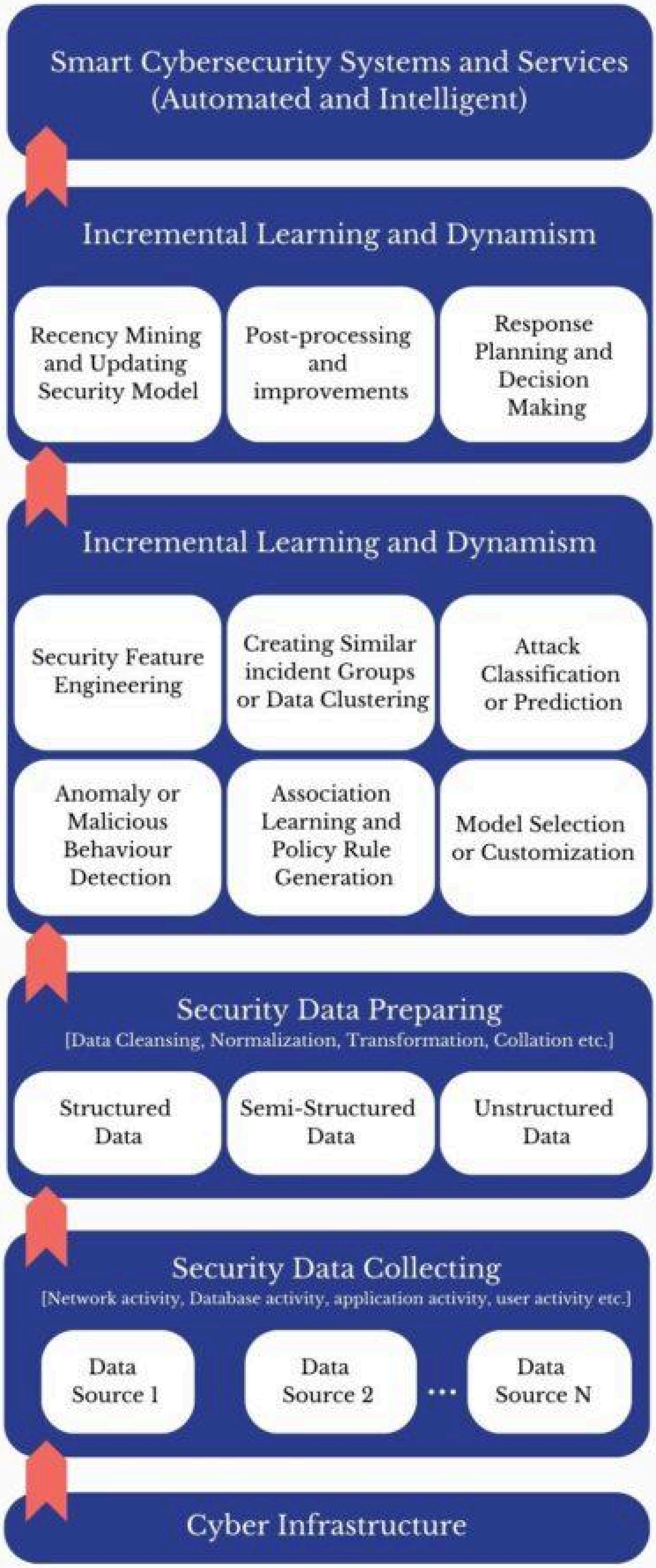
An overview of cyber security data science from a perspective of machine learning

Machine learning tasks in cyber security

Machine learning (ML) is sometimes regarded as a subset of “Artificial Intelligence,” and it is strongly related to data science, data mining, and computational statistics. It focuses on teaching computers to recognize patterns from data. **Machine learning models,** which could be crucial in the field of cyber security, often consist of a collection of rules, techniques, or intricate “transfer functions” that can be used to uncover interesting data patterns or to recognize or anticipate behavior. Here, we’ll go through various approaches for handling machine learning problems and how they relate to cyber security issues (Assistance, 2022).

PHD ASSISTANCE



Smart Cybersecurity Systems and Services (Automated and Intelligent)

Incremental Learning and Dynamism

- Recency Mining and Updating Security Model
- Post-processing and improvements
- Response Planning and Decision Making

Incremental Learning and Dynamism

- Security Feature Engineering
- Creating Similar incident Groups or Data Clustering
- Attack Classification or Prediction
- Anomaly or Malicious Behaviour Detection
- Association Learning and Policy Rule Generation
- Model Selection or Customization

Security Data Preparing

[Data Cleansing, Normalization, Transformation, Collation etc.]

- Structured Data
- Semi-Structured Data
- Unstructured Data

Security Data Collecting

[Network activity, Database activity, application activity, user activity etc.]

- Data Source 1
- Data Source 2
- ...
- Data Source N

Cyber Infrastructure

Neural networks and deep learning

Deep learning is a type of **machine learning**, a subset of **artificial intelligence** that takes cues from biological neural networks seen in the human brain. The most widely used neural network algorithm is back propagation, and artificial neural networks (ANN) are extensively employed in deep learning (Aversano et al., 2021). It executes learning on an input layer, one or more hidden layers, and an output layer of a multi-layer feed-forward neural network. Deep learning performs better as the volume of security data increases, which is the primary distinction between it and traditional machine learning. Typically, deep learning algorithms work best with vast amounts of data, whereas machine learning techniques work well with smaller datasets.

Read more