Machine Learning in 5 minutes

Machine Learning (ML): machine learning does not involve consciousness like human learning. ML is a learning approach that finds patterns (inferences) in the data via statistics, mathematics, technology, and trial and error. The type of learning task (question to answer) determines the learning approach. Effective ML requires extensive data and sufficient computing power to train the algorithms.

empowerment quality engineering

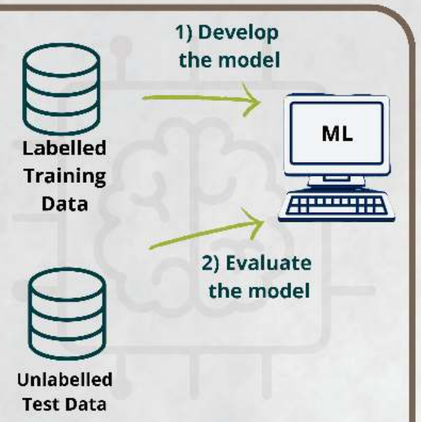
Supervised Learning

1) Maps an input variable to an output variable based on labelled examples or a training data set.

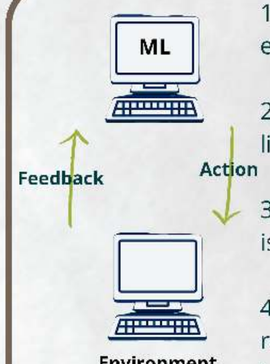
2) The algorithm identifies unlabelled examples from the test data set.

3) The machine develops a rule to categorise new examples from a test set by processing the training set examples.

This requires 'structured' data to learn from.



Reinforced Learning



1) The algorithm receives feedback from the environment (user).

2) If the feedback is positive, the algorithm is likely to perform the action again.

3) If the feedback is negative, then algorithm is less likely to repeat the action.

4) The actions taken by the algorithm are reinforced based on feedback provided. A Learning by doing approach. No training data is required as the 'training' is via try, fail, repeat for given tasks.

Data Fundamentals

A variable is a feature or characteristic within a dataset that represents a measurable element of the data, influencing the model's predictions or outcomes.

For large datasets, it is not feasible to assess each data item. A sample is taken from the data set population. Based on the sampled analysis, inferences are made on the total population.

- **Data Quantity**: If the sample is too small, or the sample does not represent all the variables used, then a valid inference cannot be made.
- **Data Quality**: Is the level of error within the data, such as bias, measurement errors, or inaccuracy that will impact the reference (prediction) of the data population?

The more data types available to analyse then the more data is required to make the inferences.

Different data types will involve different ML approaches:

- **Structured Data**: data with a predefined format and length. Supervised ML is suited to processing structured data.
- **Unstructured Data**: data that does not have a predefined data model such as images, free text and audio. Works best with unsupervised learning.

Data types will impact the type of ML to be used for the solution.

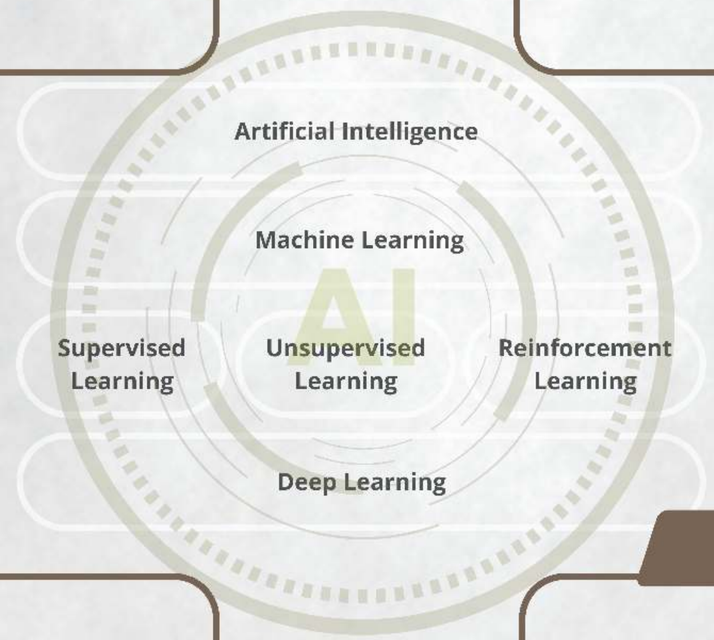
1) Prepare data: collect, prepare, split for training & testing.

ML Workflow

2) Train model: features & model is chosen.

3) Evaluate Model: trained model is assessed & improved.

4) Deploy model: to predict on new data. Monitor performance & retrain.

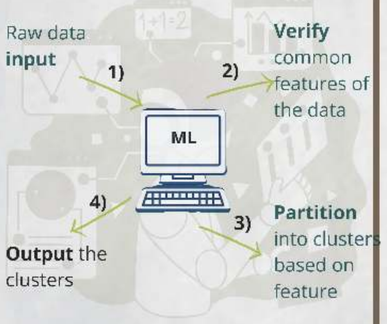


Unsupervised Learning

1) The algorithm learns to map an input variable to an output variable without using any labelled examples. There is no training set provided to the algorithm. The machine learns without being told what to learn.

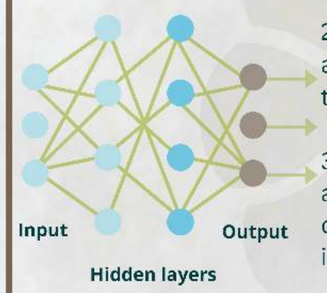
2) The algorithm tries to model the distribution of the data population in a test data set by looking for hidden patterns.

3) The most common method is clustering, where the disorganised data is clustered based in similar features.



Deep Learning

1) This is a neural network architecture, comprising of layers of networks with many connections between the layers.



2) The algorithm processes input data according to parameters, which are the 'weights' of the connections.

3) Learning is via the algorithm adjusting the weights of each connection based on the input data to improve and predict the output.

4) It can be combined with any of the other methods listed here.

e. info@empowermentqe.com | w. www.empowermentqe.com | t. +44 (0) 28 82 251 950