Banche e Sicurezza
25 e 26 Maggio 2021

Intelligenza artificiale un cambio di paradigma nella difesa digitale

Marco Ramilli, CEO Yoroi – Tinexta Cyber

Definitions

Artificial Intelligence. A broad concept. A **Science** of **making things smart** or, in other words, human tasks performed by machines (e.g., Visual Recognition, NLP, etc.)

Machine Learning. An **Approach** (just one of many approaches) to AI that uses a system that is capable of **learning from experience**.

Deep Learning. A set of **Techniques** for implementing machine learning that recognize patterns of patterns, like **image recognition**

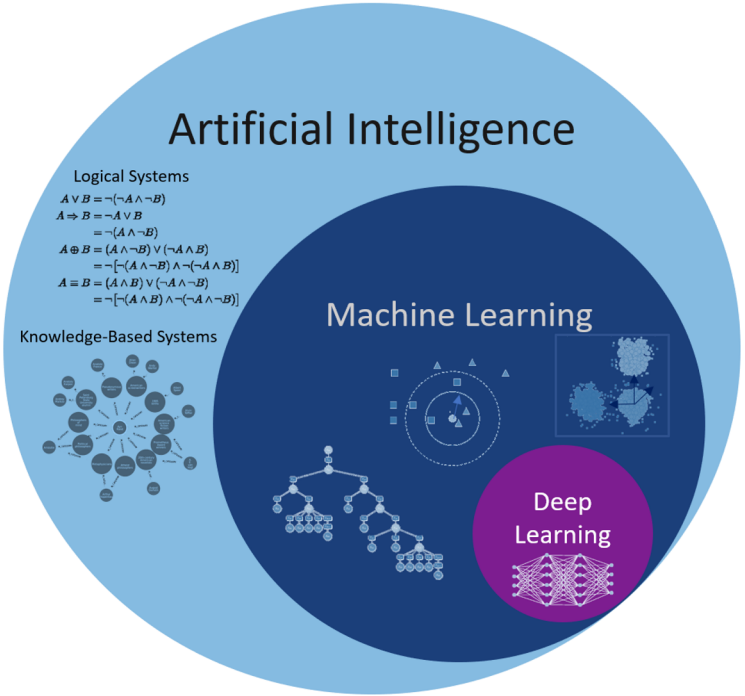


Image from: https://data-science-blog.com

Machine Learning

- **Supervised Learning.** Machine is trained on past data and its output are used to learn the hidden patterns. Usually used for predictive analyses
- **Unsupervised Learning.** Output are not known priori. Machine is not trained on any data. Mainly used for clustering
- **Reinforcement Learning.** Machine Agent explores its environment for decision making. Learning is based on reward and punishments policy. Used for making decisions in problems to maximise rewards

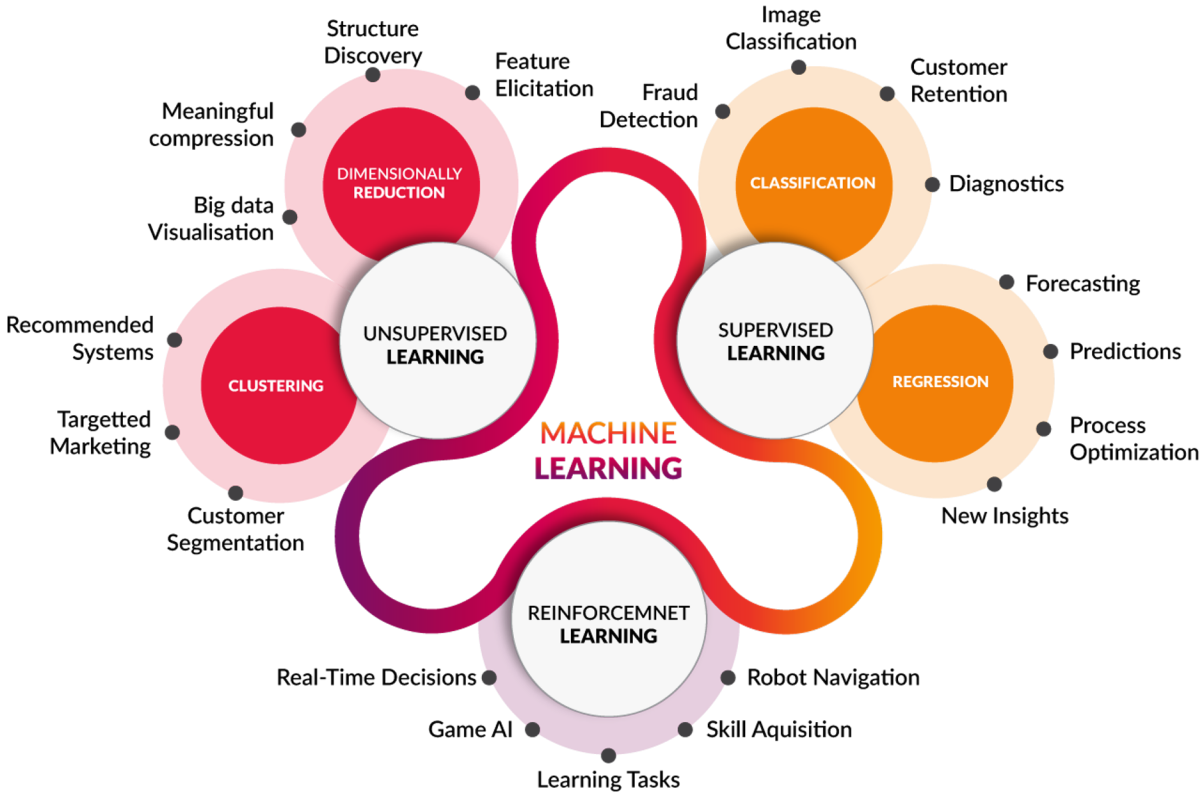


Image From: http://www.cognub.com/index.php/cognitive-platform/

ML on Cybersecurity

- **Regression (or prediction).** The knowledge about the existing data is utilized to have an idea of the new data.

  Linear regression, Polynomial regression, Ridge regression, Decision trees, SVR (Support Vector Regression), Random forest

- **Classification.** The supervised learning approach is usually used for classification where examples of certain groups are known. All classes should be defined in the beginning. Spam and Ham are the typical cybersecurity scenario.

  LogisticRegression (LR), K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), KernelSVM, NaiveBayes, DecisionTreeClassification, Random Forest Classification

- **Clustering.** Clustering is similar to classification with the only but major difference. The information about the classes of the data is unknown. There is no idea whether this data can be classified. This is unsupervised learning.

  K-nearest neighbours (KNN), K-means, Mixturemodel(LDA), DBSCn, Bayesian, Gaussian Mixture Model,...

- **Association Rule Learning**. In cybersecurity, this principle can be used primarily for incident response. If a company faces a wave of incidents and offers various types of responses, a system learns a type of response for a particular incident

  Apriori, Euclat, FP-Growth

ML on Cybersecurity

- **Dimensionality Reduction.** Used to cut or to handle unnecessary features, used to be the core algorithm for visual recognition.

  Principal Component Analysis (PCA), Singular-value decomposition (SVD), T-distributed Stochastic Neighbor Embedding (T-SNE), Linear Discriminant Analysis (LDA), Latent Semantic Analysis (LSA), Factor Analysis (FA), Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF)

- **Generative Models.** Used to create data and not to test data. Used in Cybersecurity to generate input for specific software in order to test injection paths

  Markov Chains, Genetic algorithms

ML applied on Cybersecurity

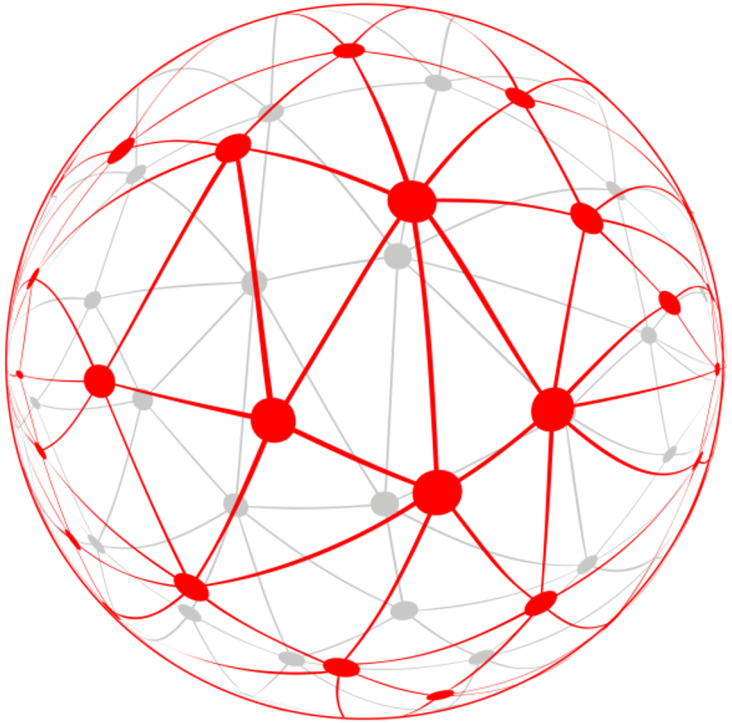


Image From: http://podatki.mieroslawska.pl

- Network Protection
- Endpoint Protection
- Application Protection
- Human Behavior Protection

  - **regression** to predict the parameters and compare them with the normal ones
  - **classification** to identify different classes of such as scanning and spoofing
  - **clustering** for forensic analysis

Machine Learning Attacks

- **Direct Gradient Based Attack**

The attacker needs to know model structure and model weights.

The attacker needs to know model structure and model weights.

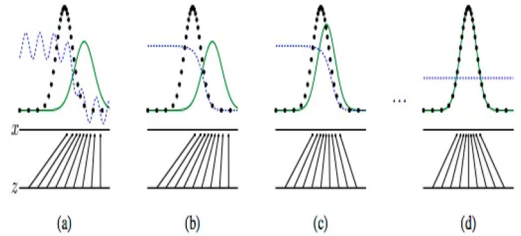


Image From: “Generative Adversial Nets” paper

- **Score Model Attack**

The attack set is based on the score systems.

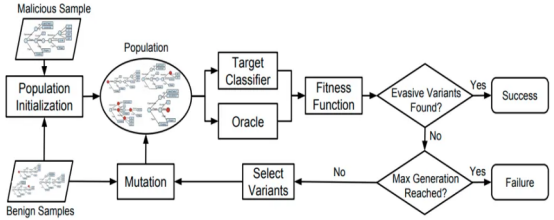
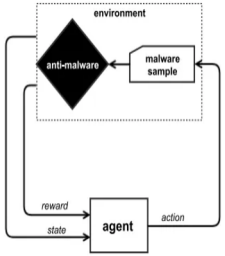


Fig. 1. Generic classifier evasion method.

Image From: “Automatically evading classifiers” paper

- **Binary Black-Box Attack**

The attacker has no idea about the Machine Learning Model and the applied Weights, he has also no idea about the scoring system but he have unlimited access to probe the Machine Learning Model.



The Complexity in Digital Era

The Human Factor

Thank you
Q&A

La presenza Yoroi nel mondo dei Social Media e nella Blogosfera

Facebook LinkedIN Twitter Youtube Blog Yoroi Blog M. Ramilli