Practice ML with Examples

No-Code

PETRO-ANALYST

Ahmed Abd Elgawad

About ME | My Portfolio

Ahmed Elsayed Abdelgawad

Junior reservoir engineer , Data analyst , Python developer and ML developer.

I am very interested an passionate about digitalization and digital transformation in the petroleum industry. I have built and still building several project connecting my programming skill with the petroleum engineering knowledge like (Jet Pump Optimization desktop application, Customized Production Data Fetching pipeline from Outlook down to OFM and PowerBi, DCA desktop application , Artificial lift selection desktop application, Simple Nodal Analysis desktop application).

LinkedIn | Github | Youtube | Kaggle | Medium

Workshop Content

The purpose of the Workshop

- Demonstrate basics of ML and how you can apply it using No-Code.
- Practice with simple examples no everyone can follow along.

Content

- Data science life Cycle.
- What is ML and its types and overview of its types.
- Hands on practice using Orange software. (Regression & Classification).

What is will not Cover

- ML Model details
- Go in details , just overview and basics.

Session Flow

Data Science Life Cycle

Just an Introduction

What is ML

Intro

When To Do ML

Why Python

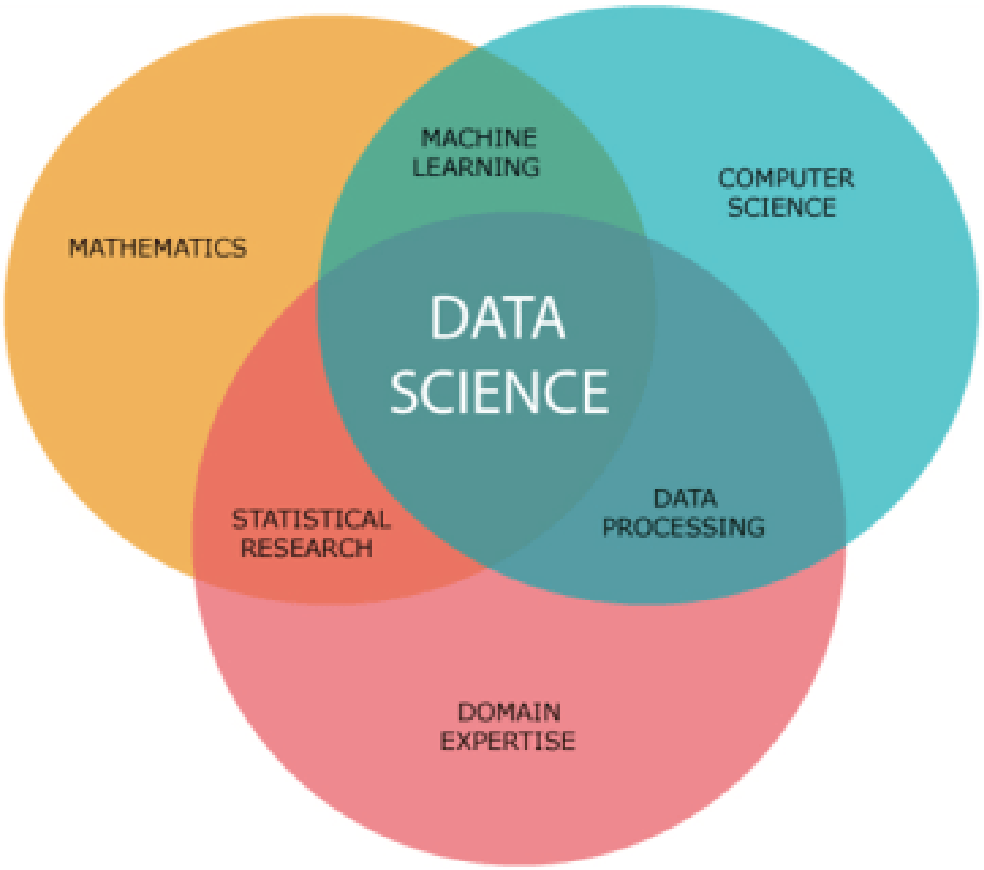
Data Science Life Cycle

Just an Introduction

PETRO-ANALYST

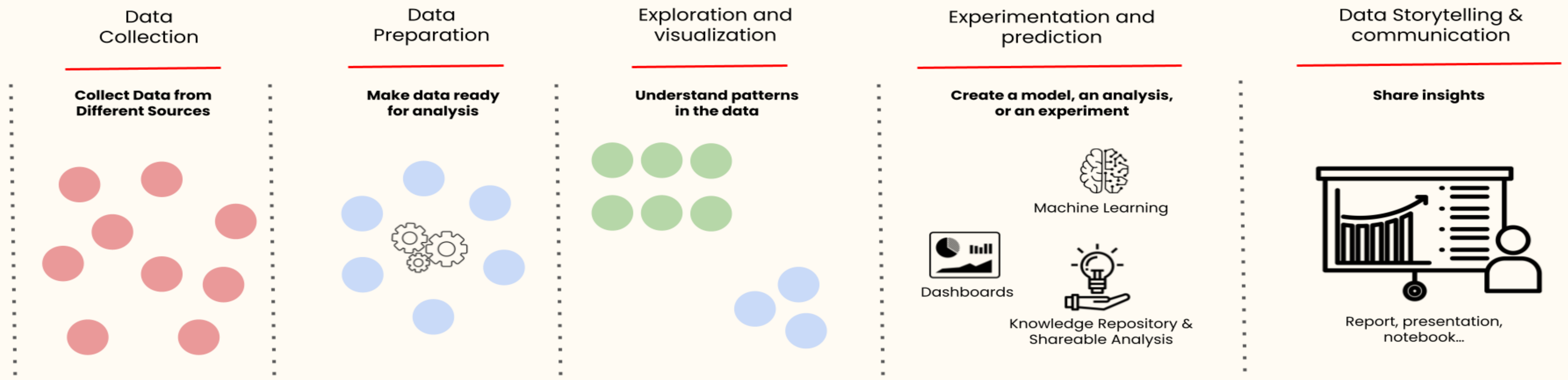
Ahmed Abd Elgawad

What is DS


MACHINE LEARNING
COMPUTER SCIENCE
MATHEMATICS
DATA SCIENCE
STATISTICAL RESEARCH
DATA PROCESSING
DOMAIN EXPERTISE

DS Life Cycle

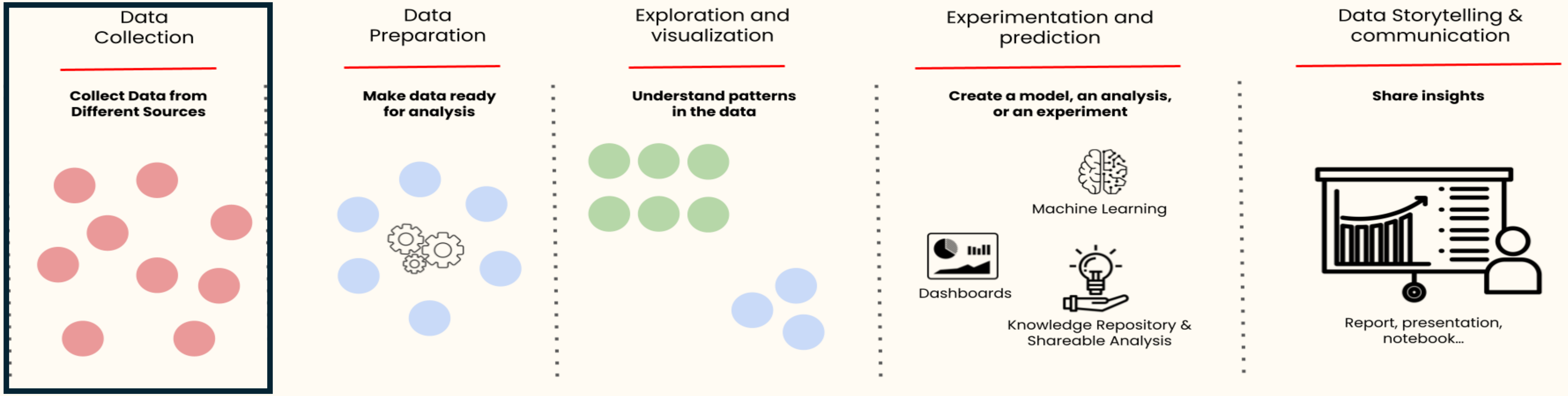
A problem that you need to solve, Using the Data!



DS Life Cycle

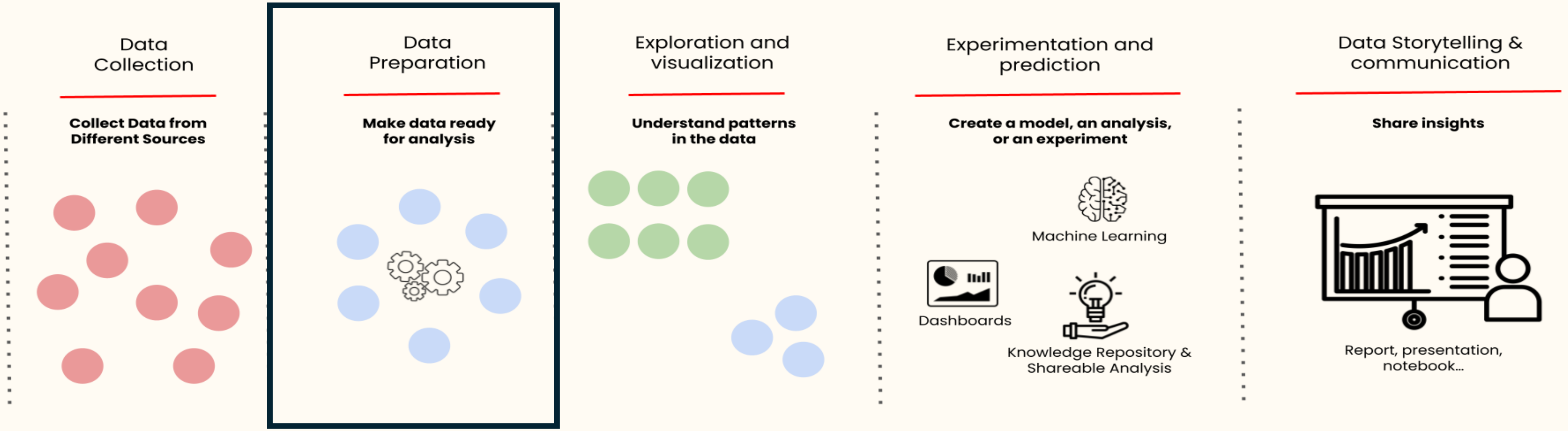
Unstructured Data

Excel File, Tables .. → Structured / Tabular Data



DS Life Cycle

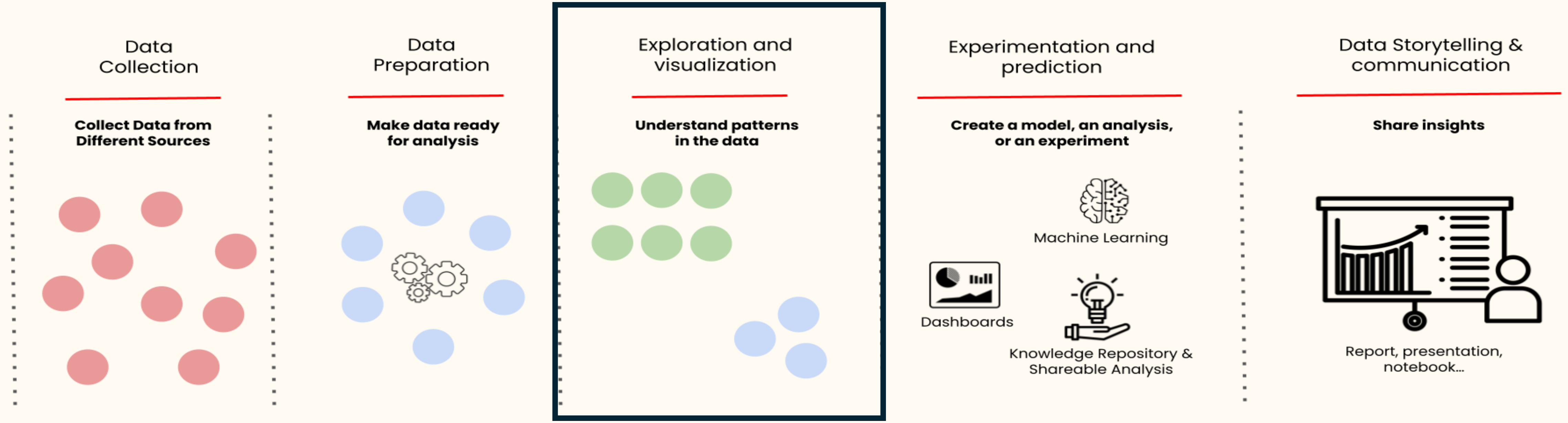
Create a clean, high quality dataset that yield accurate and reliable analytical results



Garbage In -> Model -> Garbage Out

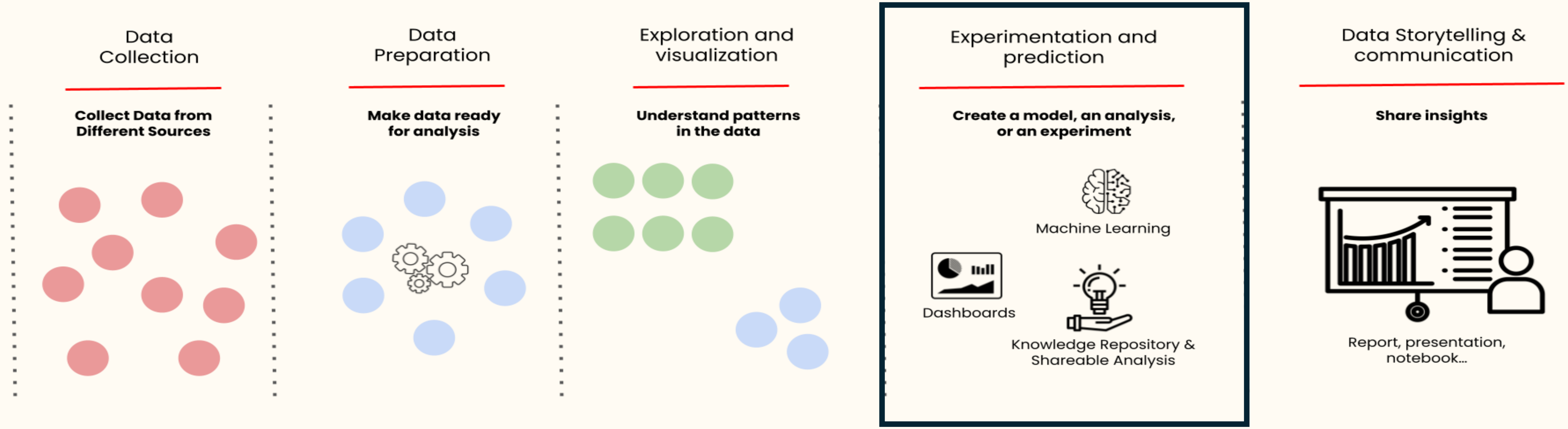
DS Life Cycle

Perform Statistical Analysis / Data Analysis , Visualizing the Data using graphs to gain insights and recognize patterns in the data.



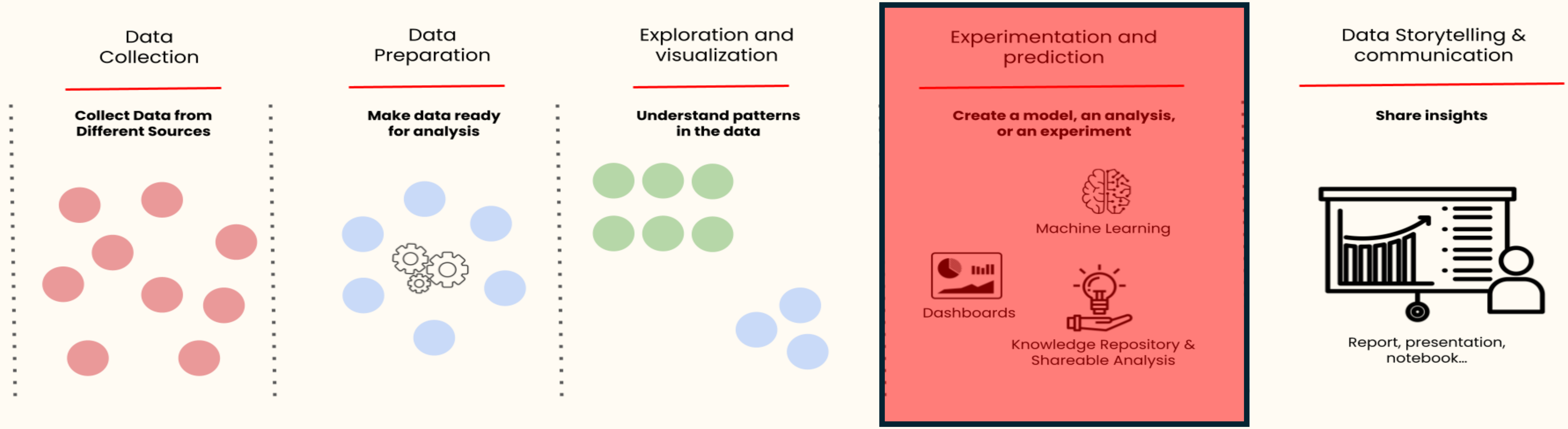
DS Life Cycle

Create Predictive Model (ML) , Dashboarding , Statistical Model.



DS Life Cycle

Create Predictive Model (ML) , Dashboarding , Statistical Model.



This step has no value without the previous steps

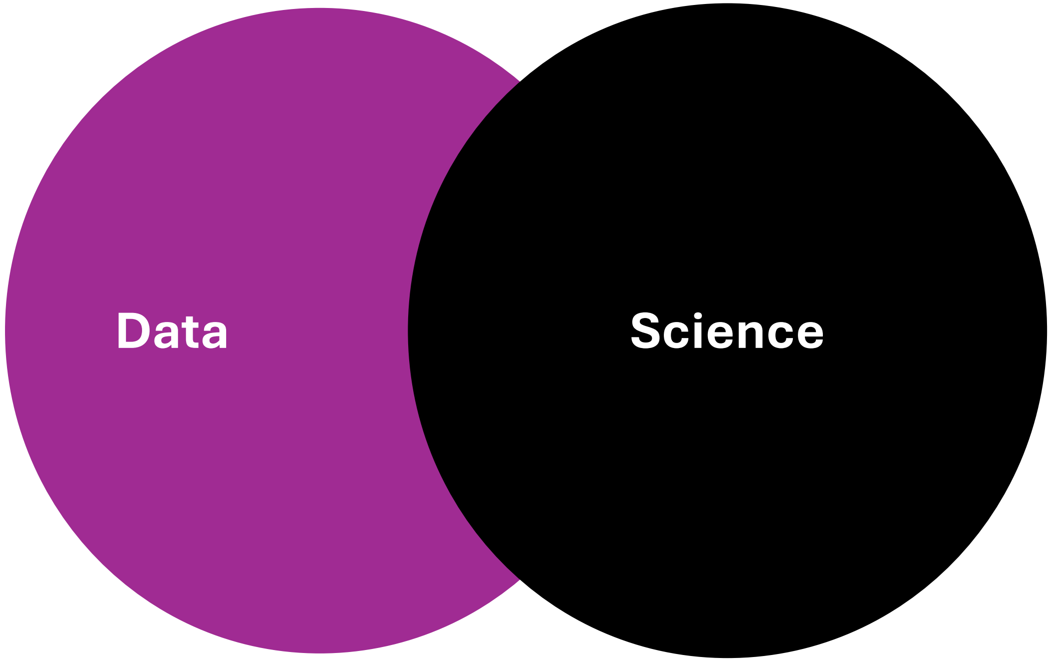
Data Preparation

Check for anomalies

Missing Data

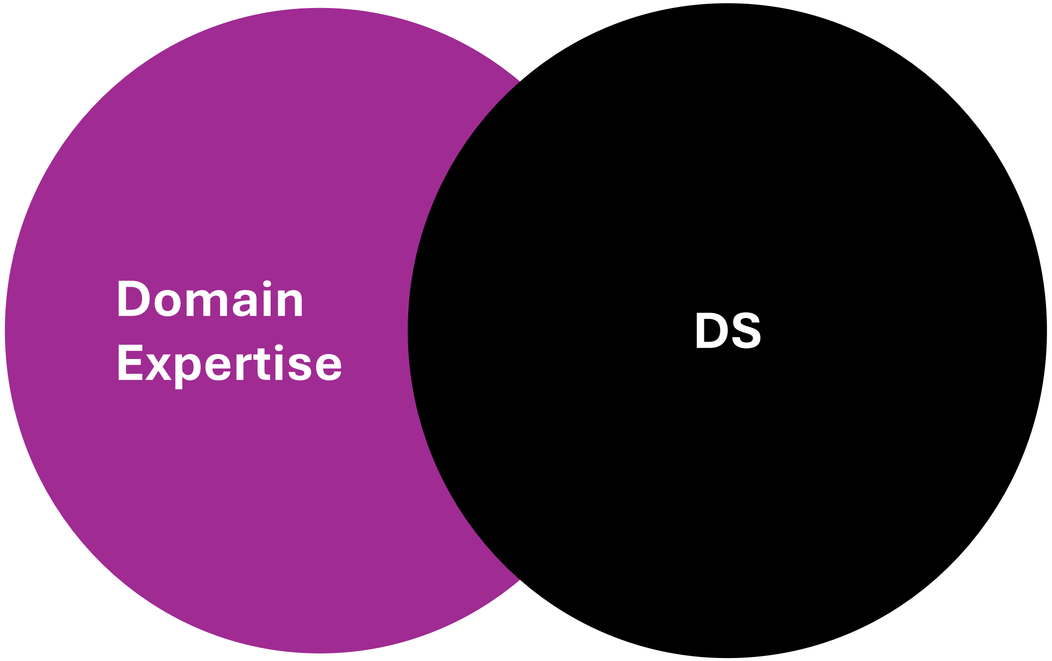
Typos

What We Can do



The Reality

- **DS/ML** are just tools.
- Through tutorials / Courses you will learn the tool.
- To work on real projects and add value you must have the **Domain Expertise**.

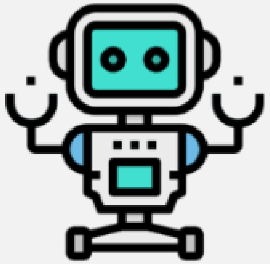


What is ML

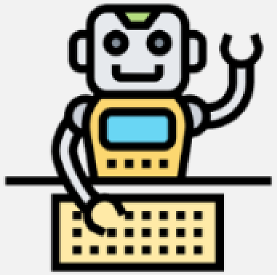
Intro

Types of ML

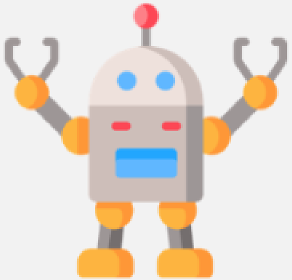
Supervised Learning



Unsupervised Learning

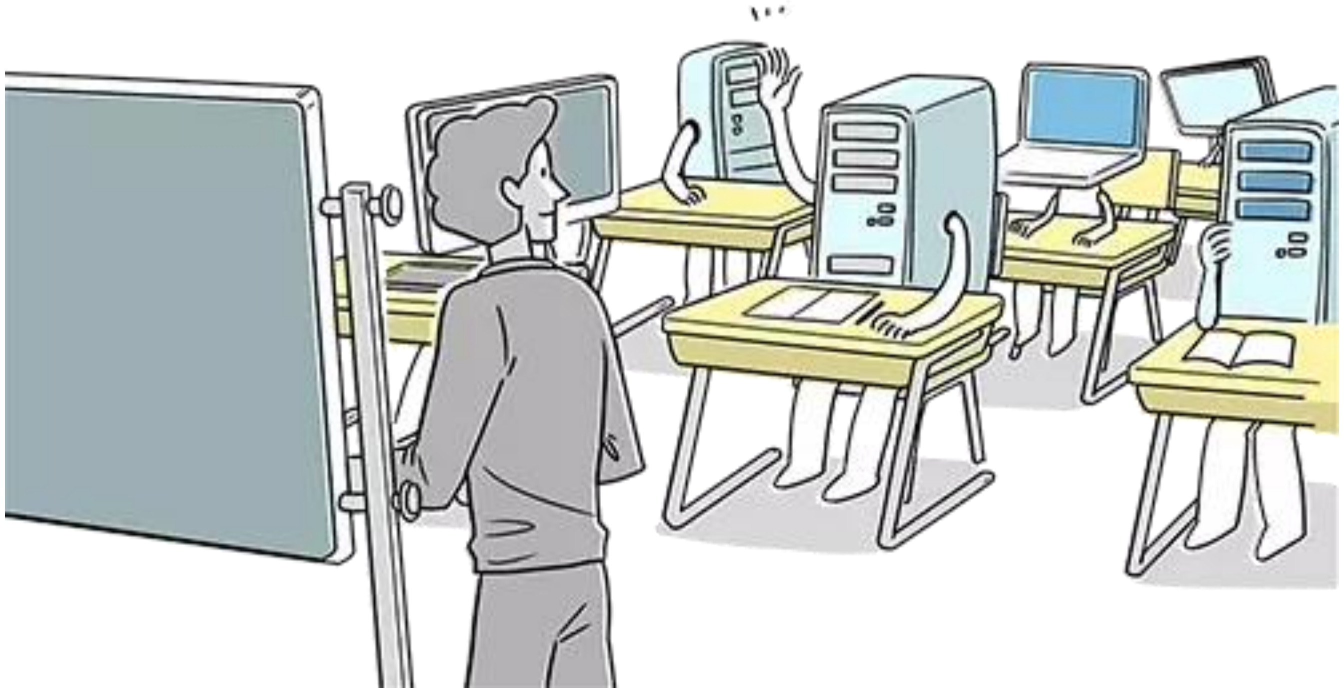


Reinforcement Learning

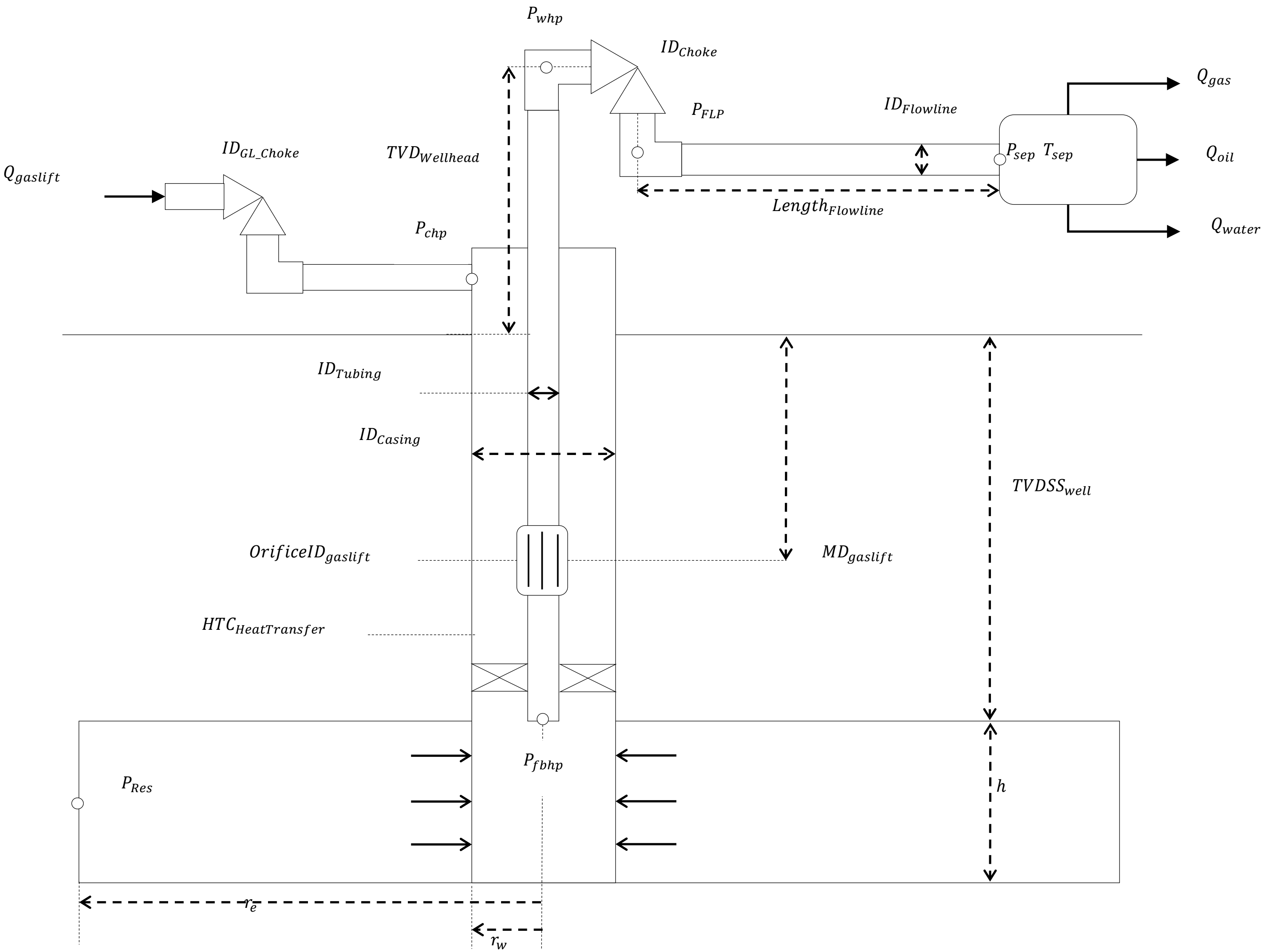


Supervised ML

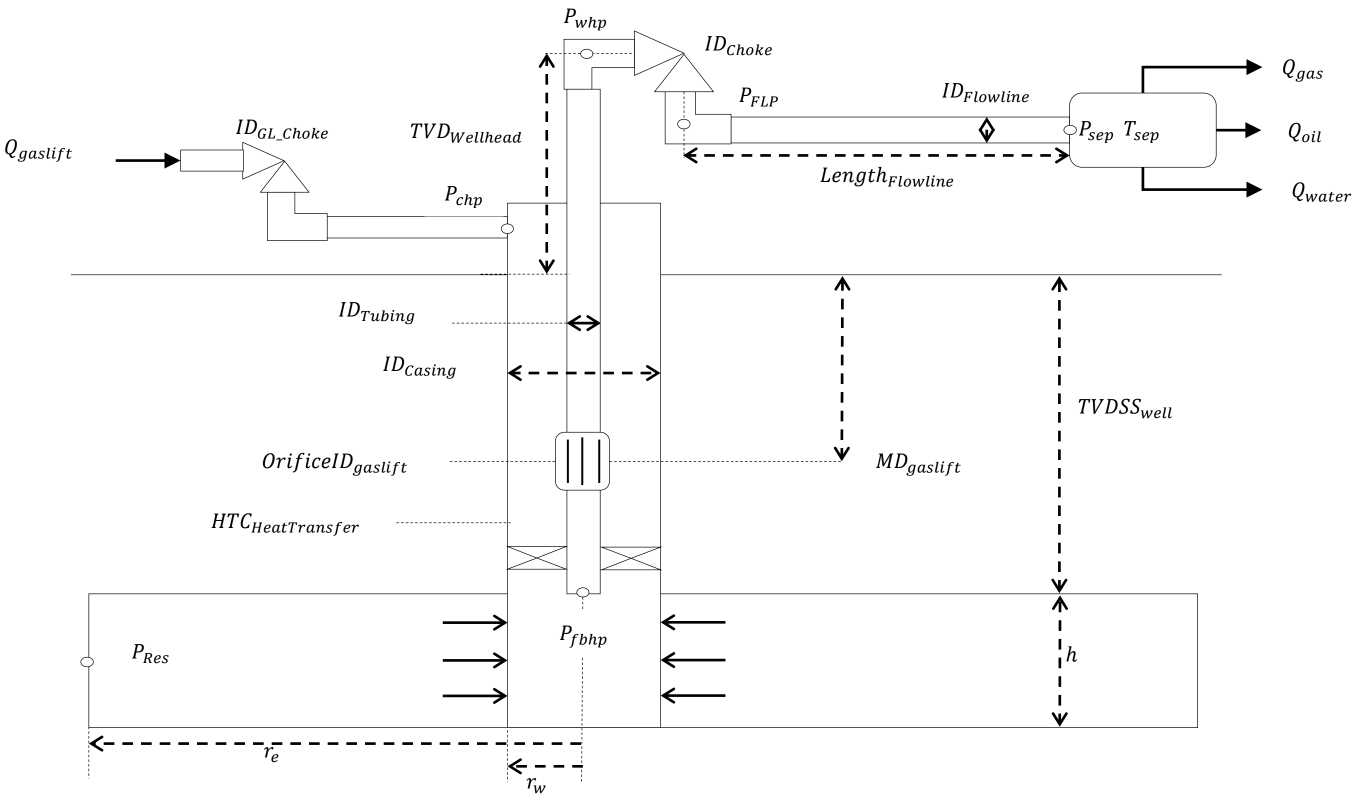
Supervised ML

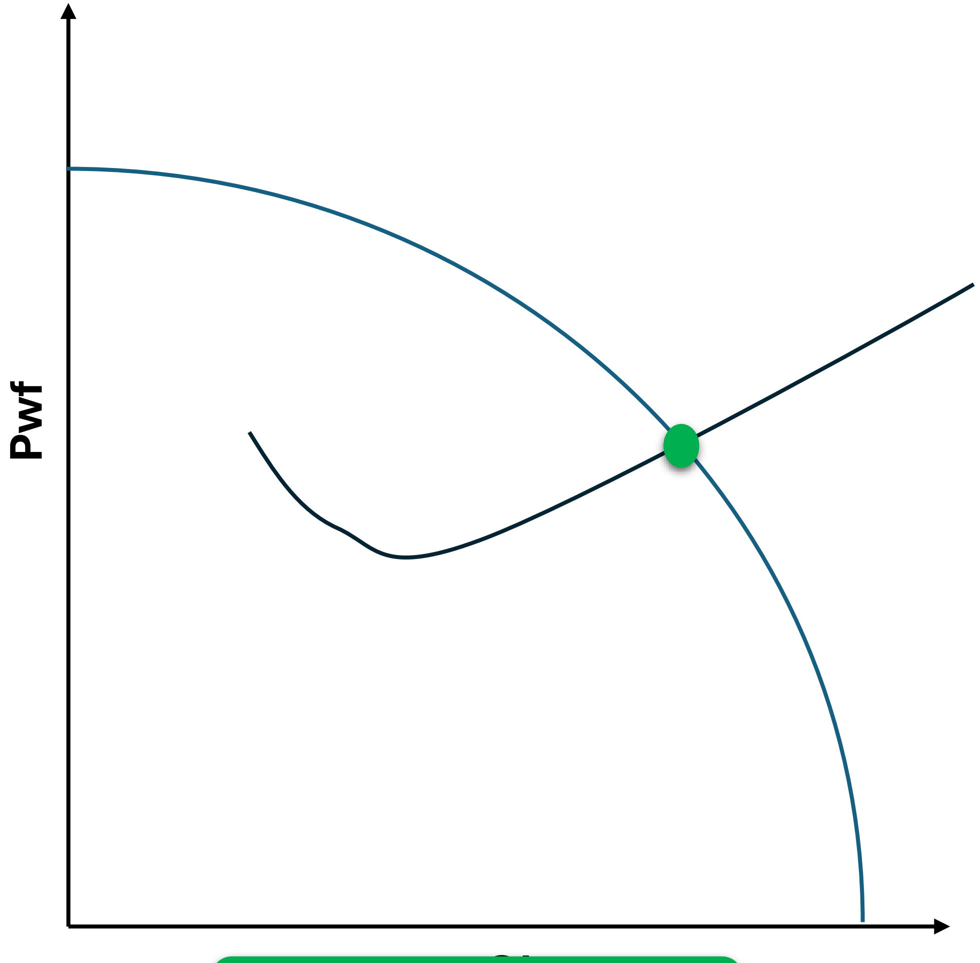


Virtual Flow Metering



Virtual Flow Metering | Physics Based





Input + Program | Rules = Output

ML Virtual Flow Metering | ML Based

Production Tests Dataset

BHP	WHP	WHT	Tsep	Psep	Choke_in	Qoil	Qwater	Qgas
5410.33	3185.75	83.3	60.32	100	0.25	954.6	0	2.39
5388	3015.38	86.88	60.65	100	0.25	801.9	200.5	2.01
5391.13	2808.8	90.83	61.19	100	0.25	634.7	423.2	1.59
5405.72	2515.99	95.11	61.97	100	0.25	448.1	672.2	1.12

Input + Output = Program | Rules

Machine Learning Vs Traditional Programming

Traditional Programming

Input + Program = Output

Using Machine learning

Input + Output = Program

Virtual Flow Metering

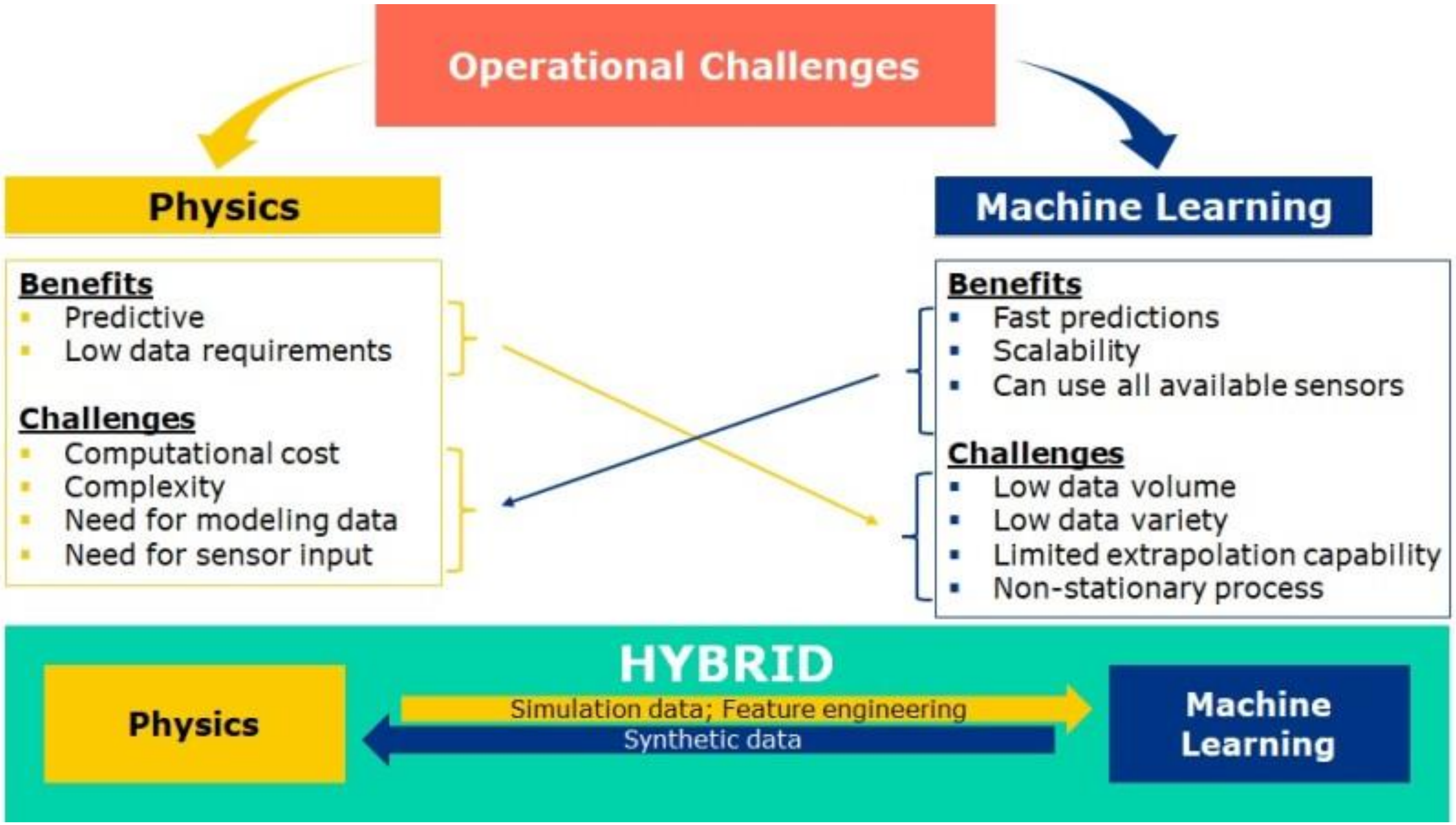
Physics Based

- Multi-phase flow simulation
- Combine thermodynamics, fluid dynamics, fluid modeling, optimization techniques.
- Consider it prosper put with sensor data ,downhole sensors (BHP, BHT), surface data (WHP , WHT)
- Calibrated on Production Test and back allocation

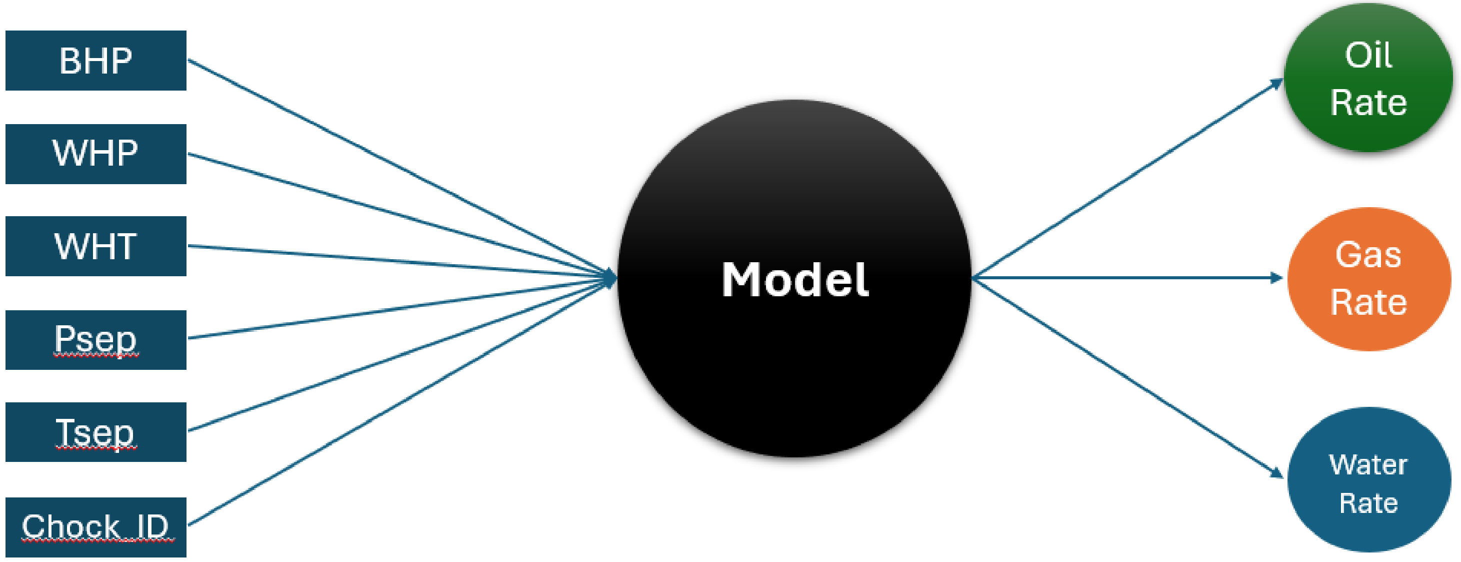
ML Based

- Real time flow rate estimation using only permeant sensor data.

Virtual Flow Metering



Model Structure



Supervised | How

X → **Model** → **Y**

data

Label

Supervision

BHP
BHT
WHP
WHT
Psep
Tsep
Chock_ID

Features

$Y = F(X)$, given x -> get Y

Oil
Gas
Water

Target

Supervised | How

Y

Label

Regression

Y -> Continuous Variable
- Pressure , Rates , Prices

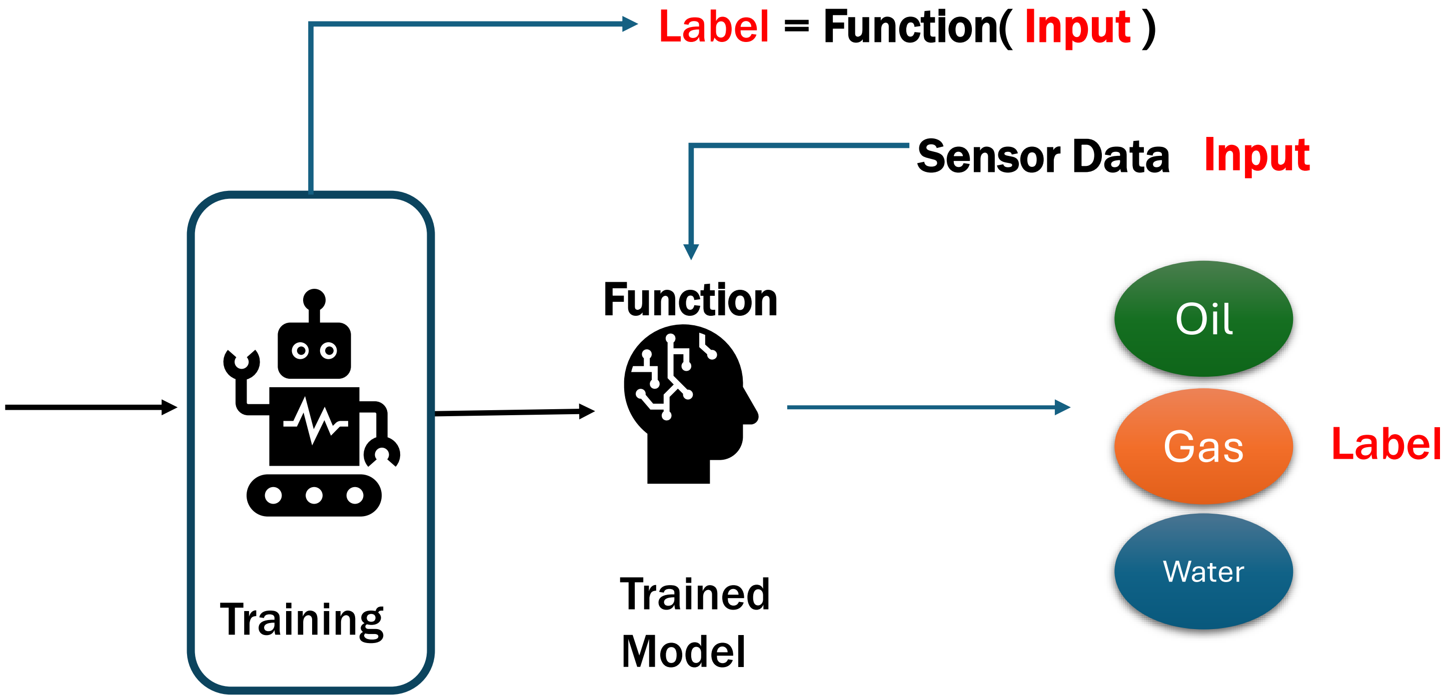
Classification

Y -> Discrete Variable
- Lithology / Facies , Classes

Model Training and Validation

BHP	WHP	WHT	Tsep	Psep	Choke_in
5410.33	3185.75	83.3	60.32	100	0.25
5388	3015.38	86.88	60.65	100	0.25
5391.13	2808.8	90.83	61.19	100	0.25
5405.72	2515.99	95.11	61.97	100	0.25

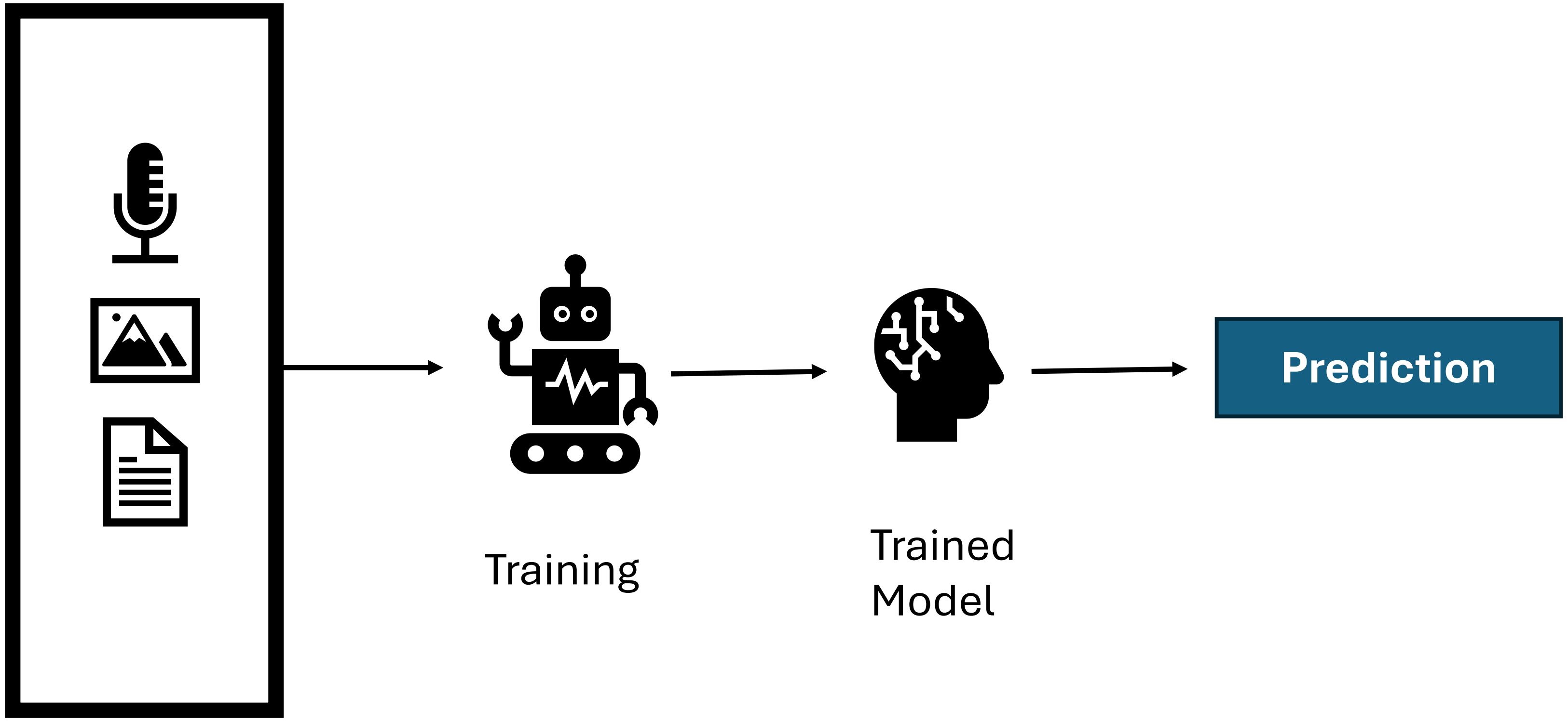
Qoil	Qwater	Qgas
954.6	0	2.39
801.9	200.5	2.01
634.7	423.2	1.59
448.1	672.2	1.12



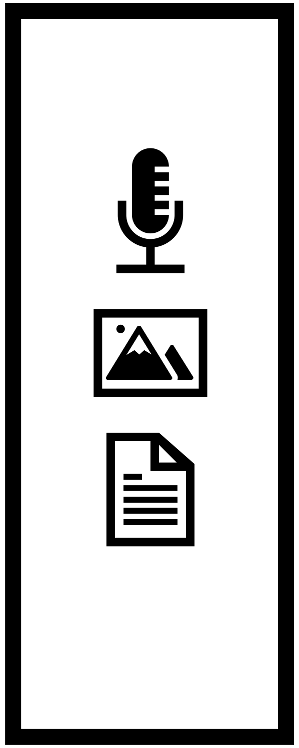
ML Workflow

Steps

ML Workflow



ML Workflow



What type of data?

- Structured data
- Unstructured data

Structured data

Feature 1	Feature 2	Feature 3	Feature 4	Target feature (Y)

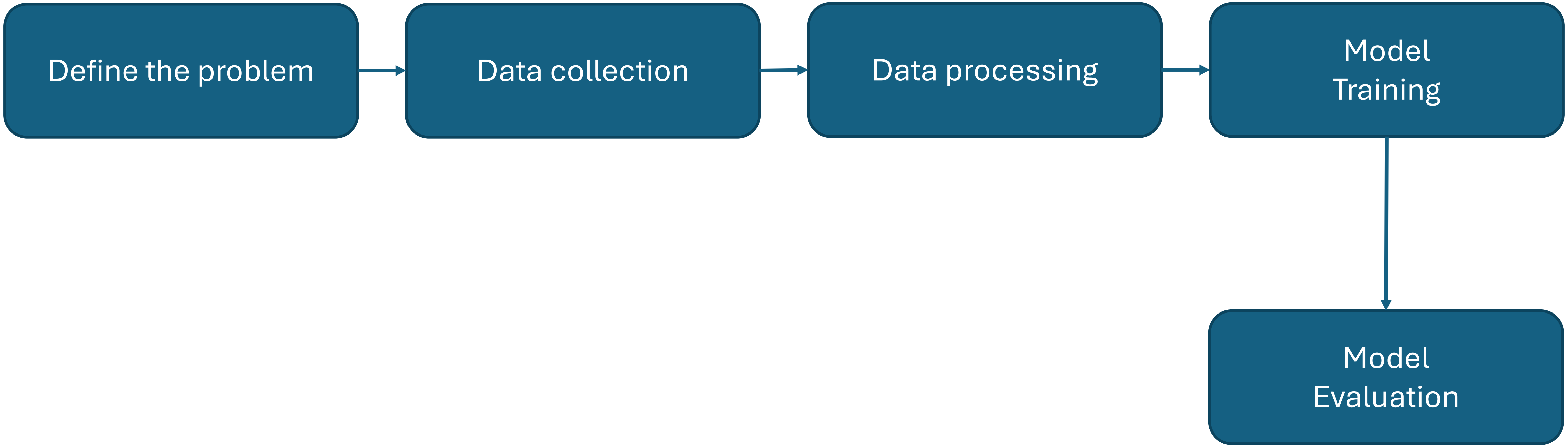
Unstructured data

Images

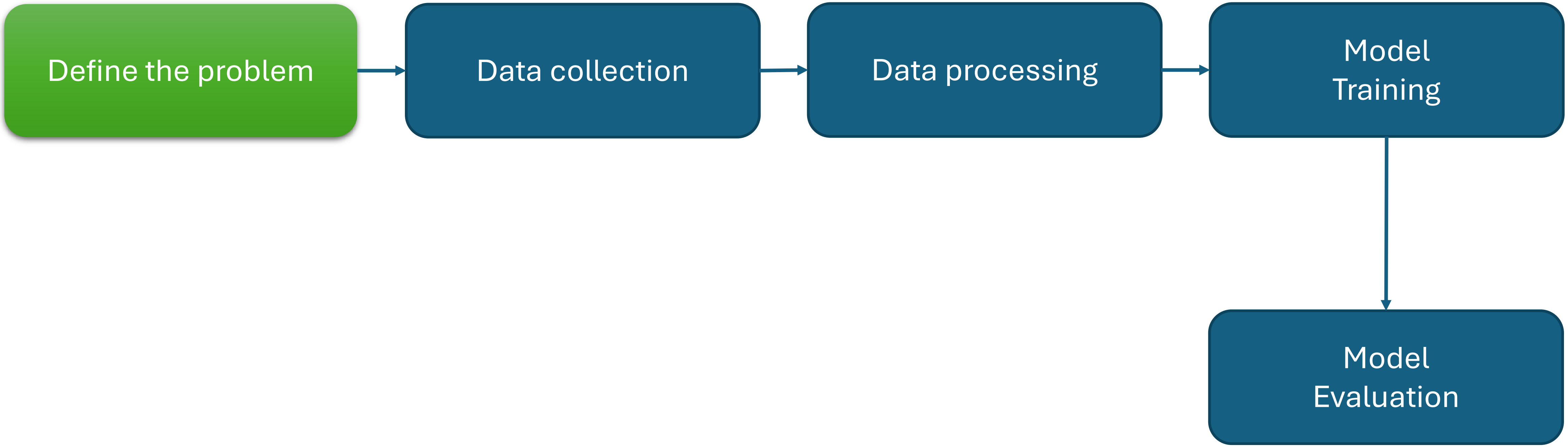
Audio

Text

ML Steps


Define the problem
Data collection
Data processing
Model
Training
Model
Evaluation

ML Steps


Define the problem
Data collection
Data processing
Model
Training
Model
Evaluation

Define the problem

Problem type

Classification

Regression

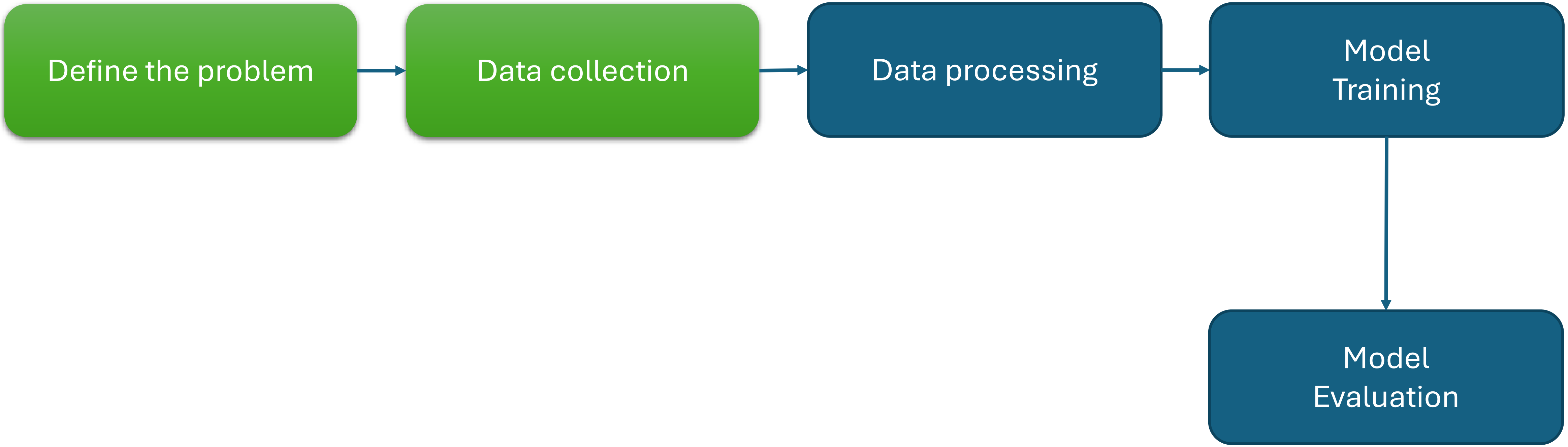
other

Baseline:

Beat the baseline

Successful project

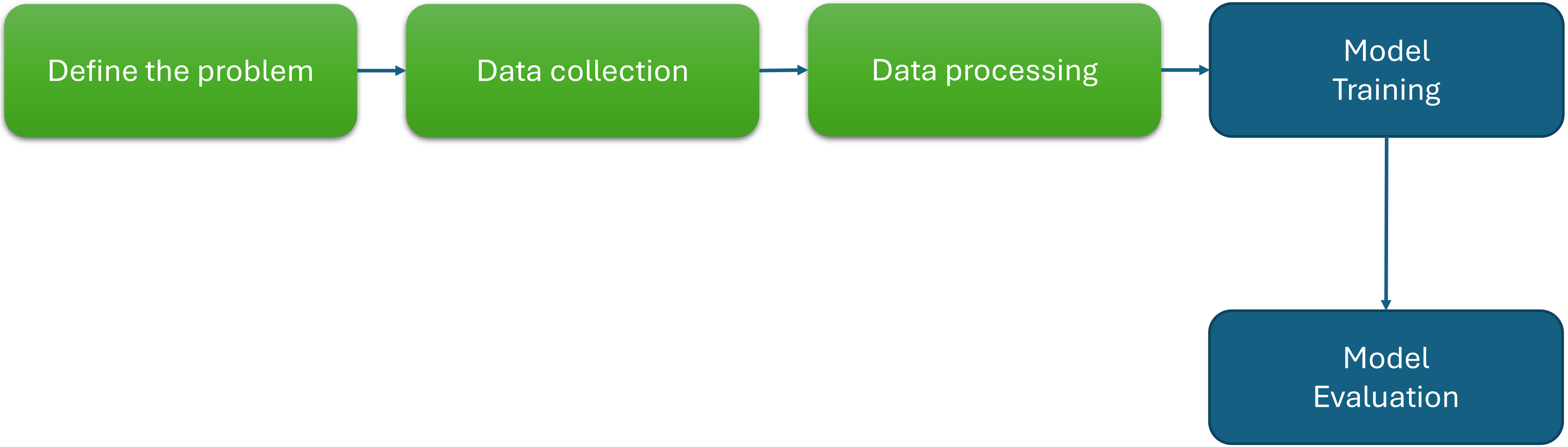
ML Steps


Define the problem
Data collection
Data processing
Model
Training
Model
Evaluation

Data Collection

Qoil	Qwater	Qgas	BHP	WHP	WHT	Tsep	Psep	Choke_in
954.6	0	2.39	5410.33	3185.75	83.3	60.32	100	0.25
801.9	200.5	2.01	5388	3015.38	86.88	60.65	100	0.25
634.7	423.2	1.59	5391.13	2808.8	90.83	61.19	100	0.25
448.1	672.2	1.12	5405.72	2515.99	95.11	61.97	100	0.25
238.9	955.7	0.6	5433.8	2059.65	99.88	63.1	100	0.25
954.6	0	2.39	5410.33	3185.75	83.3	60.32	200	0.25
801.9	200.5	2.01	5388	3015.38	86.88	60.65	200	0.25
634.7	423.2	1.59	5391.13	2808.8	90.83	61.19	200	0.25
448.1	672.2	1.12	5405.72	2515.99	95.11	61.97	200	0.25
238.9	955.7	0.6	5433.8	2059.65	99.88	63.1	200	0.25
954.6	0	2.39	5410.33	3185.75	83.3	60.32	300	0.25
801.9	200.5	2.01	5388	3015.38	86.88	60.65	300	0.25

ML Steps


Define the problem
Data collection
Data processing
Model Training
Model Evaluation

Data Processing

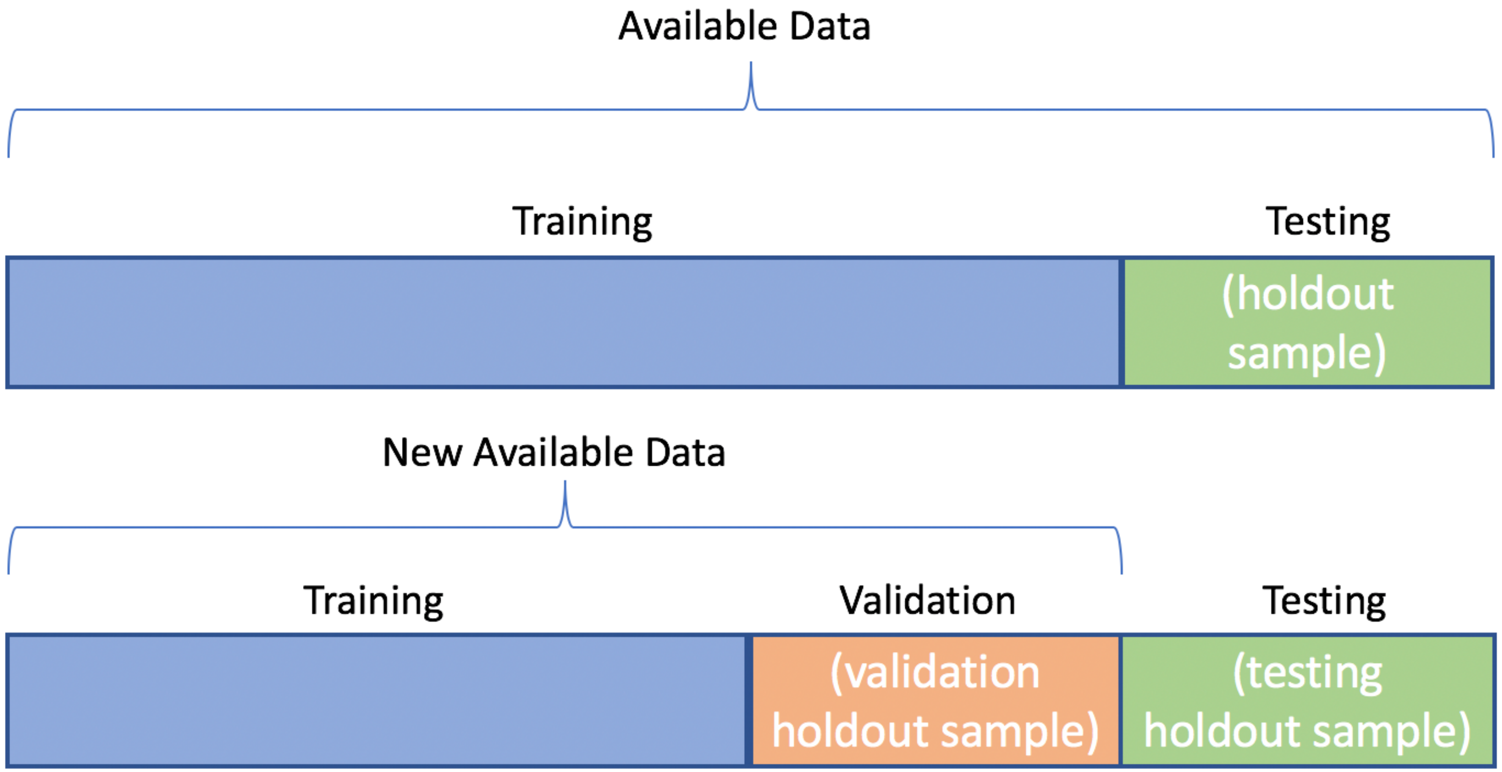
- Dealing with:
  - Missing data
  - Irrelevant data
  - Outliers

- Feature Engineering:
  - Deleting features / columns
  - Adding new features using existing one
  - Grouping features

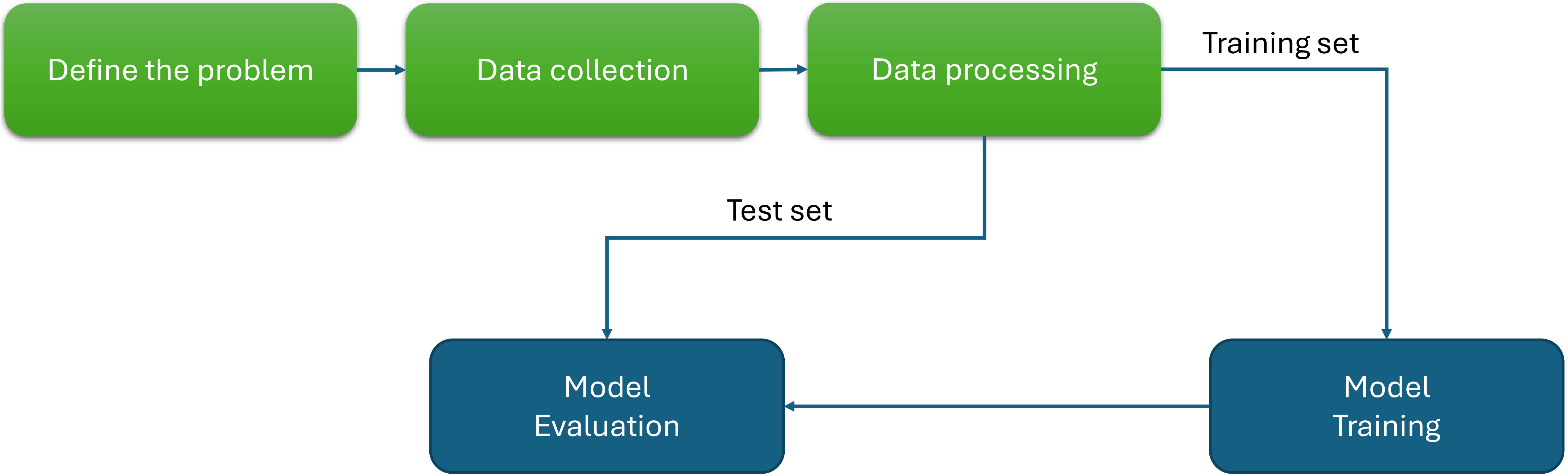
- Data Transformation:
  - normalizing the data
  - converting categorical data into numerical data
  - reducing the dimensionality of the data.

Data Preparation

- This step involves splitting the data into training and testing datasets.
  - The training dataset is used to train the ML model
  - Testing dataset is used to evaluate the performance of the model.



ML Steps


Define the problem
Data collection
Data processing
Training set
Test set
Model
Evaluation
Model
Training

Model Evaluation

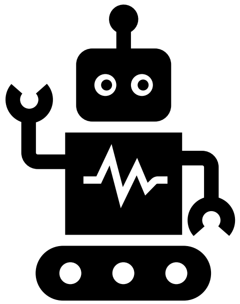
Training set

X

BHP	WHP	WHT	Tsep	Psep	Choke_in
5410.33	3185.75	83.3	60.32	100	0.25
5388	3015.38	86.88	60.65	100	0.25
5391.13	2808.8	90.83	61.19	100	0.25
5405.72	2515.99	95.11	61.97	100	0.25

Y

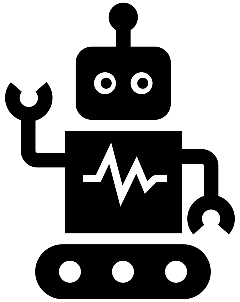
Qoil	Qwater	Qgas
954.6	0	2.39
801.9	200.5	2.01
634.7	423.2	1.59
448.1	672.2	1.12



Test set

X

BHP	WHP	WHT	Tsep	Psep	Choke_in
4140.61	982.36	170.84	120.94	300	1.94
4238.67	881.64	176.61	130.02	300	1.94
4436.07	724.66	181.97	139.03	300	1.94
4234.42	1178.02	159.75	105.42	400	1.94



Qoil	Qwater	Qgas
4007.2	2671.4	12.02
2919.7	4379.5	8.76
1605.2	6420.9	4.82
5967.3	0	17.9

Qoil	Qwater	Qgas
4010	2696	12
2930	4365	9.1
1610	6425	5.2
5960	0	15.3

Model Evaluation

Predicted

Qoil	Qwater	Qgas
4007.2	2671.4	12.02
2919.7	4379.5	8.76
1605.2	6420.9	4.82
5967.3	0	17.9

Actual

Qoil	Qwater	Qgas
4010	2696	12
2930	4365	9.1
1610	6425	5.2
5960	0	15.3

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

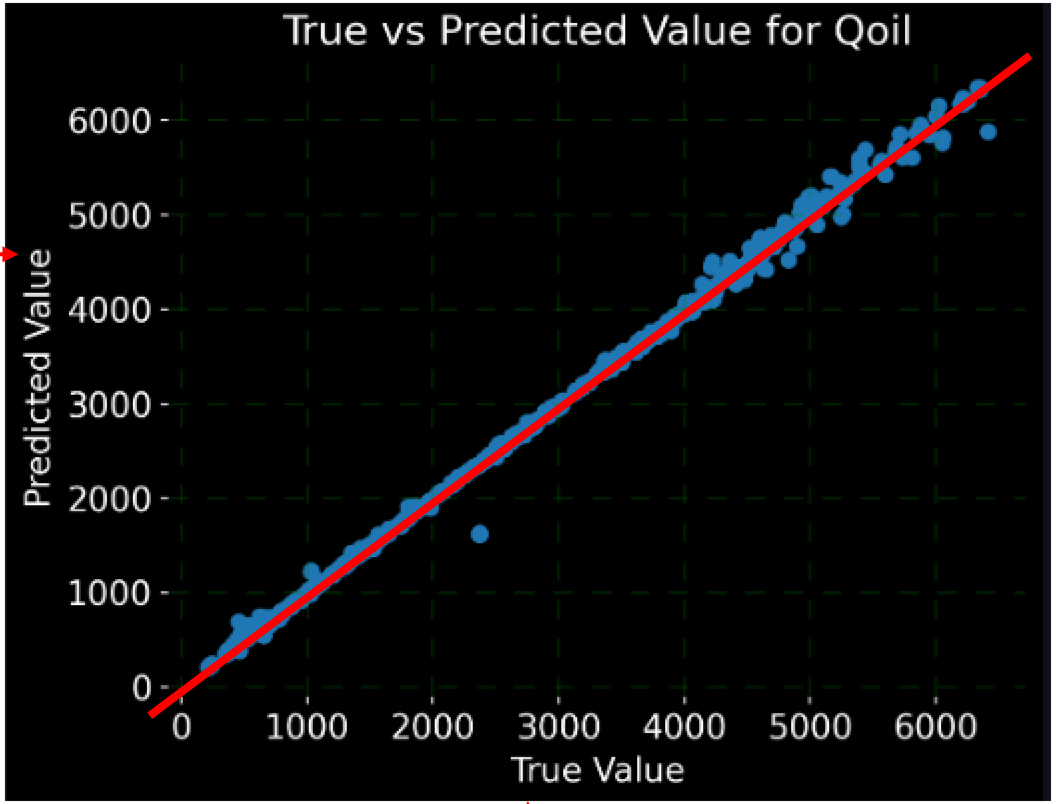
Model Evaluation

Predicted

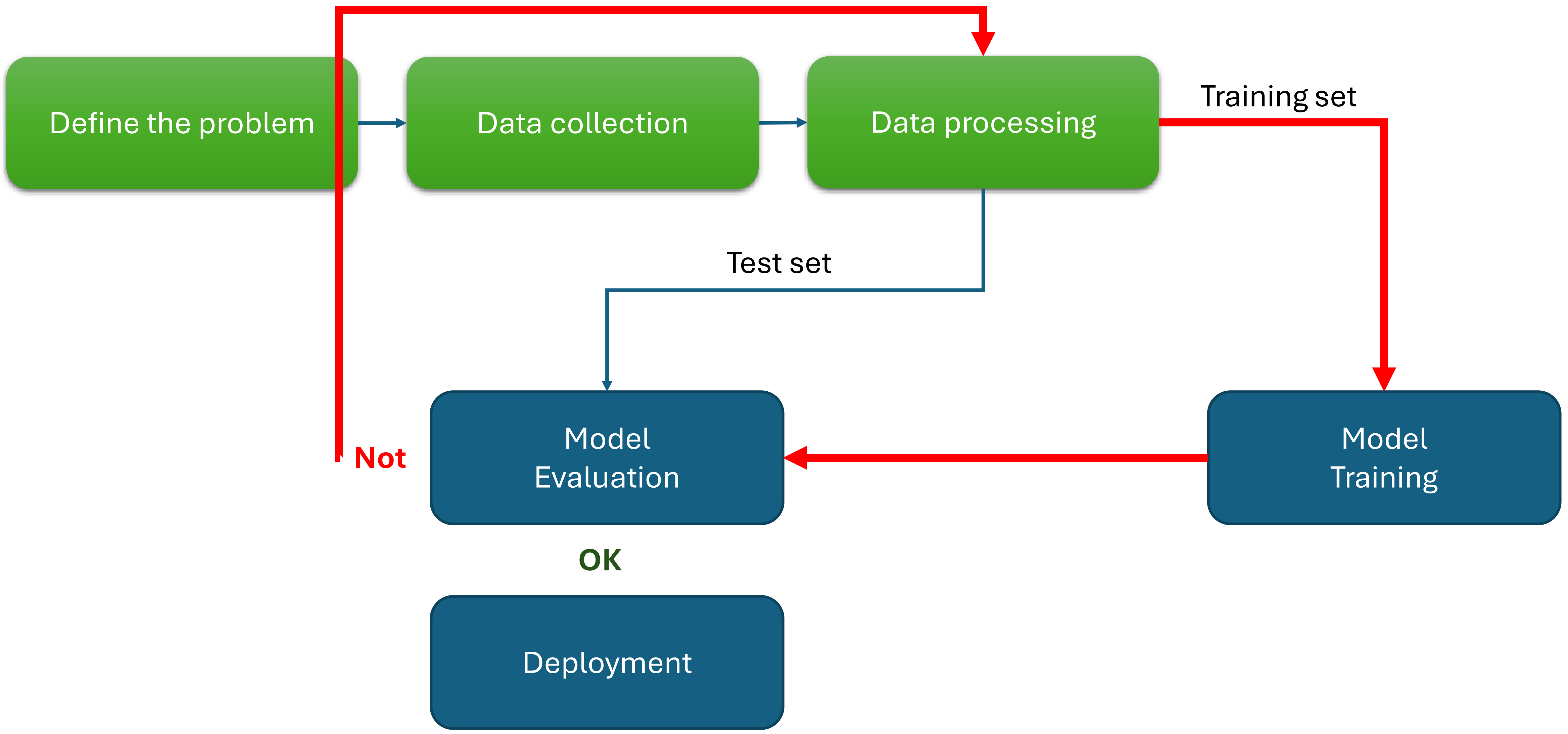
Qoil	Qwater	Qgas
4007.2	2671.4	12.02
2919.7	4379.5	8.76
1605.2	6420.9	4.82
5967.3	0	17.9

Actual

Qoil	Qwater	Qgas
4010	2696	12
2930	4365	9.1
1610	6425	5.2
5960	0	15.3

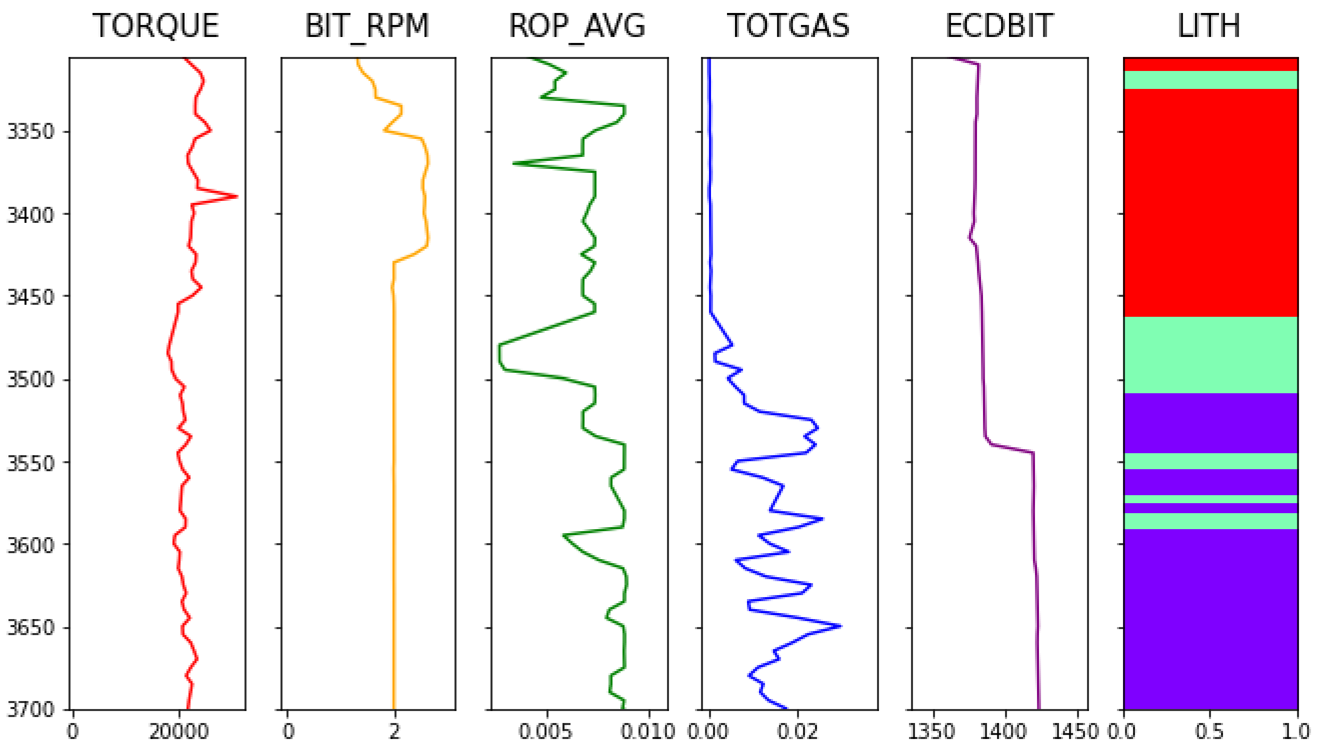


ML Steps

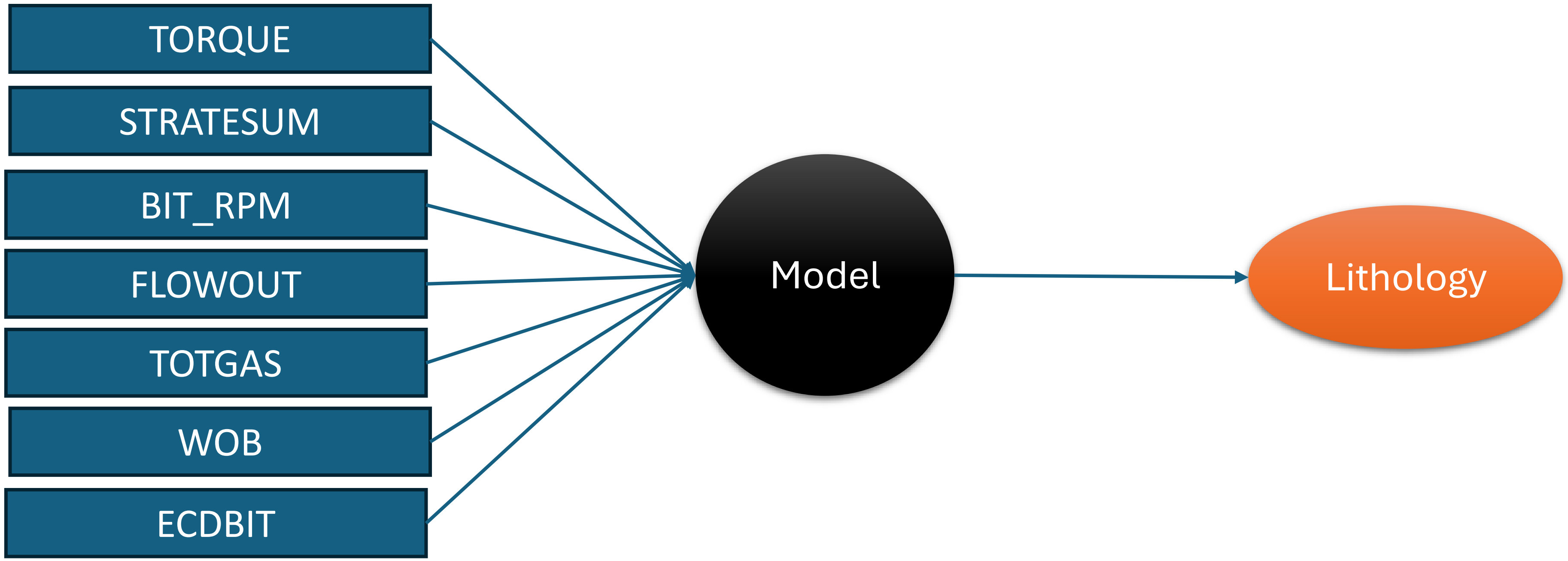

Define the problem
Data collection
Data processing
Training set
Test set
Model Evaluation
Model Training
Not
OK
Deployment

LITHO Prediction

- Features:
  - TORQUE: Surface torque average (N.m)
  - STRATESUM: Pump total stroke rate (Hz)
  - BIT_RPM: Rotation per minute at drill bit (c/s)
  - PUMP: Pump pressure (Pa)
  - FLOWOUT: Mud flow out (m3/s)
  - ROP_AVG: Depth averaged rotation per minute (m/s)
  - TOTGAS: Total gas content (ppm)
  - WOB: Weight on bit (N)
  - ECDBIT: Effective circulating density on bit (kg/m3)
- Target:
  - Lithology



LITHO Prediction | Model



Classification

$$\textbf{Accuracy} = \frac{num\ of\ correct\ predictions}{total\ num\ of\ samples}$$

Model

Data

90% is Sandstone

Just predict it is **Sandstone**

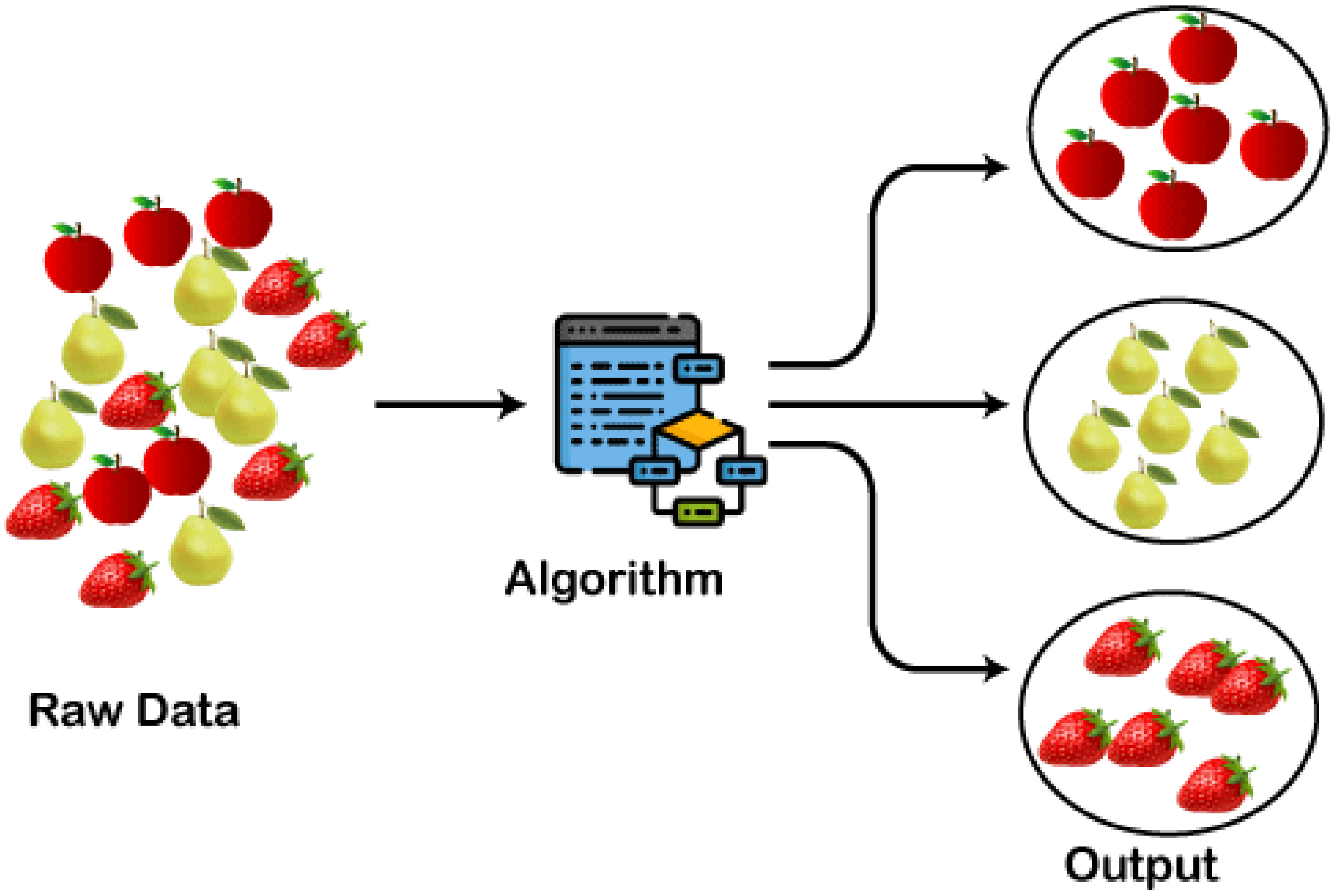
Unsupervised ML

Unsupervised ML

- No Supervision , **No label , No Y.**
- There are different type of algoritms, they do certain tasks given the data
  - Clustering
  - Association analysis
  - Dimensionality reduction

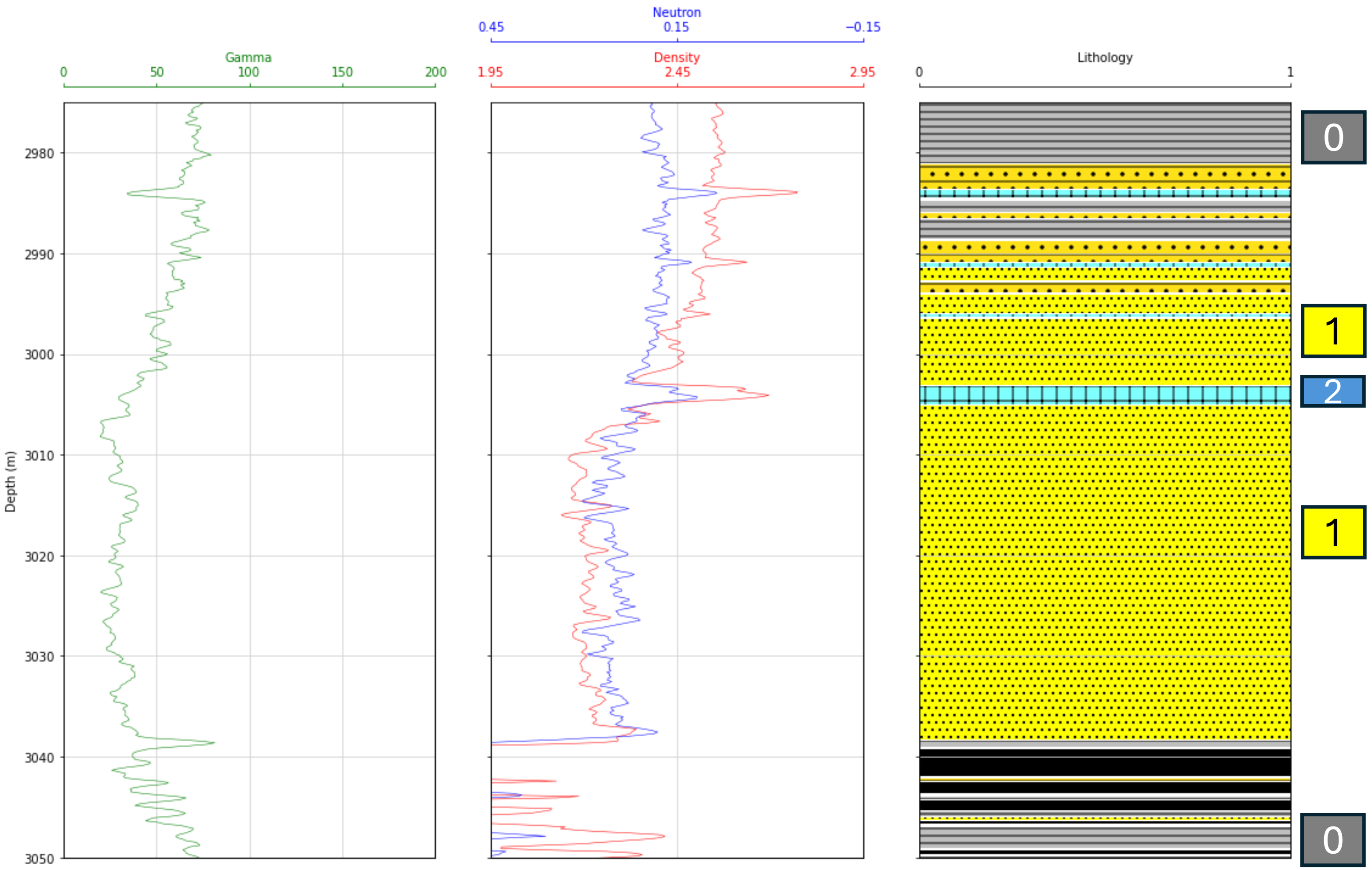
Unsupervised ML

- **Only X, no Y**
- **Clustering**
- Association
- Dimensionality reduction



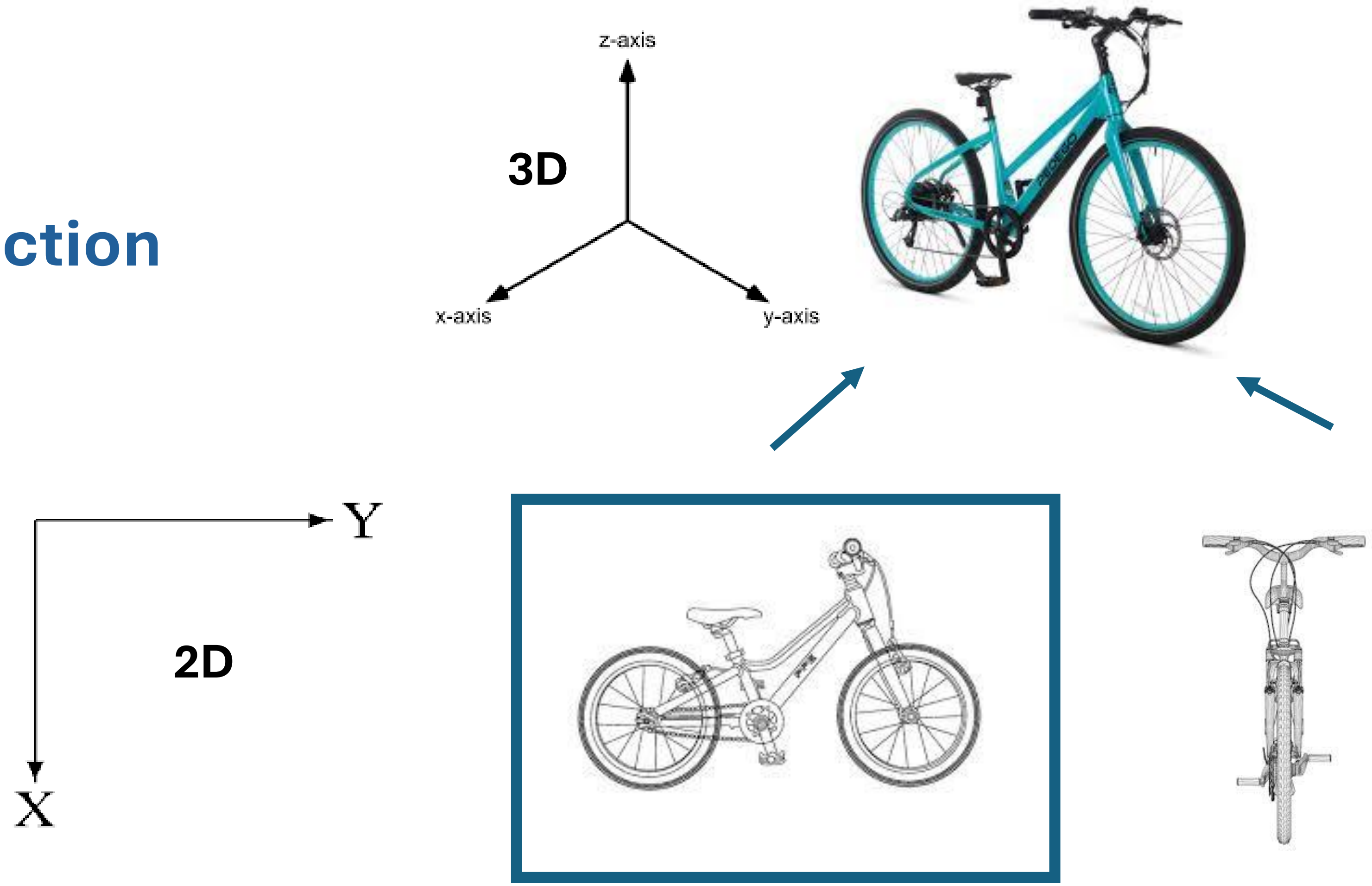
Unsupervised ML

- **Only X, no Y**
- **Clustering**
- Association
- Dimensionality reduction



Unsupervised ML

- **Only X, no Y**
- Clustering
- Association
- **Dimensionality reduction**



Reduced the dimensions of the data from **3D** to **2D** and still **preserving** most of the information

Dimensionality Reduction | PCA

10-Dimensional Dataset

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10

3-Dimensional Dataset

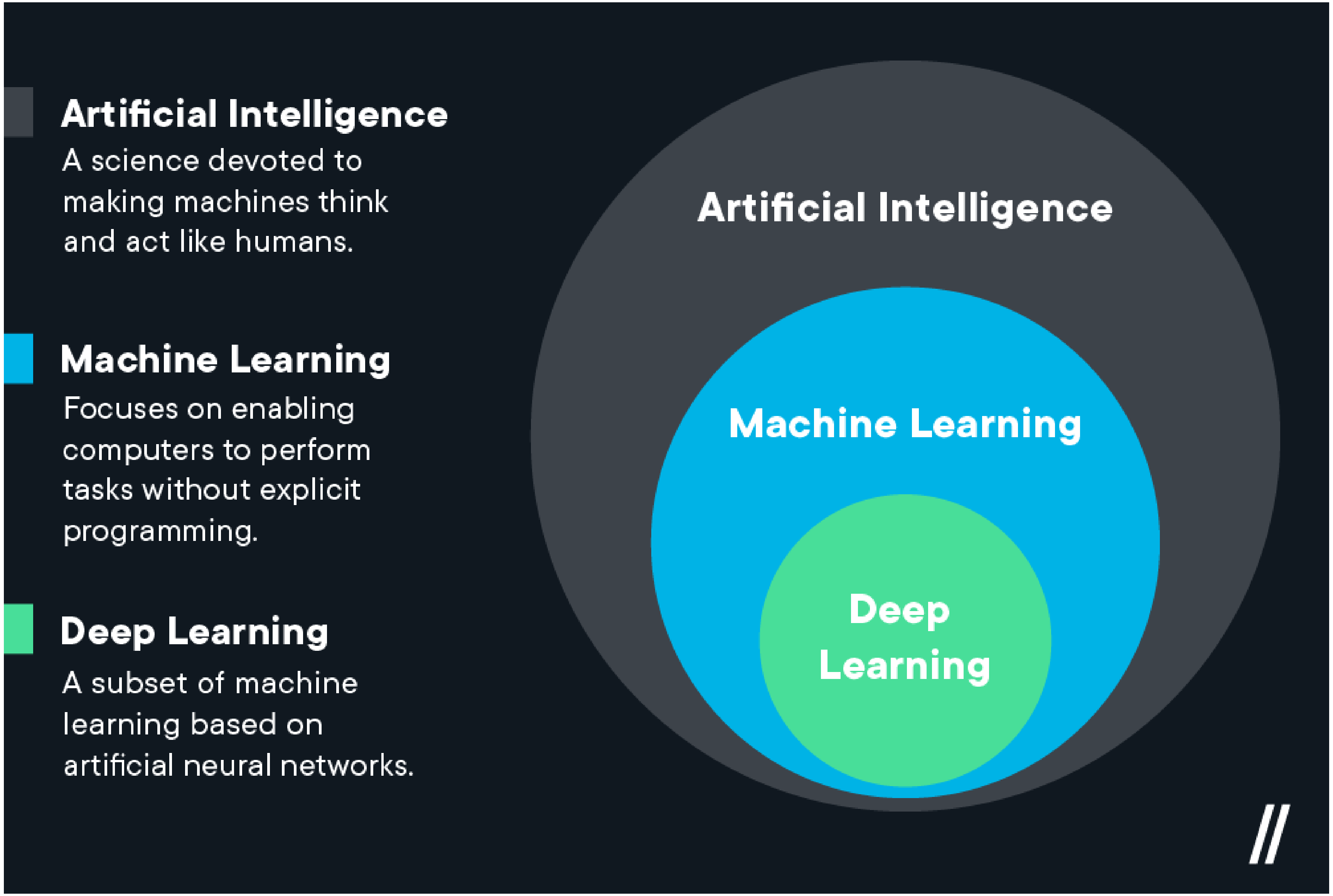
PCA1	PCA2	PCA3

Reinforcement Learning

ML | RL

- Self driving cars
- Robotics (Boston Dynamics)
- Gaming (AlphaGo)

ML | DL | AI



Artificial Intelligence
A science devoted to making machines think and act like humans.
Machine Learning
Focuses on enabling computers to perform tasks without explicit programming.
Deep Learning
A subset of machine learning based on artificial neural networks.
Artificial Intelligence
Machine Learning
Deep Learning

Two Approaches in ML

Model Centric Approach

Model Selection / Model Tuning

Data Centric Approach

Auto ML

When To Do ML

When To Do ML

When not to use ML:

- ML is not a solution for every type of problem, you don’t need ML if you can determine a target value by suing simple rules, computations, or predetermined steps that can be programmed.

Use machine learning for the following situations:

- *You cannot code the rules:*
  - Email Classification (Spam or not Spam) , cannot be adequately solved using a simple (deterministic), rule-based solution that depend on too many factors.
- *You cannot scale:*
  - You might be able to manually recognize a few hundred emails and decide whether they are spam or not. However, this task becomes tedious for millions of emails. ML solutions are effective at handling large-scale problems.

When To Do ML

Is there a problem you need to solve?

Just say the magic words

ML

Why Python

Why Python

- Use your model effectively
- You can use it for automation
- You can calculate your own calculations , programs ,applications.

Some Python Projects

- Production Data Automation Insights: A Python-Powered Journey from Outlook to PowerBI
- Capturing Uncertainty in Production Rates: Monte Carlo Simulations Using Python, OpenServer, and Prosper
- Relative Permeability Data Normalization and Denormalization APP

References

- Virtual Flow Metering (physics vs ML based)
- Lithology Prediction From Drilling Parameters
- Machine Learning for oil and gas (Book)