CM2604 Machine Learning

Introduction to Machine Learning

Week 01 | Prasan Yapa

ROBERT GORDON UNIVERSITY ABERDEEN
TEF Gold
INFORMATICS INSTITUTE OF TECHNOLOGY

Module Requirements

- Prerequisites for the Module - CM1601 or equivalent, CM1606 or equivalent
- Corequisites for the module - None
- Precluded Modules - None

Aims of Module

To provide a theoretical underpinning of a range of established machine learning (ML) algorithms with focus on implications of real-world deployment.

Learning Outcomes

1. To create a dataset for ML using data and feature engineering methods applied to a real-world data collection.
2. To critically analyze the theory including statistical and mathematical underpinning of a range of ML algorithms.
3. To use ML evaluation methodologies to compare and contrast supervised and unsupervised ML algorithms using an established machine learning framework.
4. To analyze the ethical, social, professional and legal issues associated with collecting /creating datasets and use of machine learning models in the real-world.

Module Content

- Data cleansing, missing values handling , stemming, lemming, encoding of textual data, recognition of independent / dependent variables, over fitting, under-fitting, dimensionality reduction.
- Supervised Learning Techniques: Regression techniques, Bayer's theorem, Naïve Bayer's, SVM, Decision Trees and Random Forest.
- Un-supervised Learning Techniques: Clustering, K-Means clustering, Association Mining, Apriori.
- Ensemble Techniques: Ada-Boost, Bagging, Stacking.
- Evaluation and Testing mechanisms: Precision, Recall, F-Measure, Confusion Matrices, ROC, AUC.
- Data Protection Act, BCS Code of conduct, Ethical Principles.

Module Delivery

- Lectures - 2 hours/Week
- Labs - 2 hours/Week
- Tutorial Feedback Sessions - 2 hours/Week

Assessment Plan

- Examination - 60%
- Coursework - 40%

Recommended Material

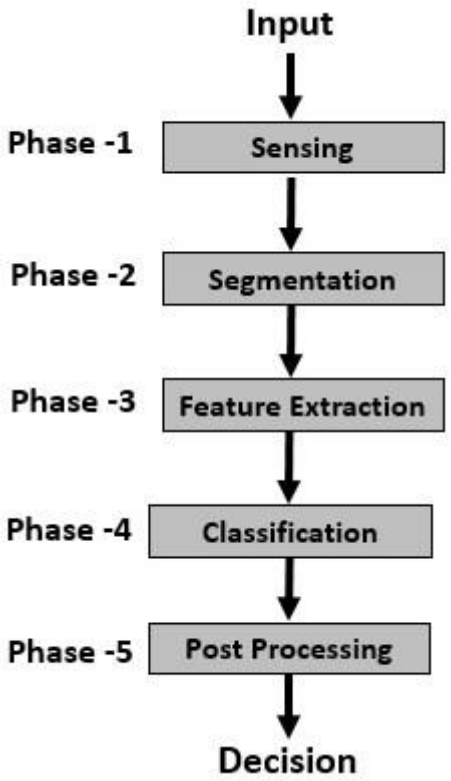
- Geron, A. 2020. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. *O'Reilly*
- Han, J. and Kamber, M. 2006. Data Mining: Concepts and Techniques. 2nd ed. *Morgan Kaufmann*
- Bishop, C. 2007. Pattern Recognition and Machine Learning. *Springer Verlag*
- Provost, F. and Fawcett, T. 2013. Data Science for Business. *O'Reilly Media*
- Tan, P., Steinbach, M. and and Kumbar, V. 2005. Introduction to Data Mining. *Addison-Wesley*

What is ML

Machine Learning

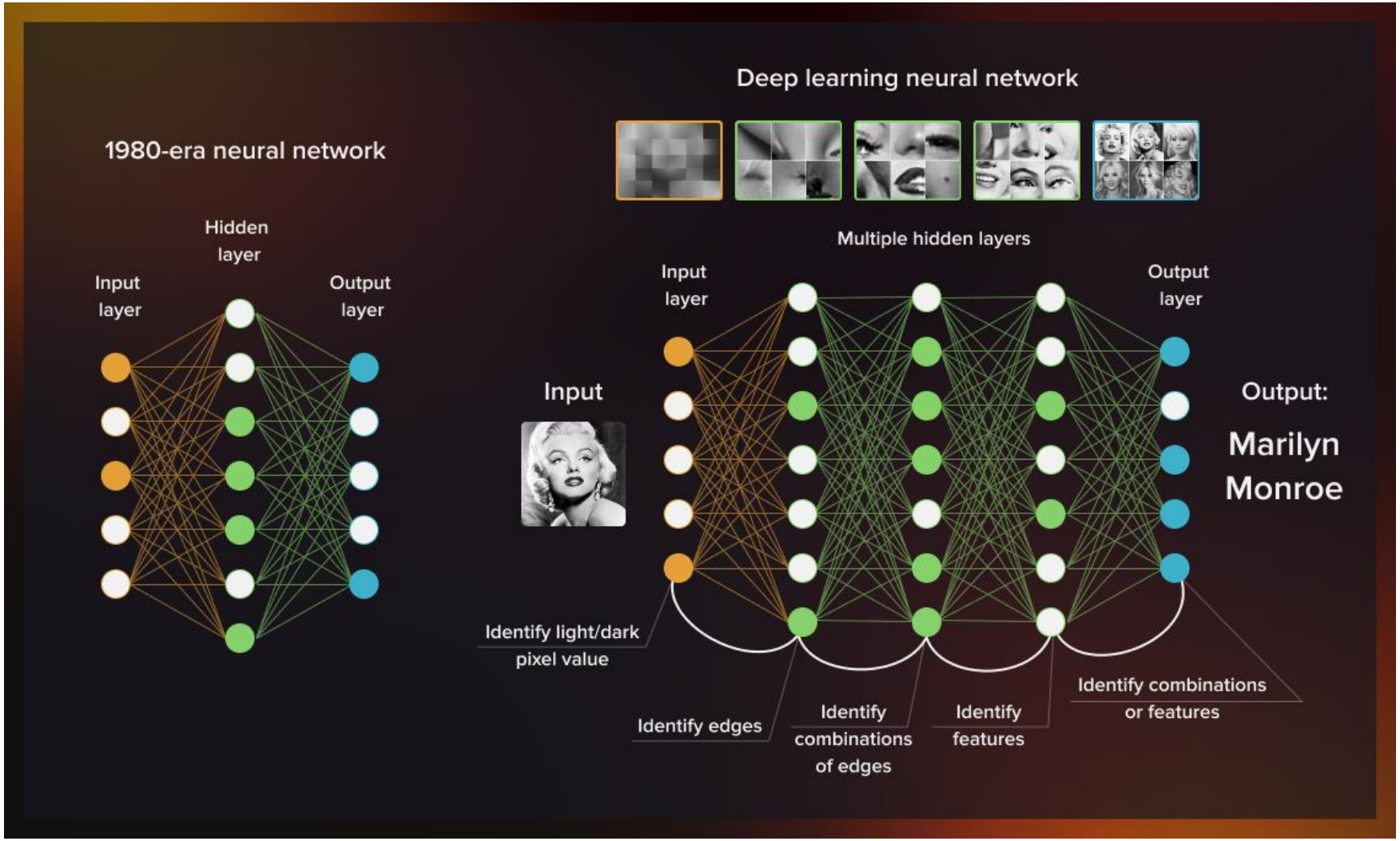
- An evolving branch of computational algorithms that are designed to emulate human intelligence.
- Techniques based on machine learning have been applied successfully in diverse fields.
- ML is widely used in software to enable an improved experience with the user.
- The main advantage is that, once an algorithm learns what to do with data, it can do its work automatically.

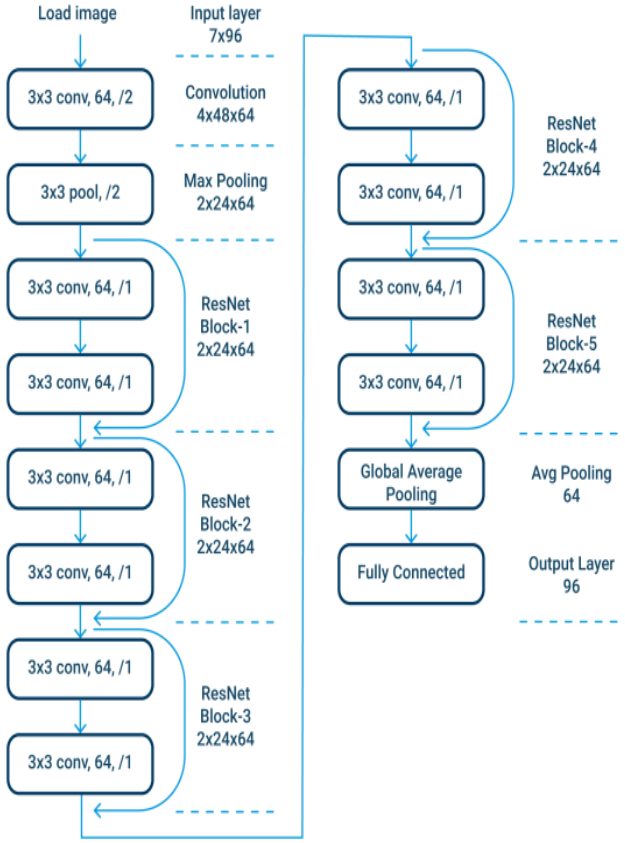
ML for Pattern Recognition



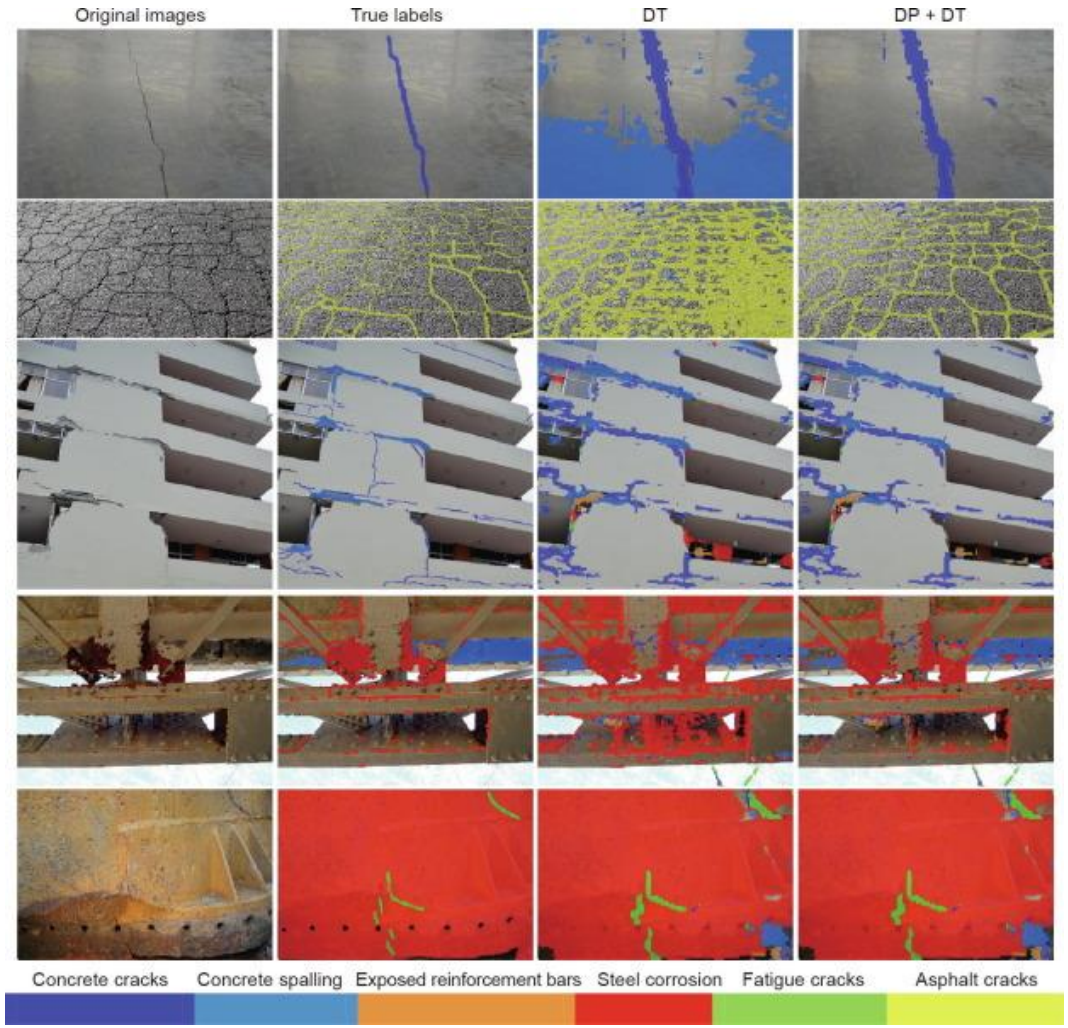
B. N. Narayanan, O. Djaneye-Boundjou and T. M. Kebede, Performance analysis of machine learning and pattern recognition algorithms for Malware classification, 2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS).

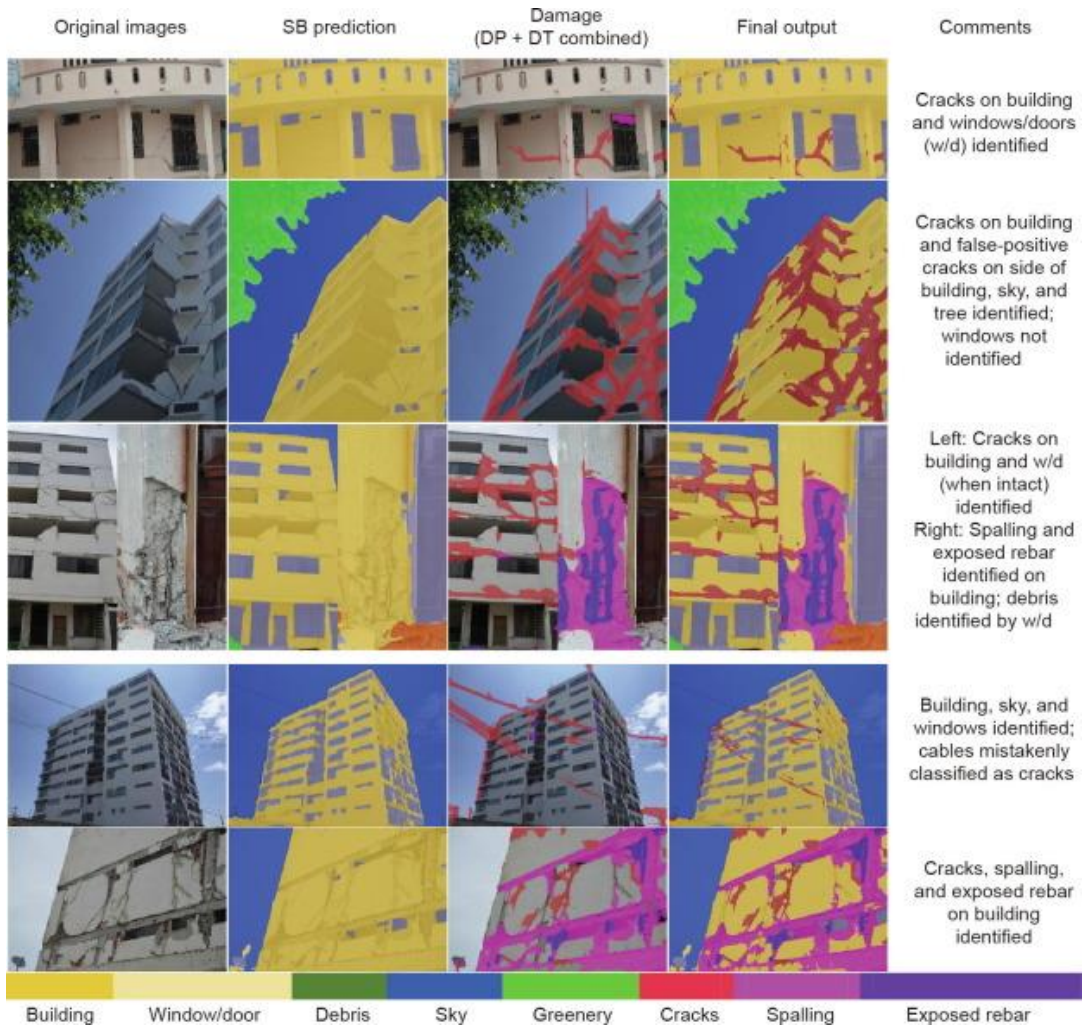
ML for Computer Vision





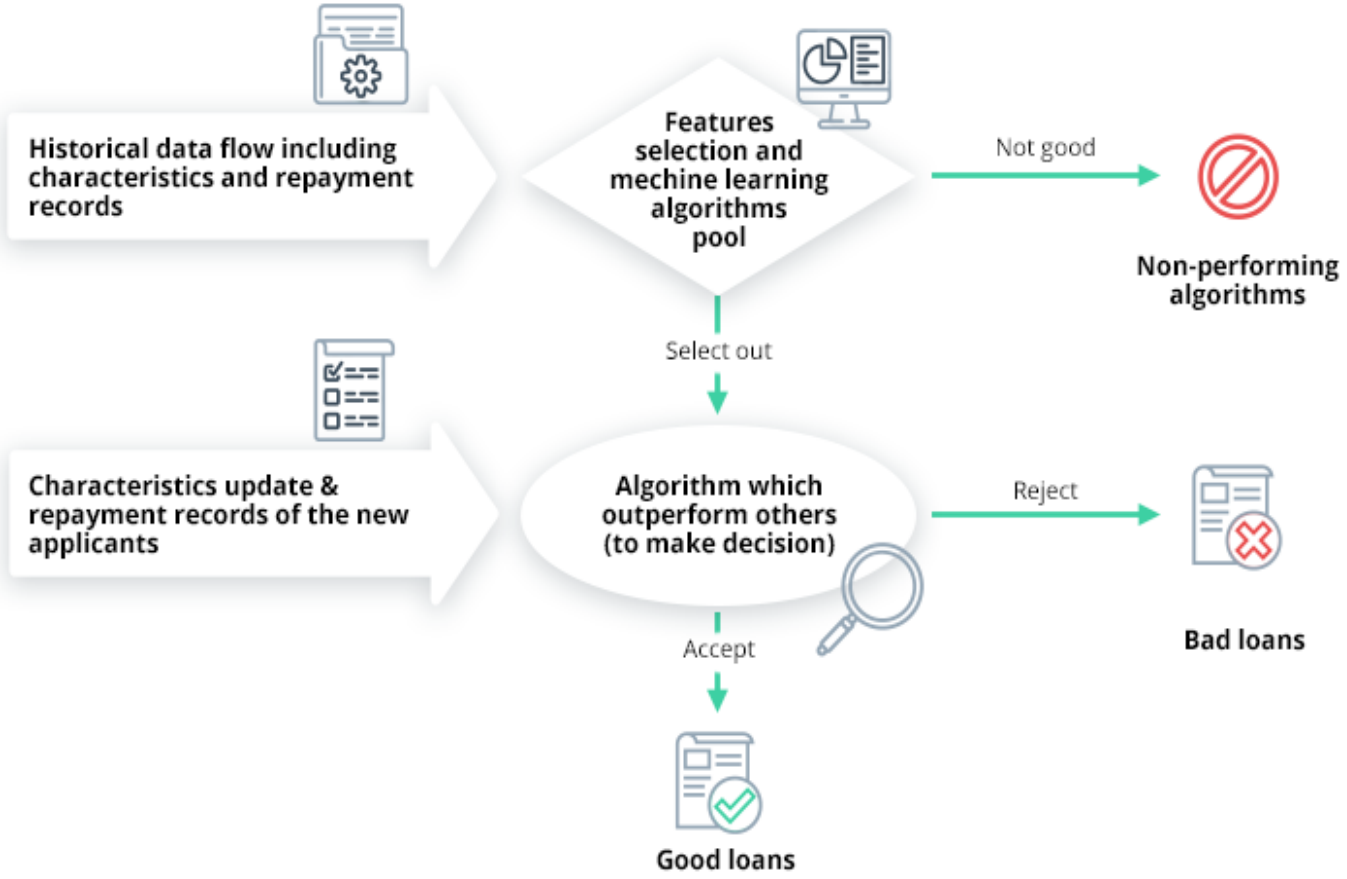
ML for Engineering

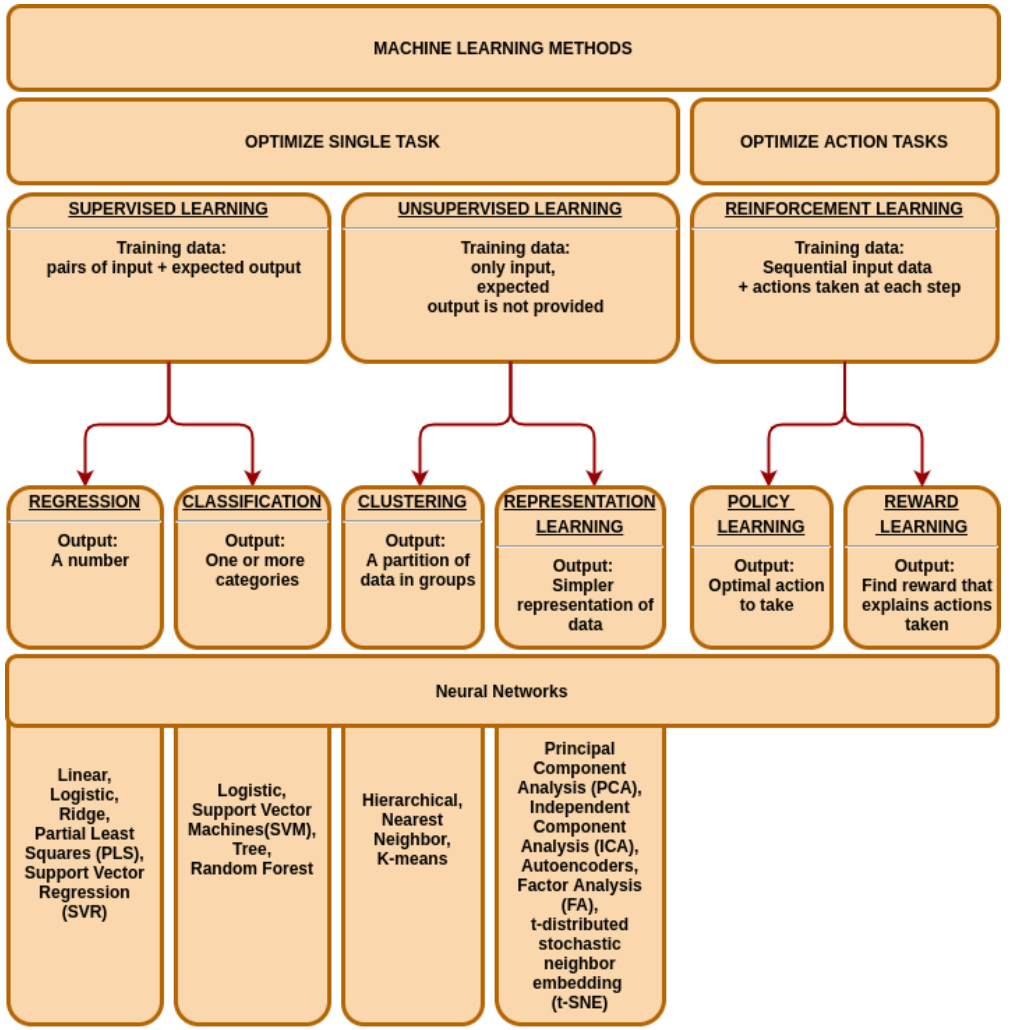




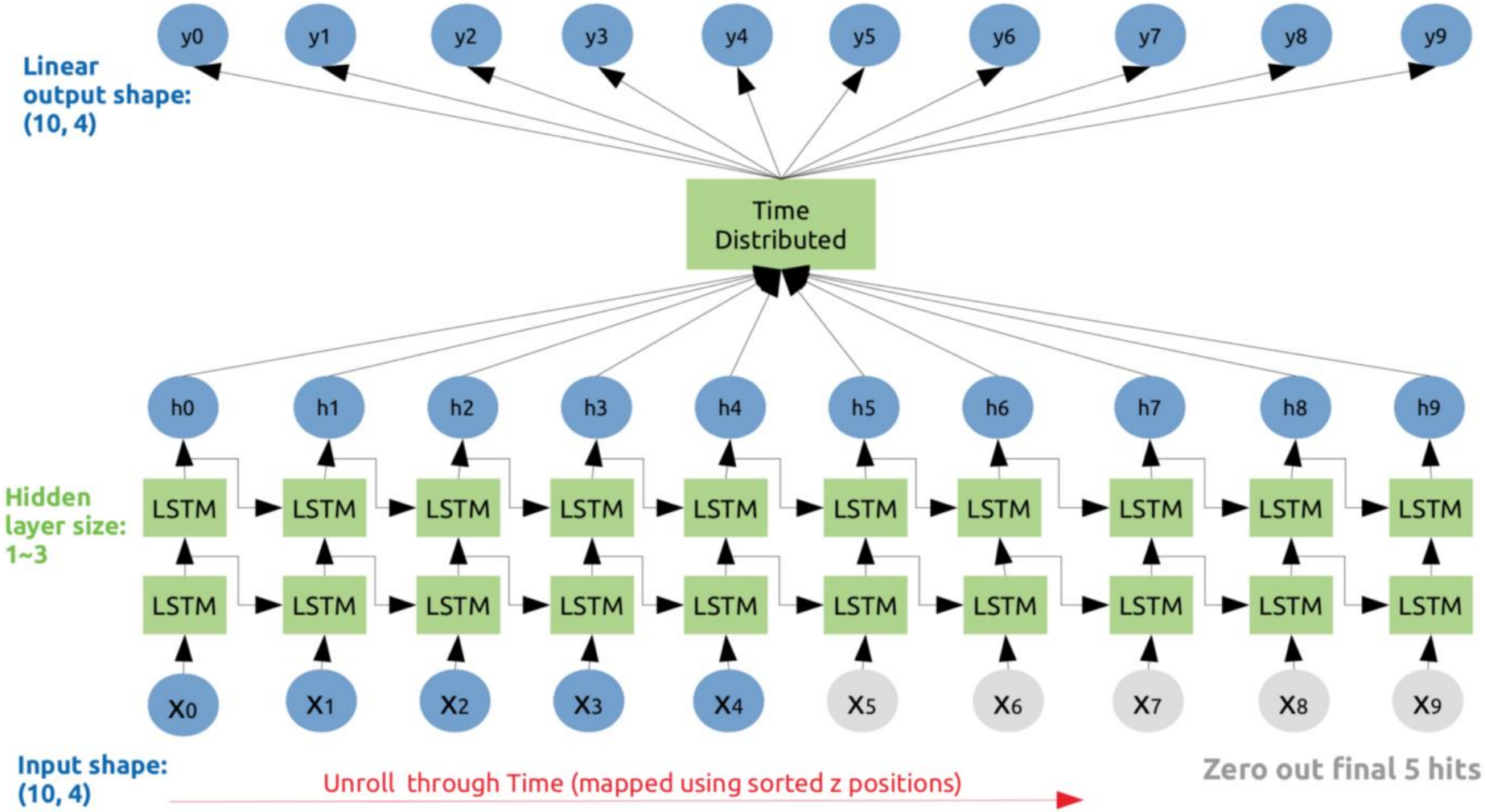
Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. 2022. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems

ML for Finance

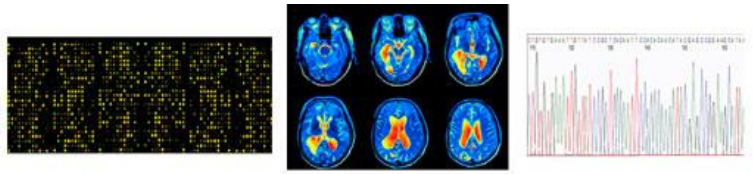




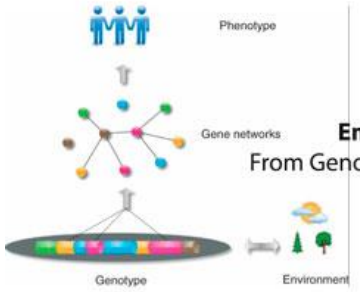
ML for Entertainment



ML for Medicine

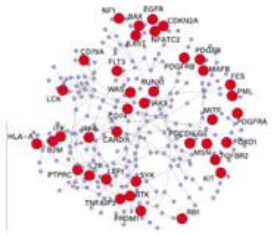


Data
Utilization and integration

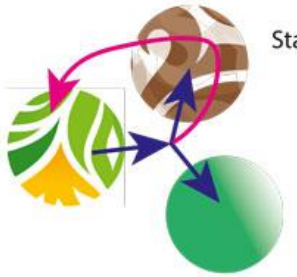


Emergence
From Genotype to Genotype

Networks
Inference and analysis

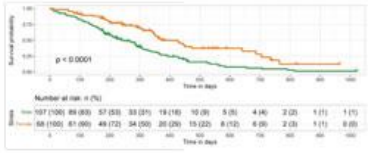


Artificial Intelligence for Molecular Medicine

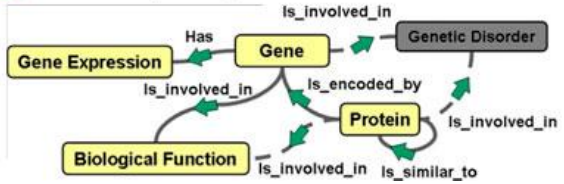


Robustness
Statistical thinking

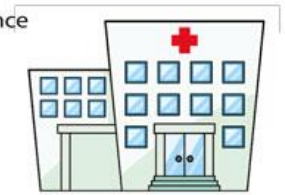
Predictions
Deep learning models

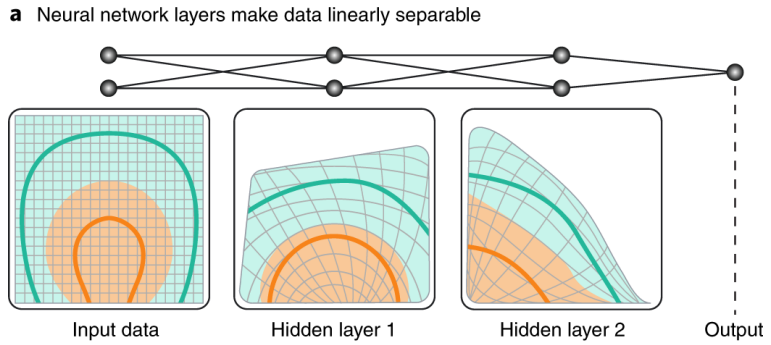


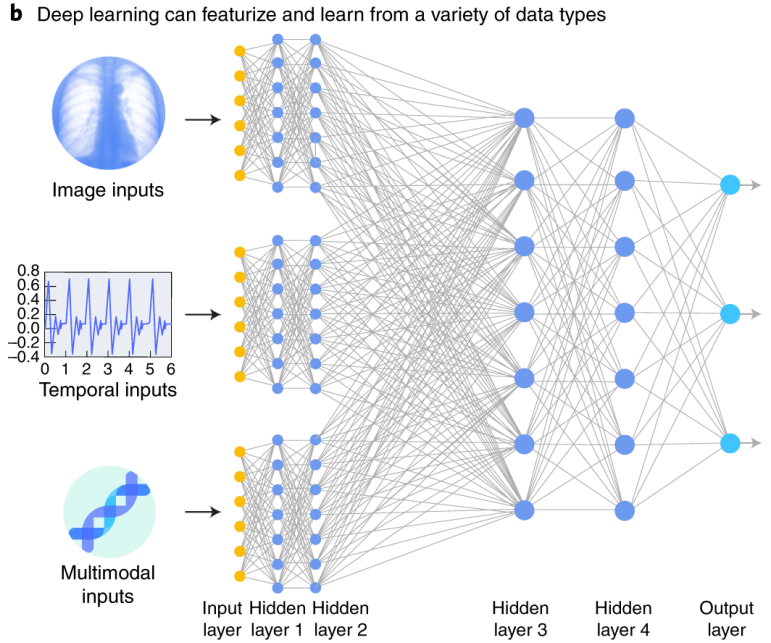
Understanding
Explaining molecular semantics



Decisions
Clinical relevance



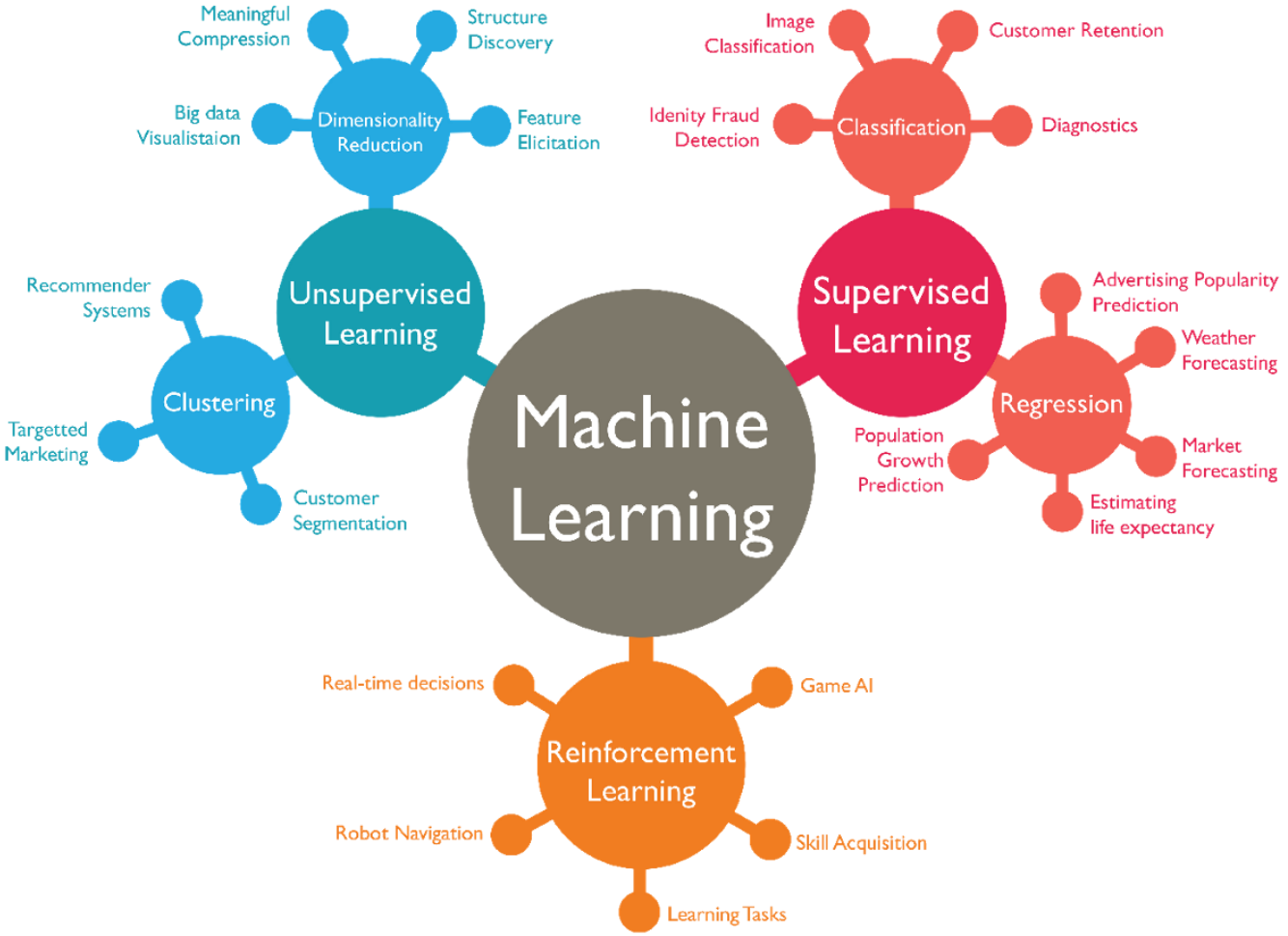




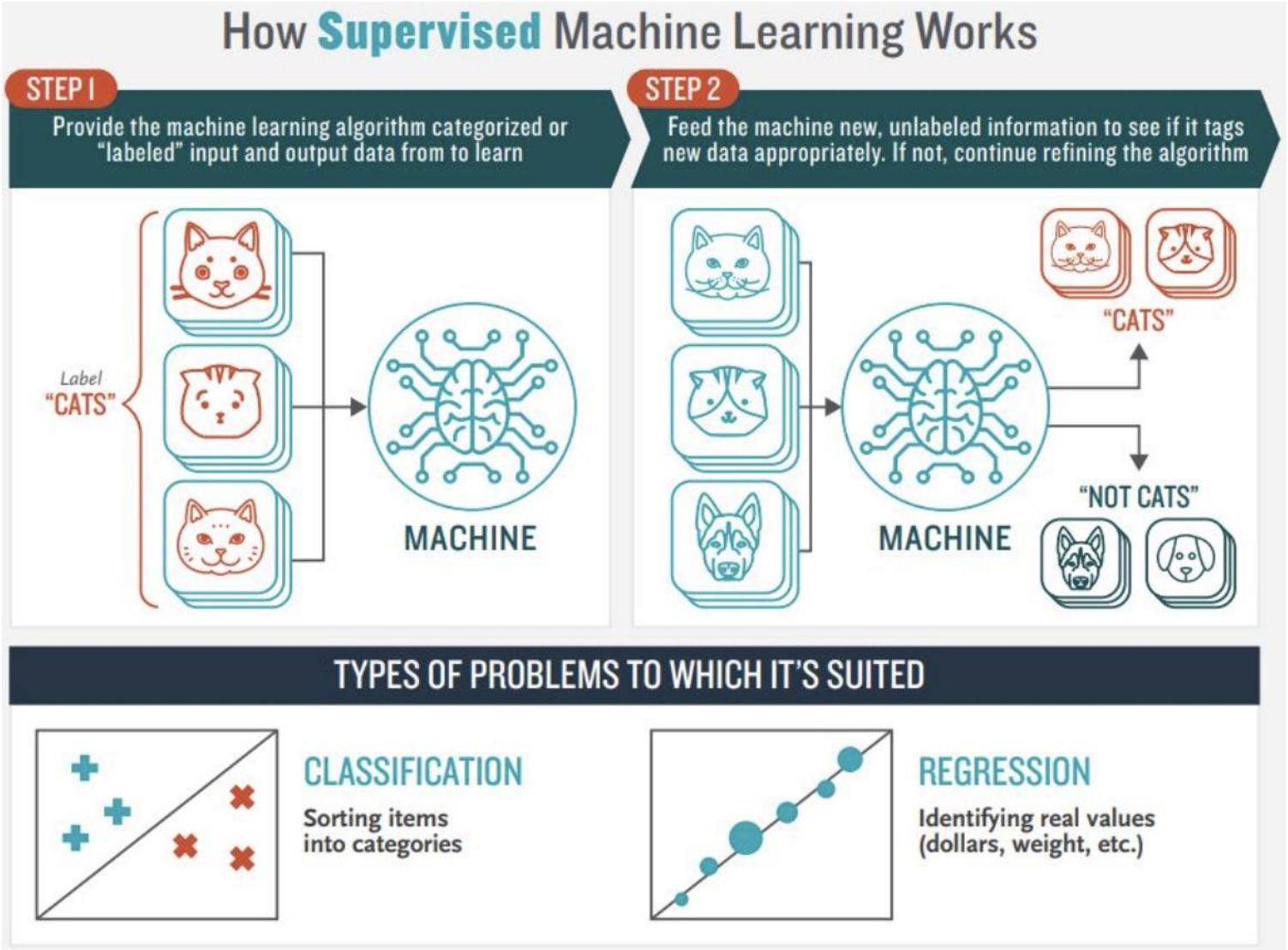
Types of ML

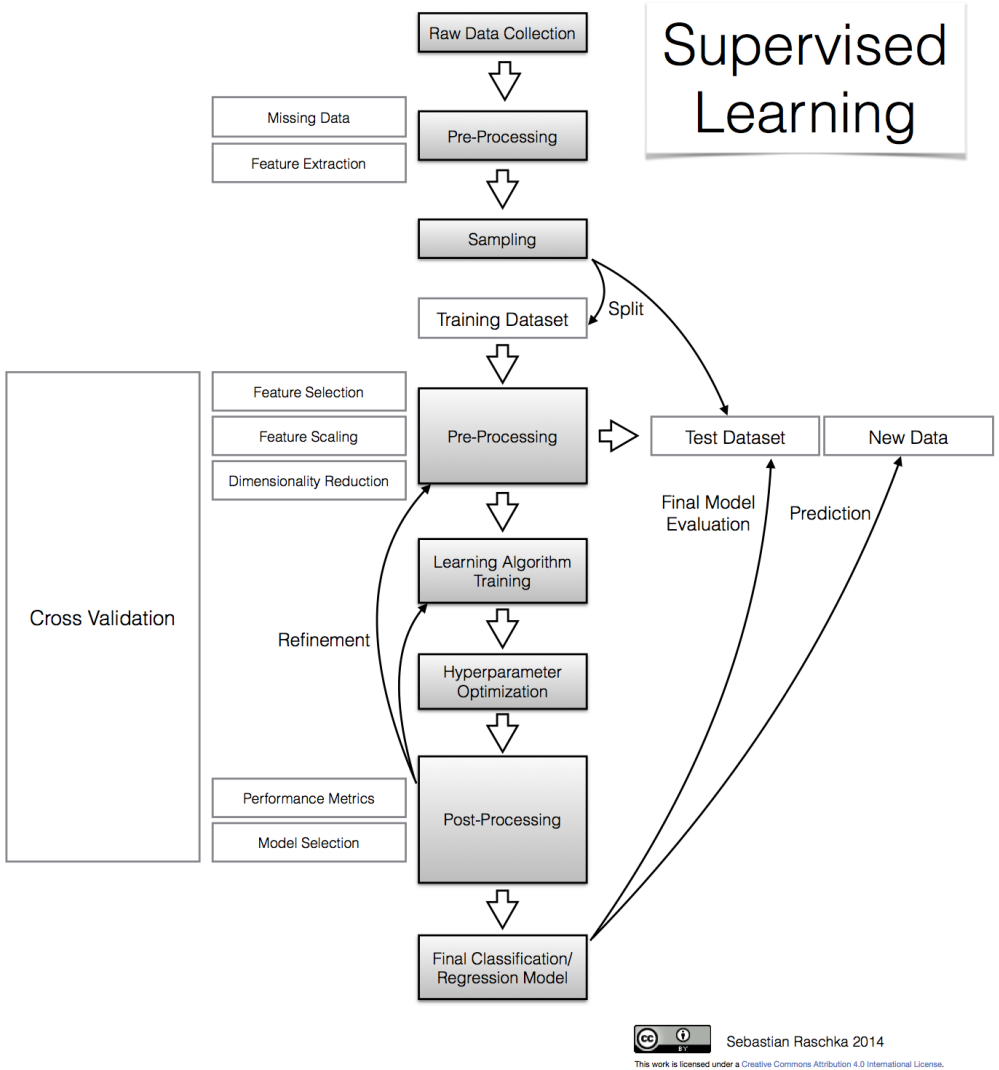
Types of Machine Learning

- Supervised Learning
- Un-supervised Learning
- Semi-supervised Learning
- Reinforcement Learning
- Ensemble Learning



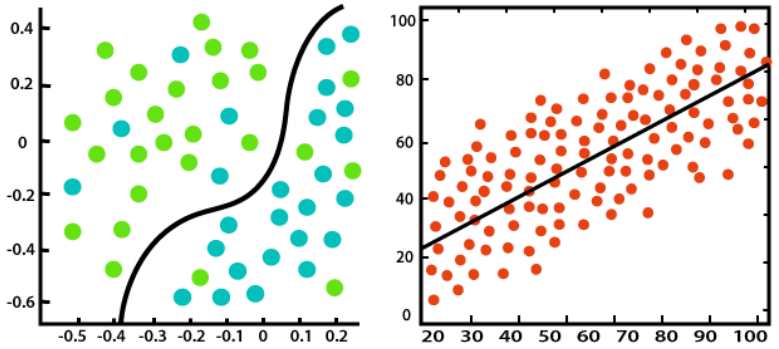
Supervised Machine Learning





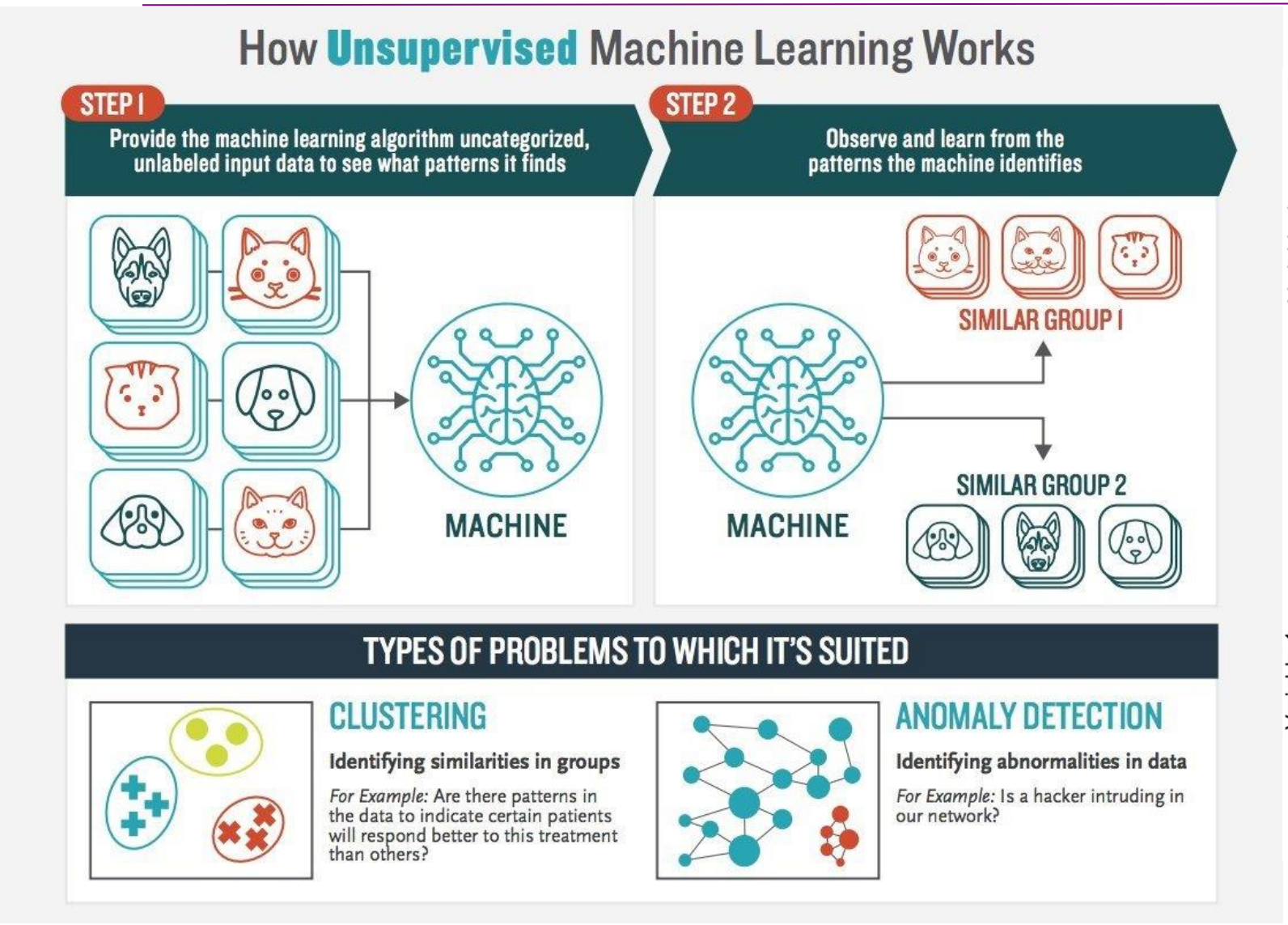
Supervised Machine Learning

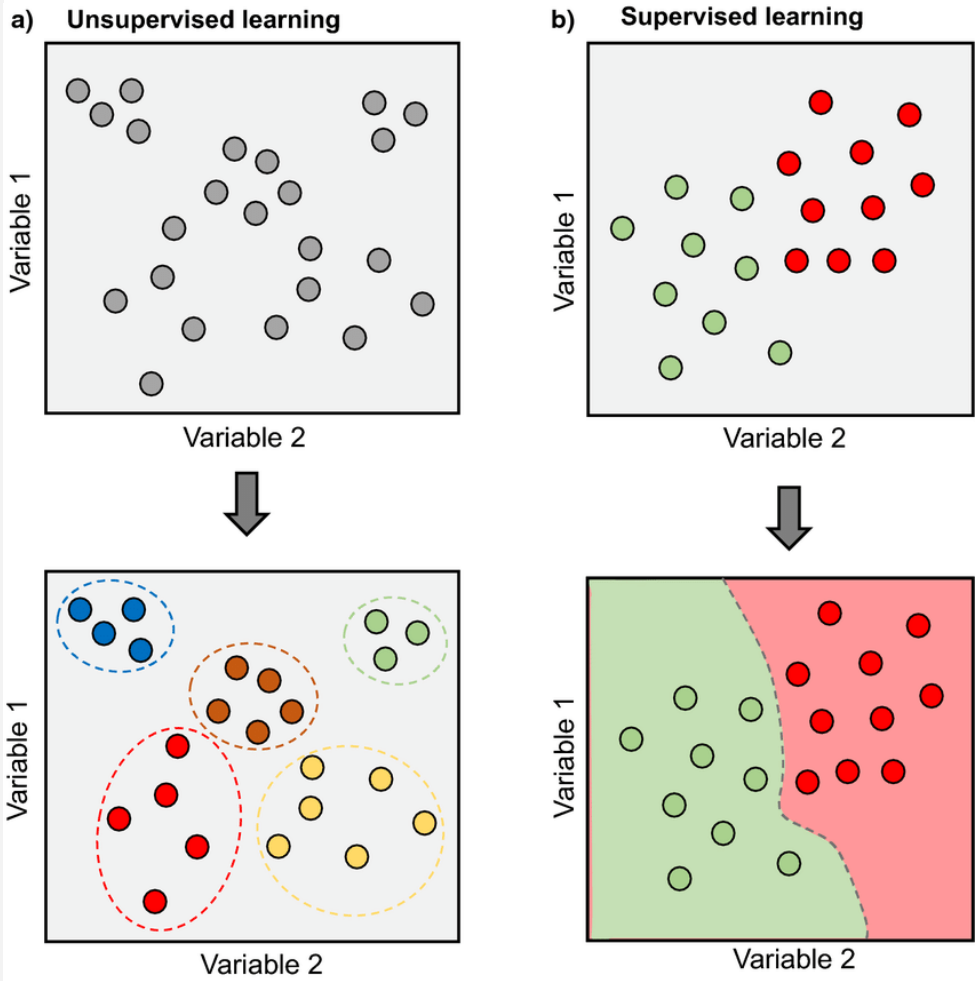
- Classification
  - Grouping the output inside a class (binary vs multiclass classification).
  - K-Nearest Neighbor, Random Forest, SVM, Decision Trees.
- Regression
  - Predicting a single output value using training data.
  - Outputs always have a probabilistic interpretation.
  - Linear Regression and Logistical Regression.



Classification Regression

Un-supervised Machine Learning

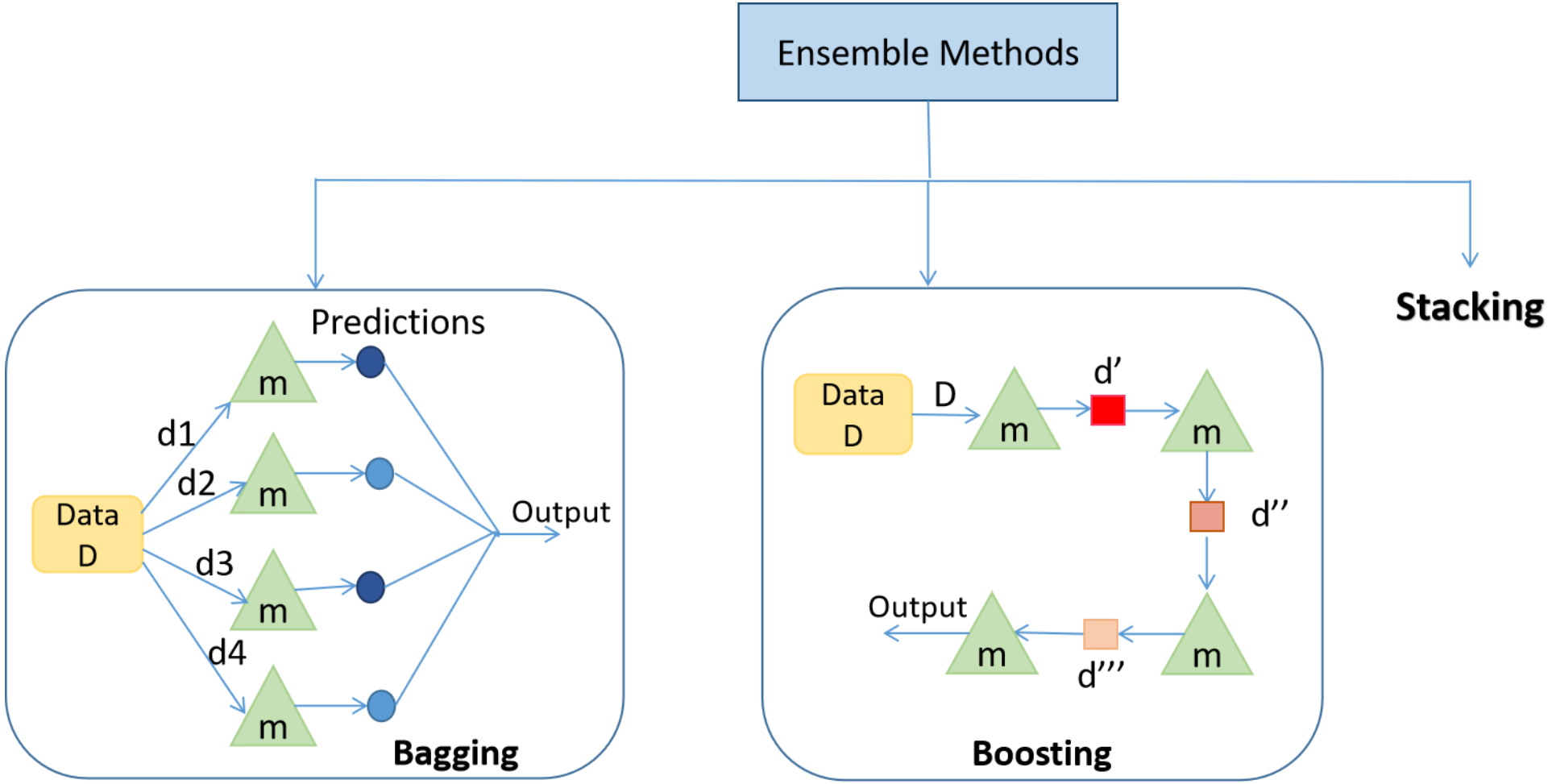




Un-supervised Machine Learning

- Clustering
  - Grouping unlabeled data based on their similarities or differences.
  - Can be categorized such as specifically exclusive, overlapping, hierarchical, and probabilistic.
  - K-means clustering and Gaussian Mixture Models.
- Association Mining
  - Finding relationships between variables in a given dataset using a rule-based method.
  - Used for market basket analysis, products categorization etc.
  - Apriori algorithms.

Ensemble Learning



Ensemble Learning

- Seeking better predictive performance by combining the predictions from multiple models.
- The three main classes of ensemble learning methods are bagging, stacking, and boosting.
- Bagging learns independently from each other in parallel and combines them.
- Stacking learns in parallel and combines them by training a meta-model to output a prediction.
- Boosting learns sequentially in a very adaptative way and combines them following a deterministic strategy.

Questions