



Introduction to Machine Learning

By Ayub Odhiambo



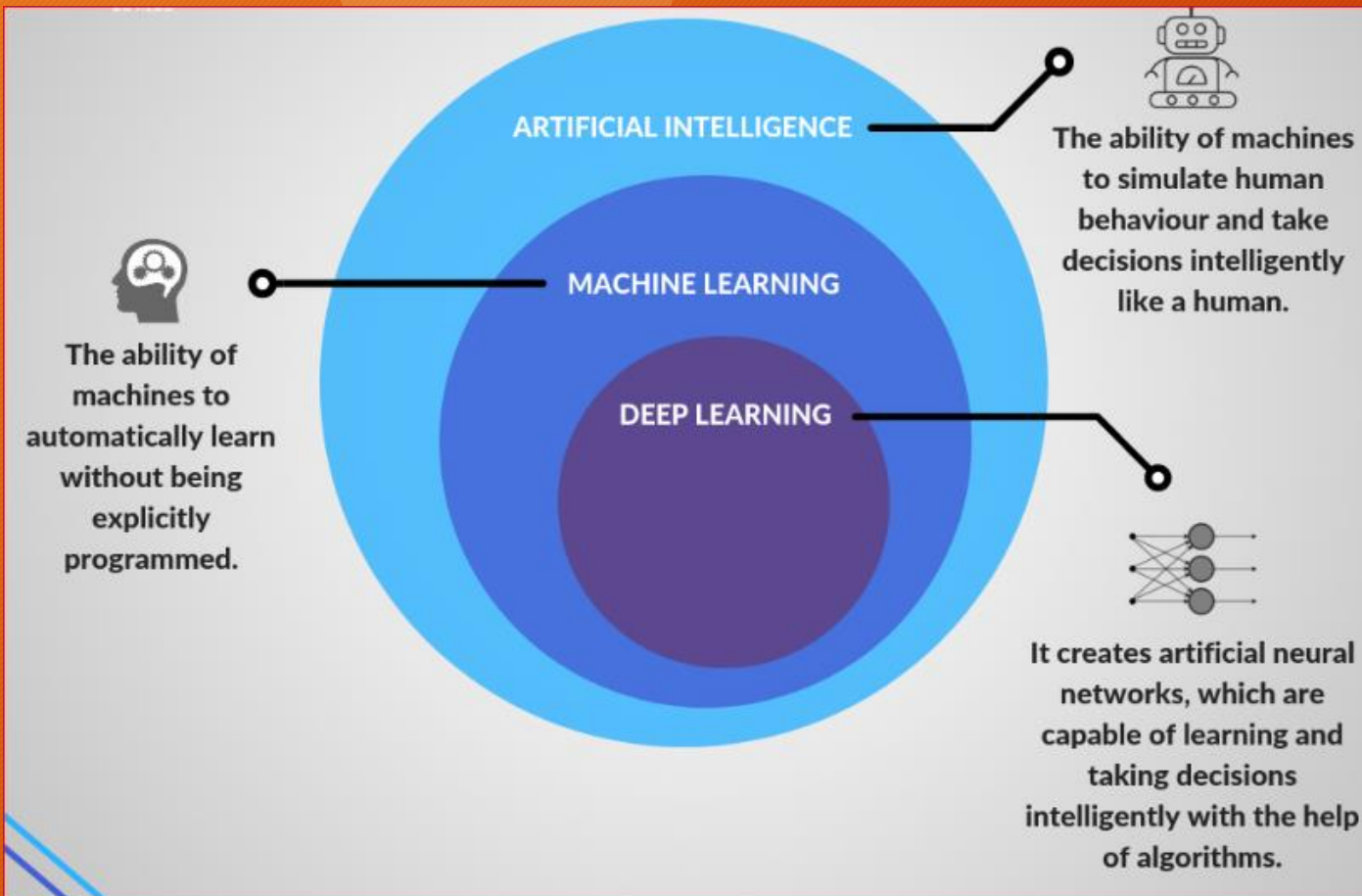
Objectives

- What is machine learning
- Machine learning terms and definitions
- Design thinking with machine learning
- Machine learning process
- What is the need for machine learning
- Types of machine learning
- Types of problems solved using machine learning
- Practical implementations of machine learning





What is machine learning



Traditional Programming



Machine Learning





Machine learning terms and definitions



- **Algorithm** - set of rules and statistical techniques used to learn patterns from data and draw significant information from it.
- **Model** - this is the main component of machine learning and it is trained using an algorithm. An algorithm maps all the decisions a model is to take based on given inputs
- **Predictor Variables** - features of the data that can be used to predict the output
- **Response Variables** - this is the output variable that needs to be predicted.
- **Training Data** - The model is built using the training dataset. The training dataset helps the model to identify key trends and patterns essential to predict the output.
- **Testing Data** - After the model is trained it needs to be tested to evaluate how it can predict an outcome. This is done using a testing dataset.



Design Thinking with Machine Learning



1. Understand

1. **Empathize** - Conduct research to develop an understanding of your users
2. **Define** - Combine all your research and observe where your user problems exist

2. Explore

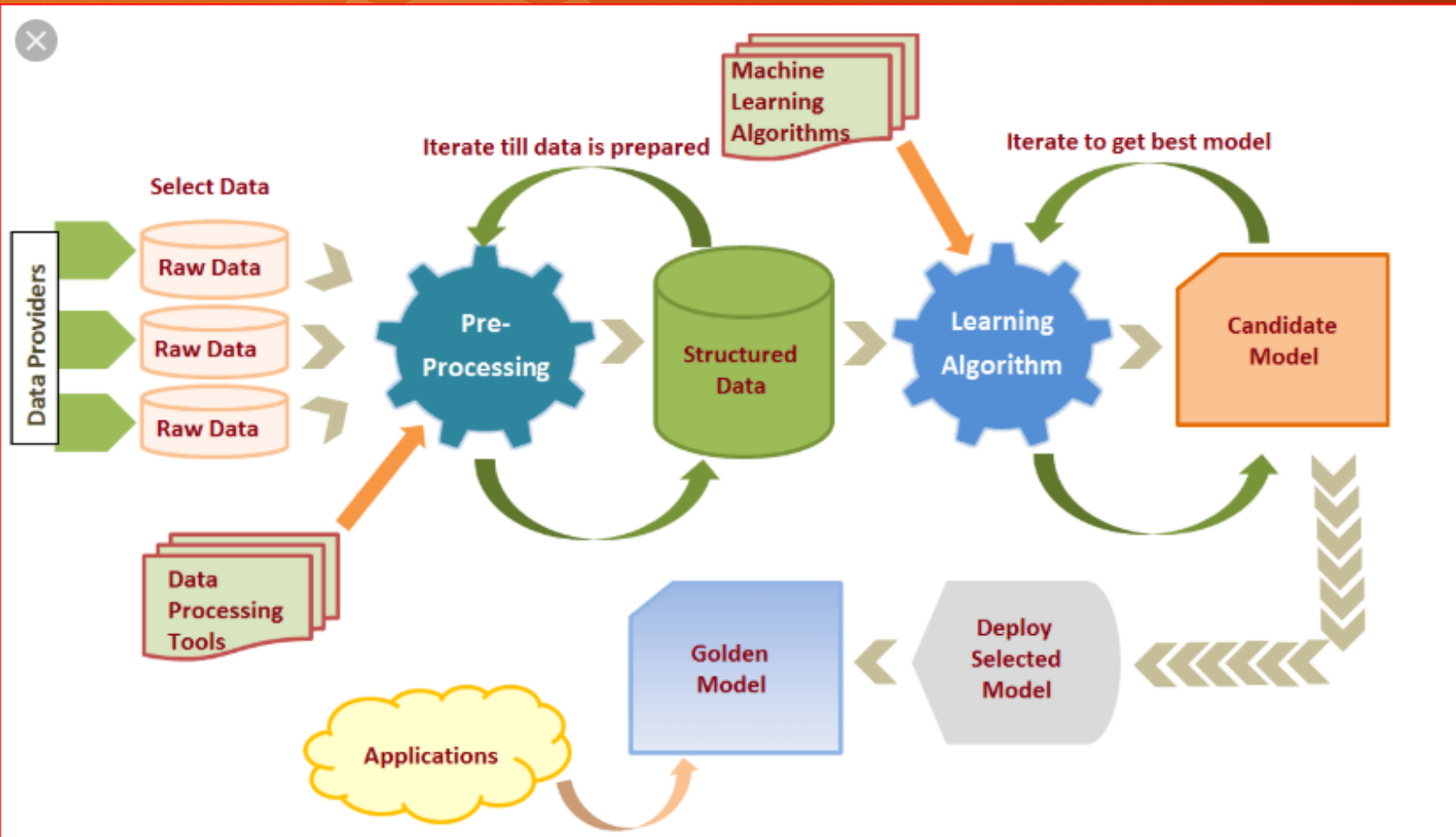
1. **Ideate** - Generate a range of crazy and creative ideas
2. **Prototype** - Build real, tactile representations for a range of your ideas

3. Materialize

1. **Test** - Return to the users for feedback
2. **Implement** - Put the vision into effect



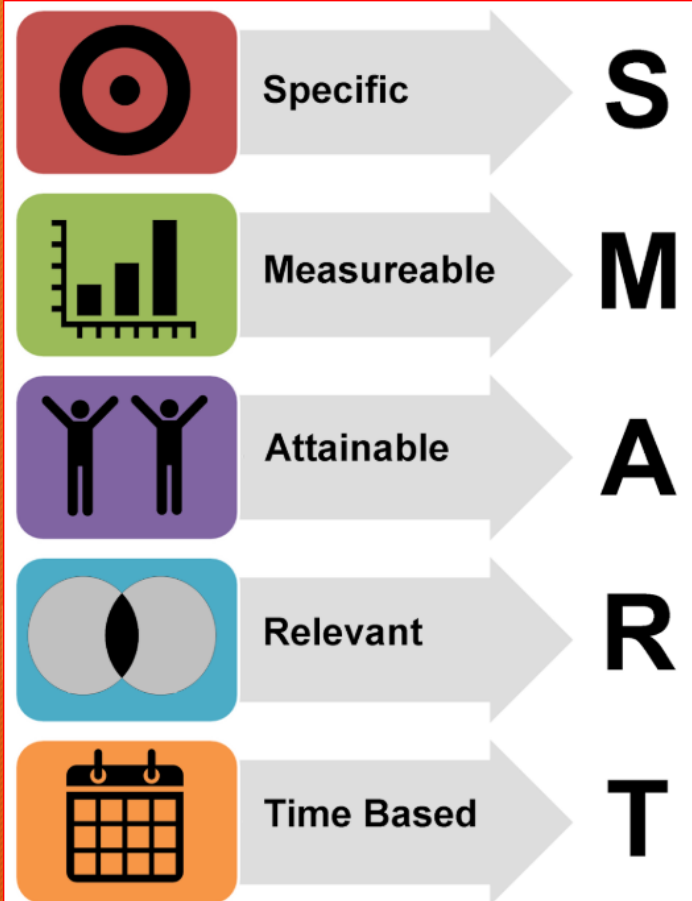
Machine learning process (MLP) Overview



1. Objective Setting
2. Data Understanding
 - Data Assessment
 - Data Exploration
3. Building the model
4. Model Evaluation
5. Predictions



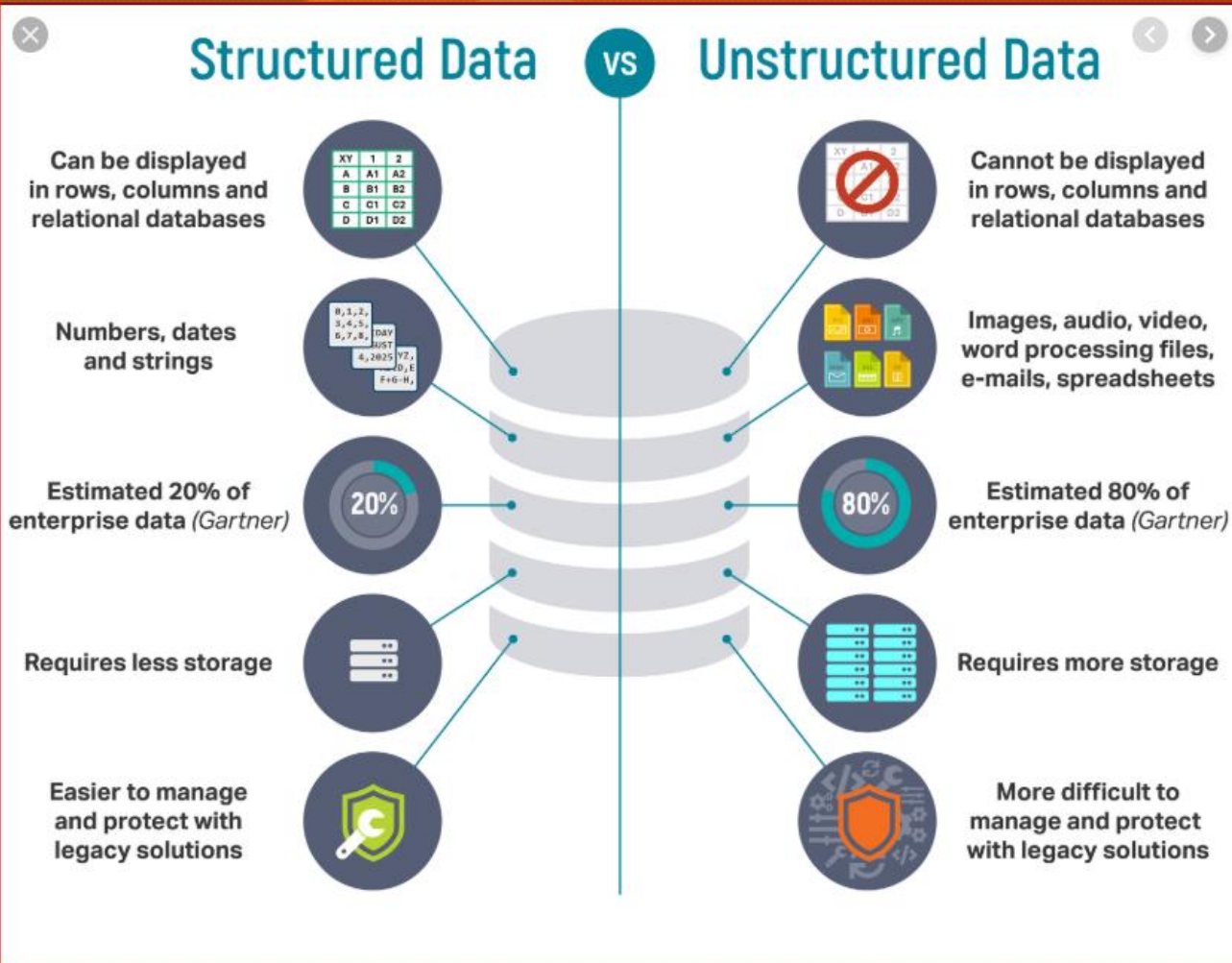
MLP Step 1 : Objective Setting



1. Have a clear understanding of what needs to be predicted
2. Document why the prediction is needed and the value it will bring
3. Take note of the duration for delivery
4. Take mental notes on what kind of data may be required to solve the problem.
5. Identify all the key stakeholders



MLP Step 2 : Data Understanding



Data Assessment

- Evaluates what data is available and how it aligns to the business problem

Data Exploration

- This is the process for creating meaningful summaries of your data and it is particularly important when you are unfamiliar with the data



MLP Step 2a : Data Assessment

The Questions

- What data is available?
- How much data is available?
- Do you have access to the ground truth, the values you are trying to predict?
- What is the data format?
- Where is the data?
- How is the data accessed?
- Which fields are important?

- **Ground truth** - This refers to labelled data with correct values and/or appropriate category that the model should learn to predict.
- **Data Relevance** - Take note of the data points that intuitively makes sense that a machine learning model could predict with
- **Data Quantity** - There is no strict formula for this but the rule of thumb should be at least you have ten times the number of parameters the model needs to train.



MLP Step 2b : Data Exploration

The Questions

- How many records do you have?
 - What are the data-types?
 - Do we have any missing values?
 - What is the distribution of each column?
 - Do we have any outliers?
 - Is the data balanced?
 - Is there any bias in the dataset?
- **Missing Values**- It is important to understand if there is a reason or pattern for the missing value.
 - **Outliers**- This is a data-point that is significantly different from other observations.
 - **Unbalanced Dataset** - A dataset is unbalanced if each class does not have similar number of examples

Exploratory Data Analysis with Jupyter

Hands-On practical sessions.

