Machine Learning for Petroleum Professionals: A Beginner-Friendly, No-Code Introduction

PETRO-ANALYST

Ahmed Abd Elgawad

About ME | My Portfolio

Ahmed Elsayed Abdelgawad

Reservoir engineer, Data analyst, Python developer and ML developer.

Junior Reservoir Engineer with expertise in data analysis, Python development, and machine learning, focused on digitalization and digital transformation in the petroleum industry. I have successfully integrated programming and data science skills into various projects, enhancing my proficiency in petroleum engineering. Passionate about leveraging cutting-edge technologies to optimize production and automate processes, I aim to drive efficiency and innovation within the industry.

LinkedIn Github Youtube Kaggle Medium

Workshop Content

- Explain the Basics of ML (What is ML using Example)
- What is ML, DS life cycle, Supervised and Unsupervised ML
- Practice with no-code software (GUI)

- Go deep in any details, just high level knowledge (Just the basics)
- We will know explain any ML model and how it works (For you investigation)

Session Flow

Data Science Life Cycle
Just an Introduction

What is ML
Just an Introduction

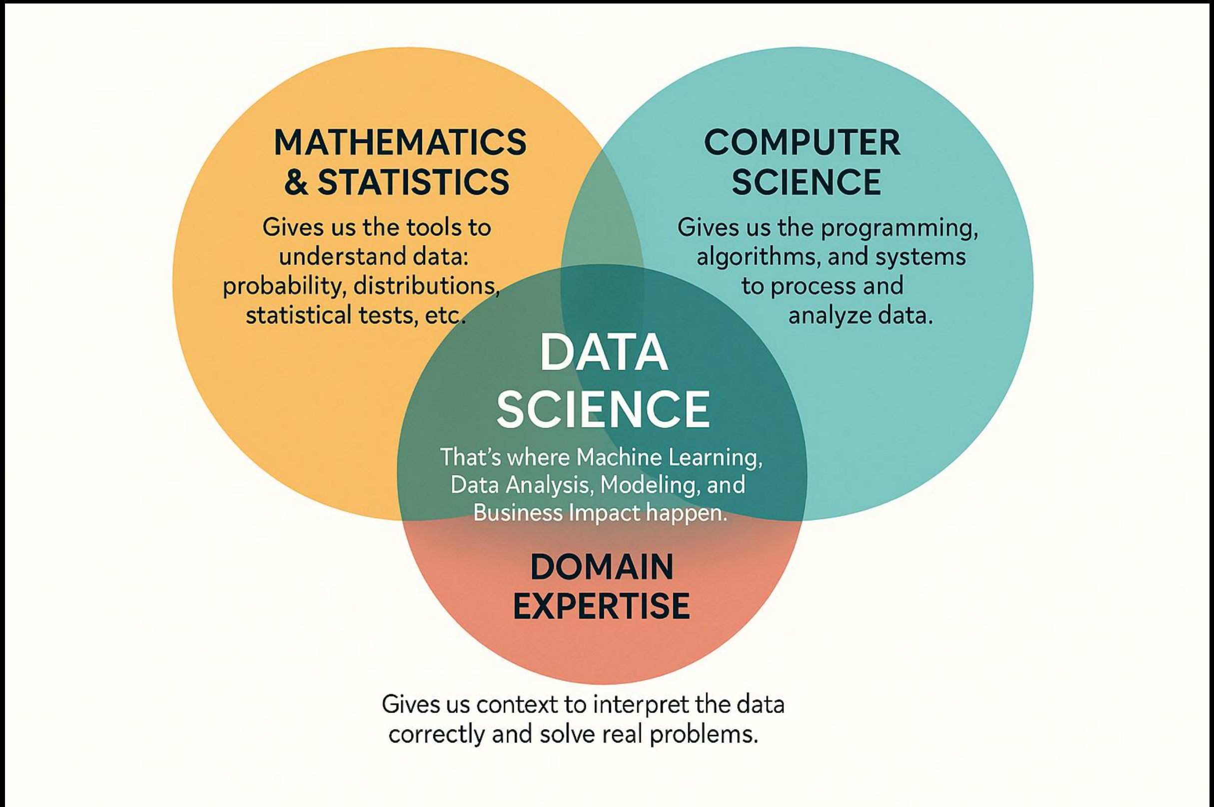
ML Workflow

When and When not to do ML

Data Science Life Cycle

Just an Introduction

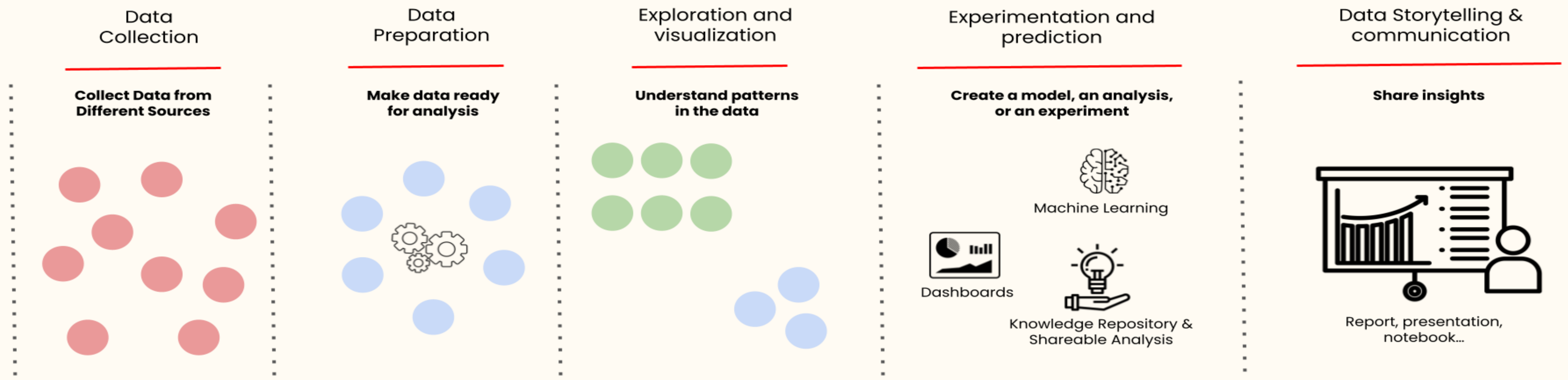
What is DS



Mathematics + Computer Science + Domain Expertise ⋙ DS

DS Life Cycle

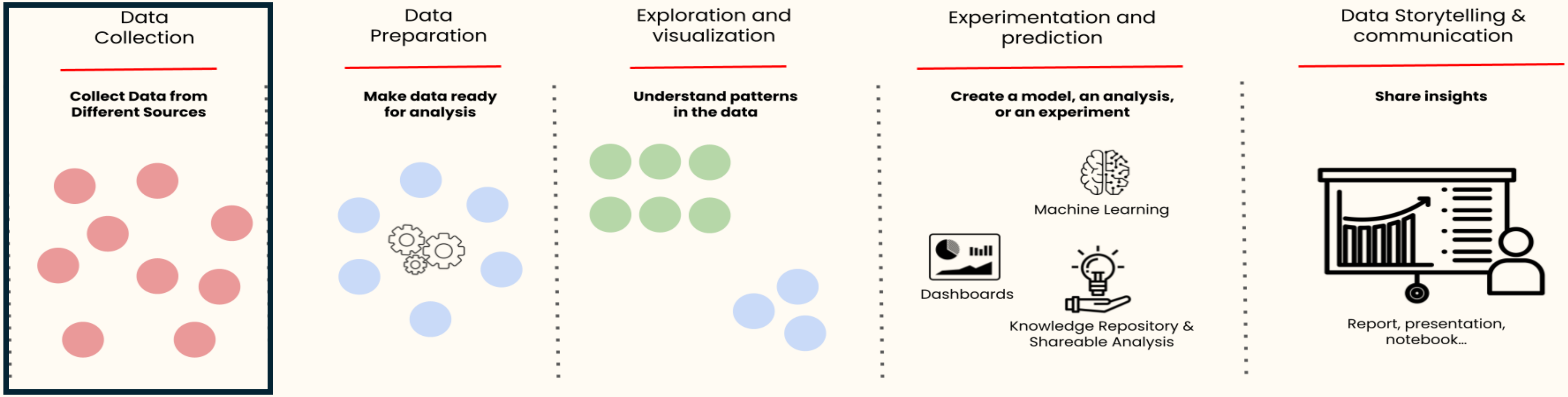
A problem that you need to solve, Using the Data!



DS Life Cycle

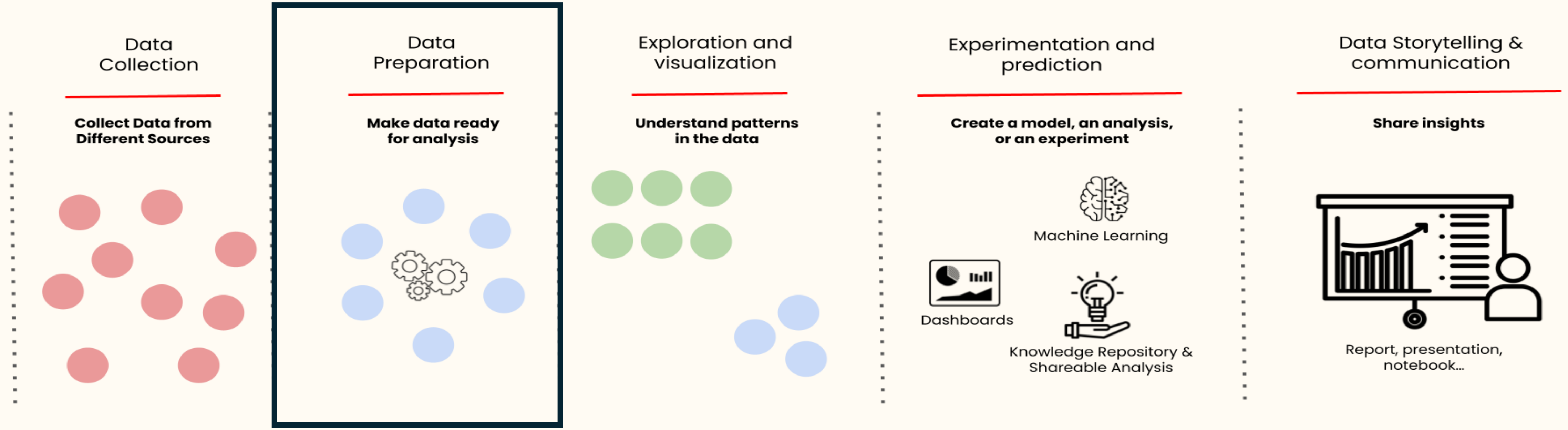
Unstructured Data

Excel File, Tables .. → Structured / Tabular Data



DS Life Cycle

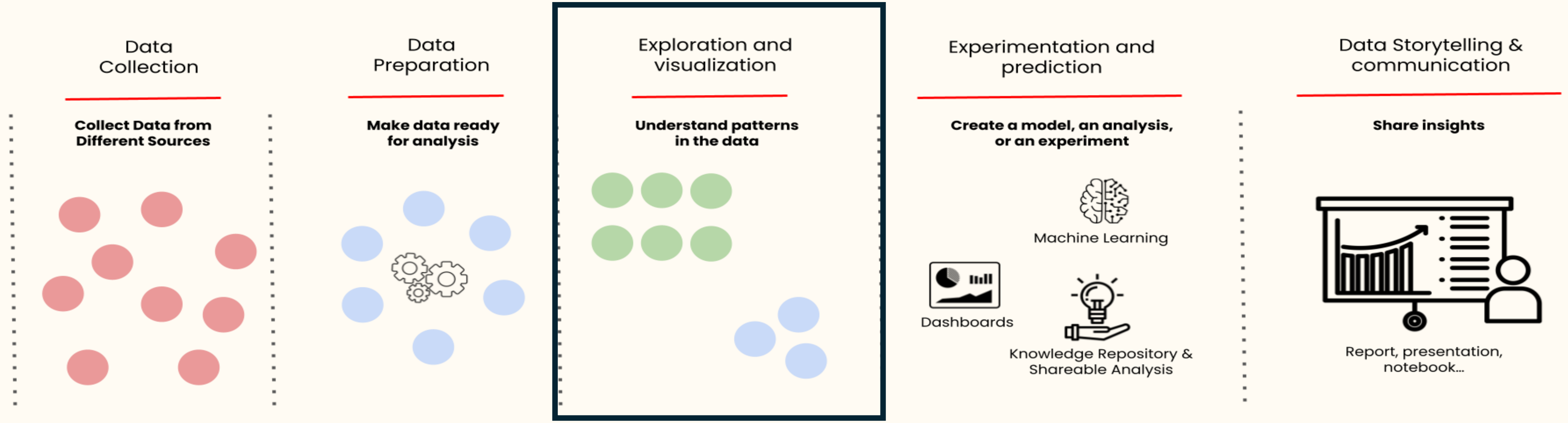
Create a clean, high quality dataset that yield accurate and reliable analytical results



Garbage In -> Model -> Garbage Out

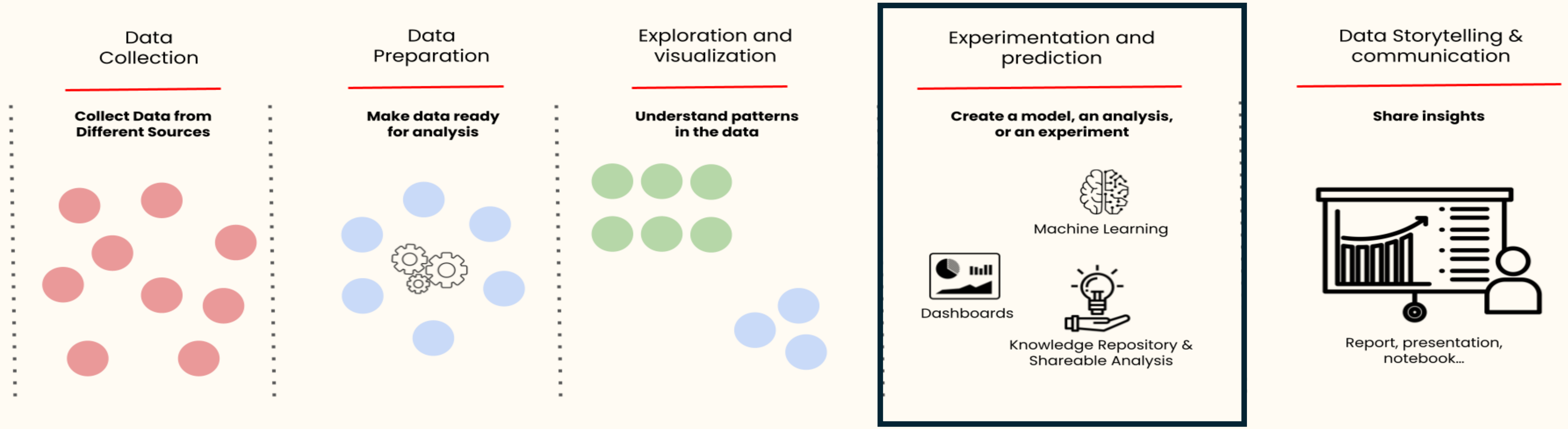
DS Life Cycle

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.

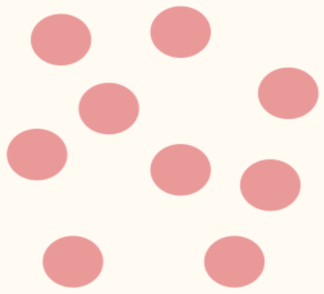


DS Life Cycle

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

Data Collection

Collect Data from Different Sources



Data Preparation

Make data ready for analysis

Exploration and visualization

Understand patterns in the data

Experimentation and prediction

Create a model, an analysis, or an experiment

Machine Learning

Dashboards

Knowledge Repository & Shareable Analysis

Data Storytelling & communication

Share insights

Report, presentation, notebook...

This step has no value without the previous steps

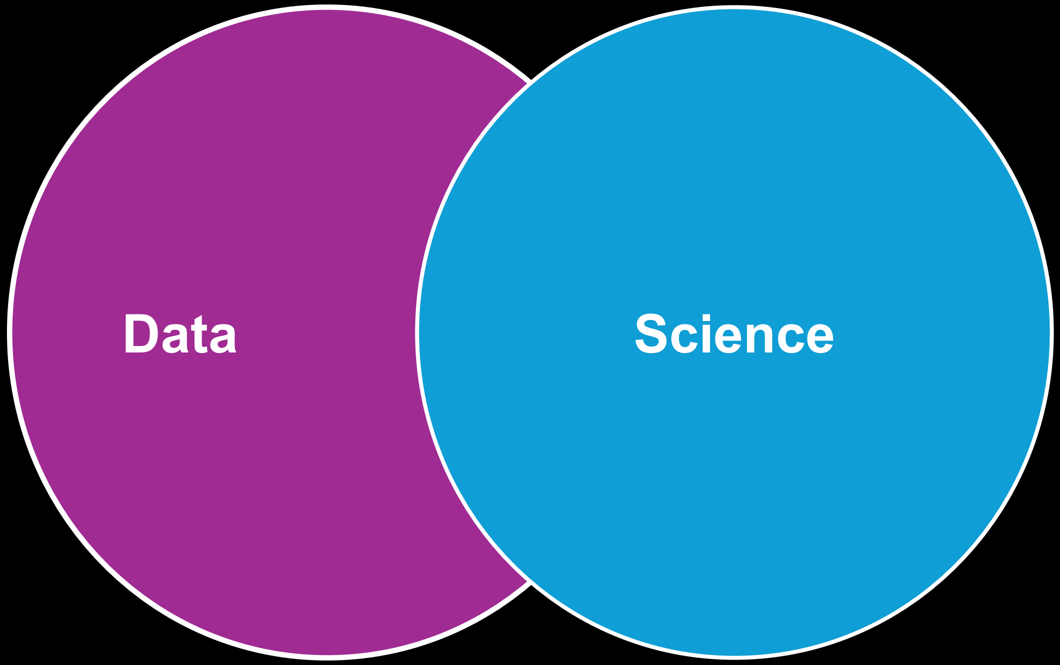
Data Preparation

Check for anomalies

Missing Data

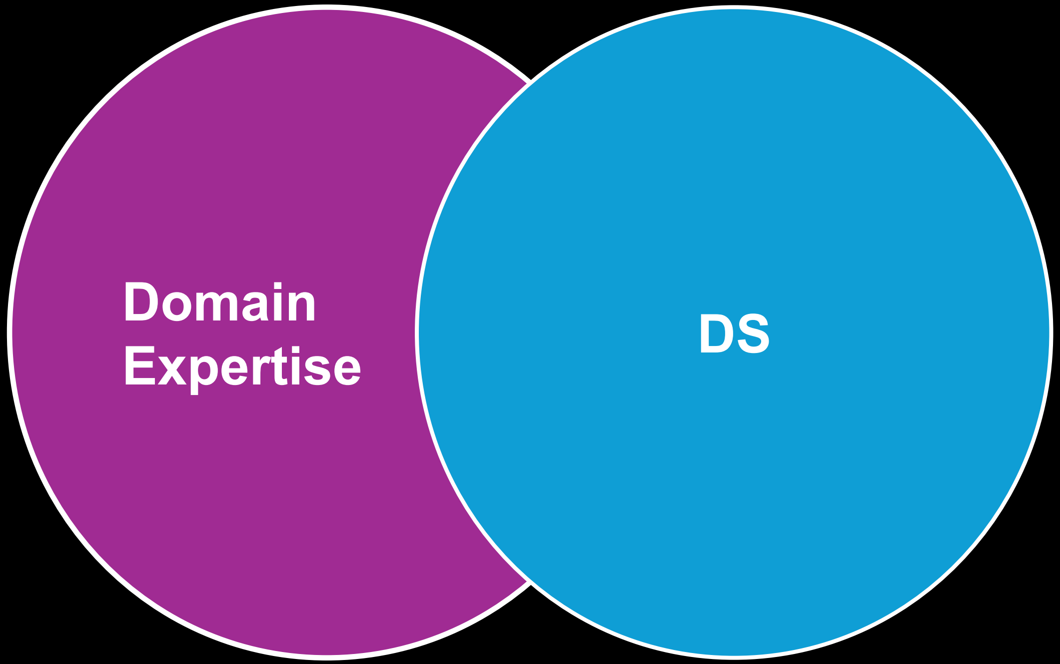
Typos

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**.



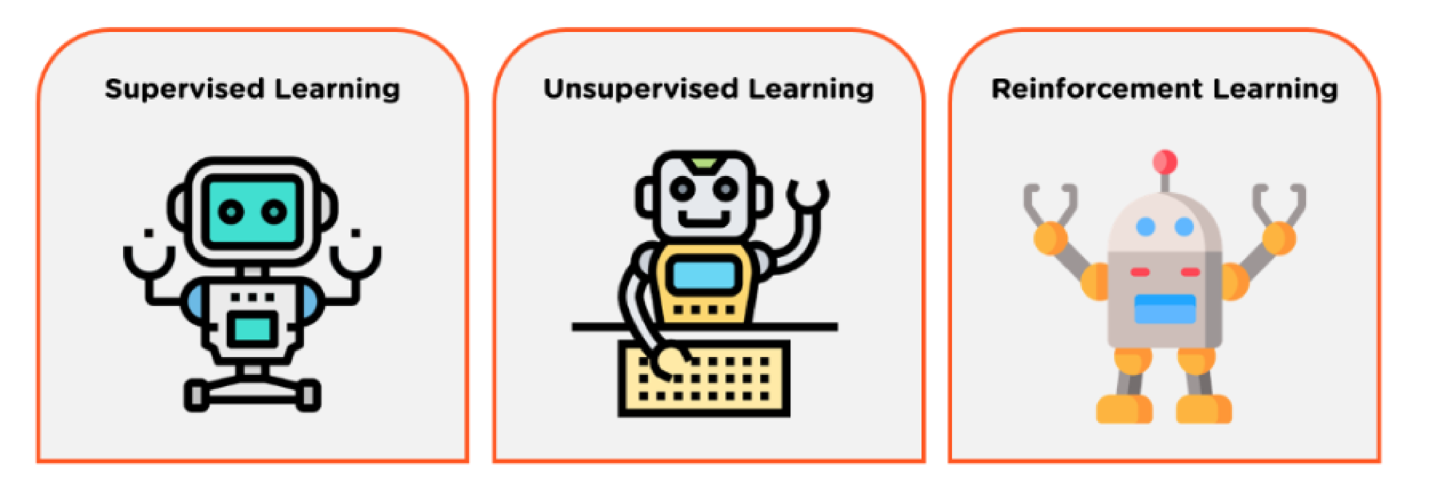
What is ML

Just an Introduction

PETRO-ANALYST

Ahmed Abd Elgawad

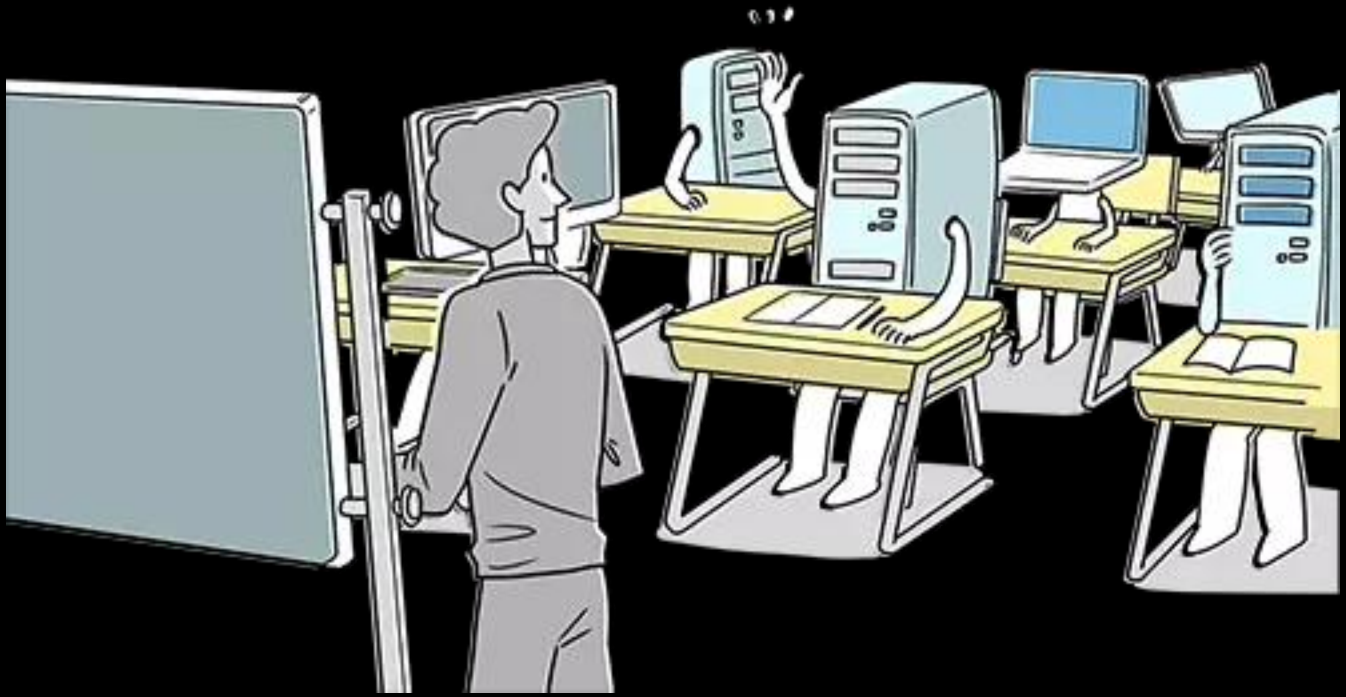
Types of ML



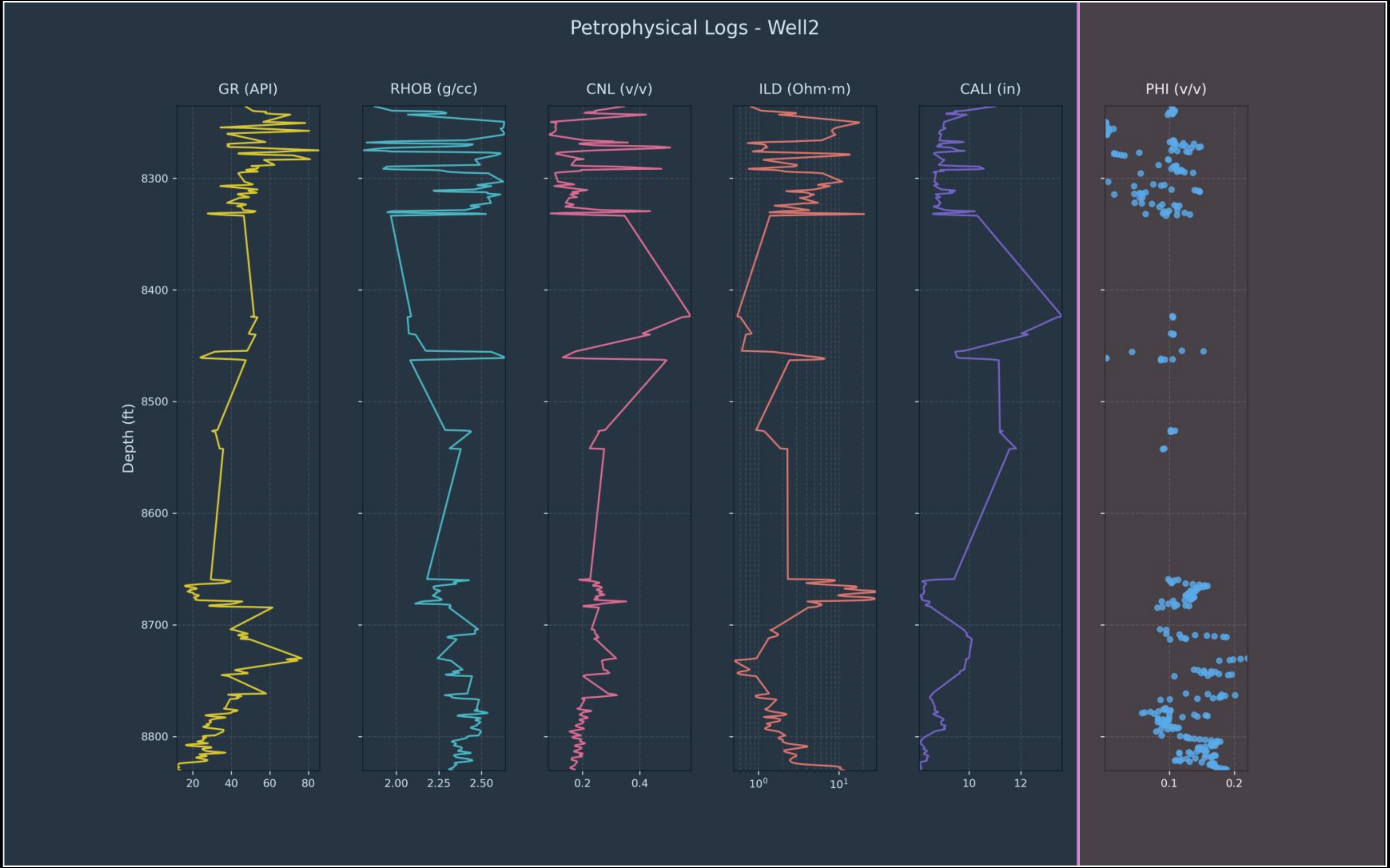
Supervised Learning
Unsupervised Learning
Reinforcement Learning

Supervised ML

Supervised ML



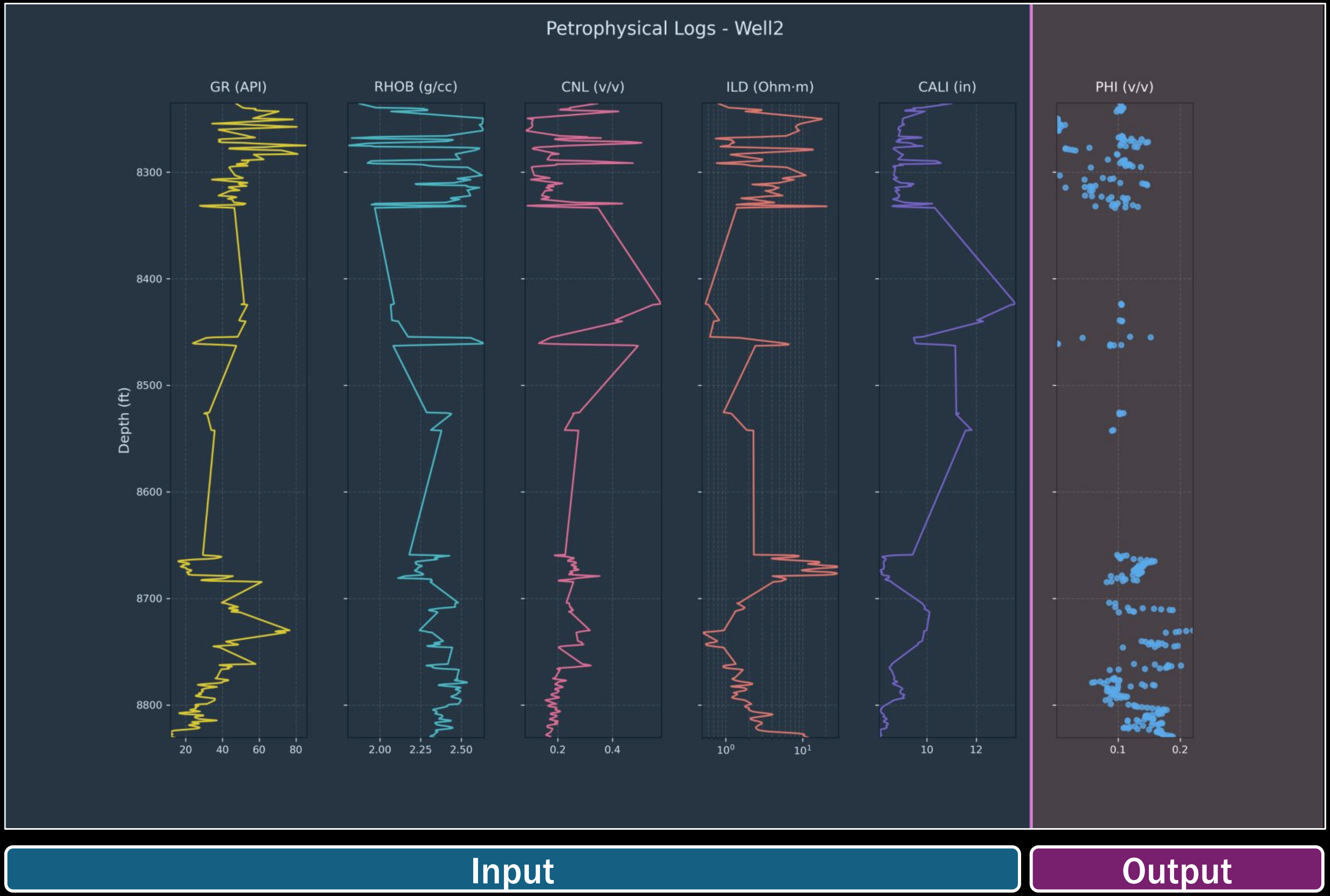
Porosity Calculation



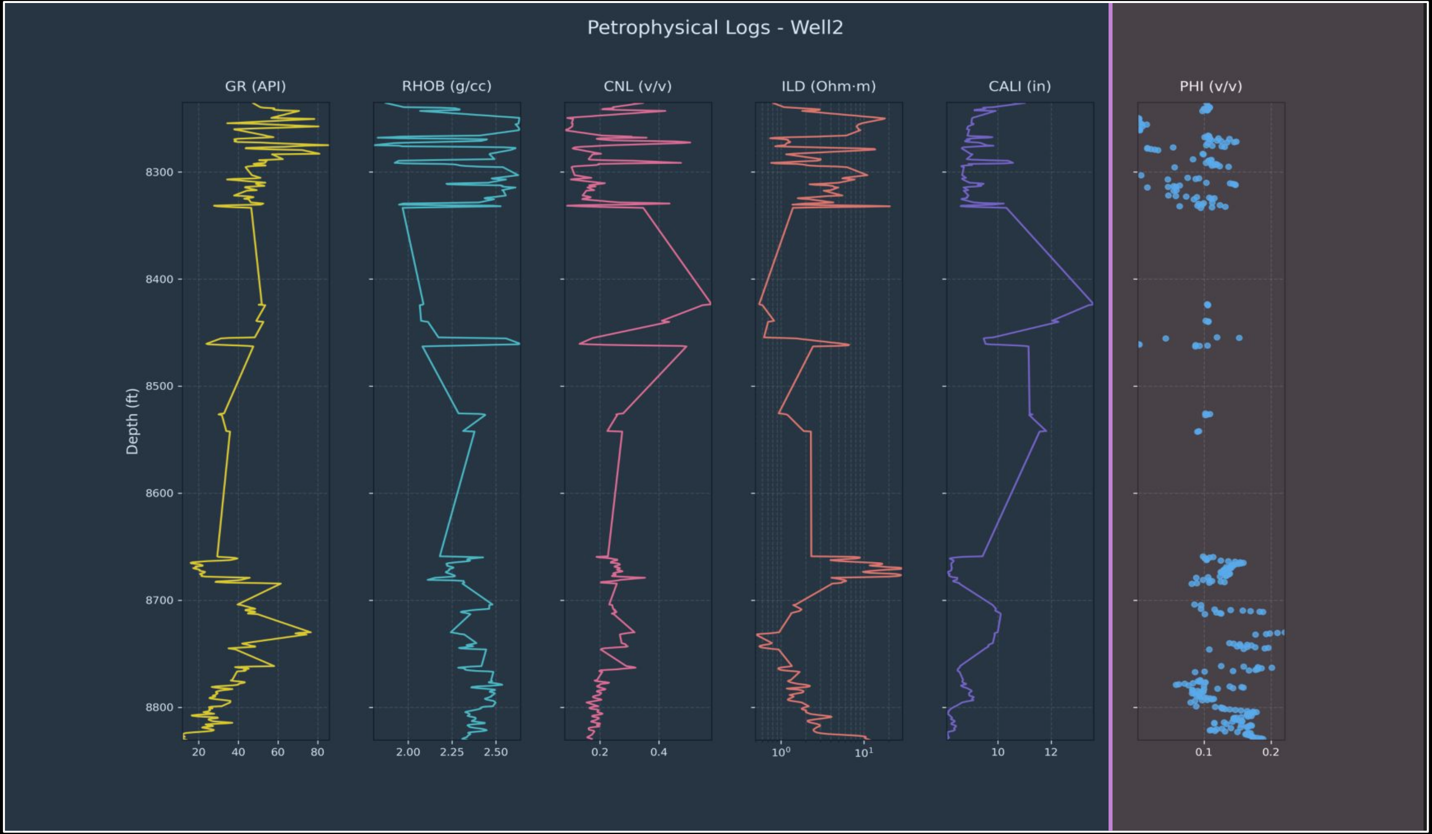
Input

Output

Porosity Calculation



Porosity Calculation

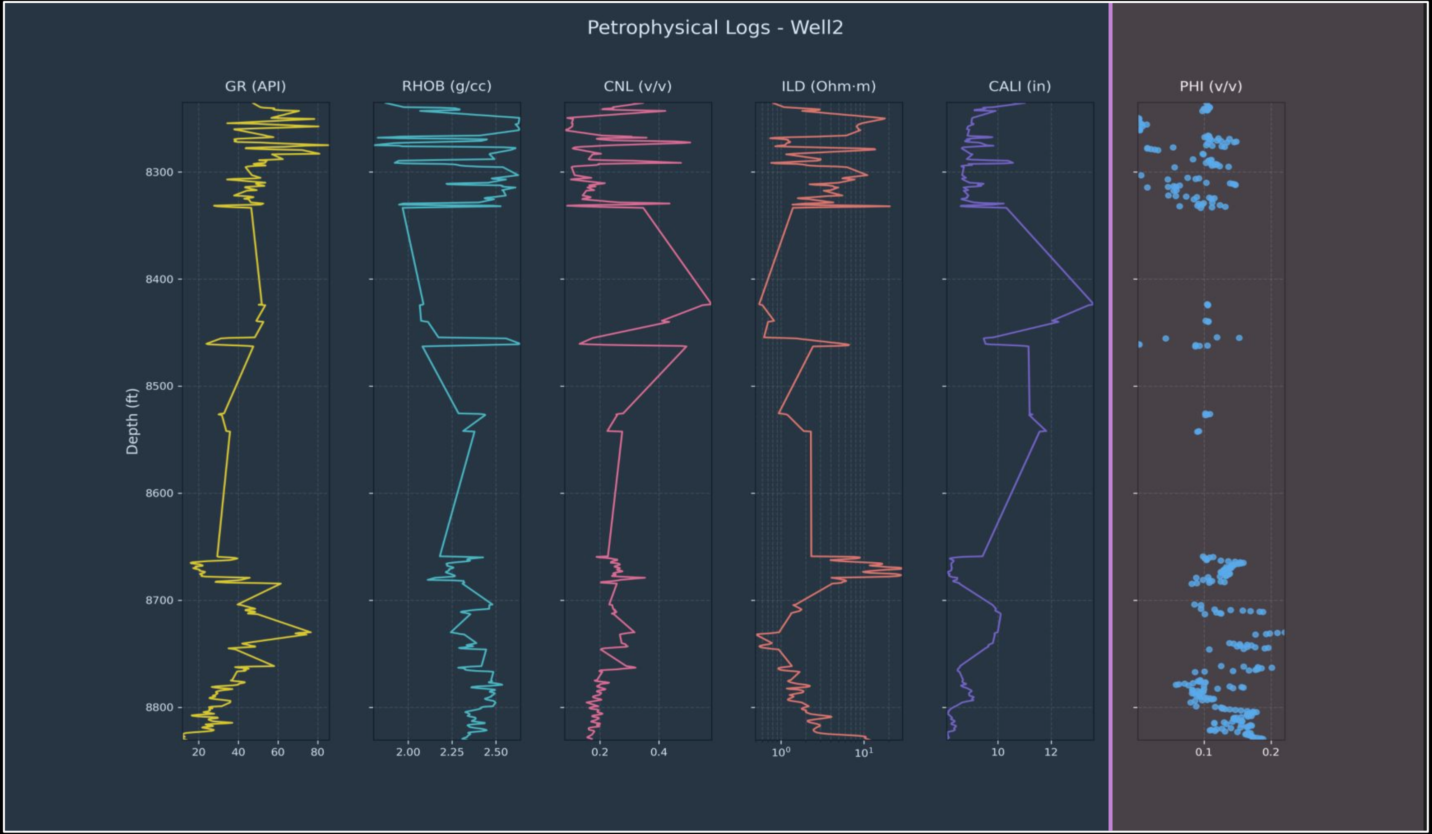


Input | Output

$$\phi = \frac{\rho_{\text{matrix}} - \rho_{\text{bulk}}}{\rho_{\text{matrix}} - \rho_{\text{fluid}}}$$

Input + Rule ⟫ Output

Porosity Calculation



Input

Output

$$\phi = \frac{\rho_{\text{matrix}} - \rho_{\text{bulk}}}{\rho_{\text{matrix}} - \rho_{\text{fluid}}}$$

Input + Output ⋙ Rule

Machine Learning Vs Traditional Programming

Traditional Programming

Input + Rule | Program ❯❯❯ Output

Using Machine learning

Input + Output ❯❯❯ Rule | Program

Porosity Calculation (ML)

RESERVOIR SOLUTIONS
CREATED TO OPTIMIZE

PETRO-ANALYST

DEPTH	ILD	RHOB	GR	CNL	CALIPER	Core_Phi
8235.00	0.81	1.89	48.03	0.31	10.56	0.10
8235.50	0.80	1.87	47.33	0.35	11.01	0.10
8239.50	1.10	1.98	51.09	0.24	9.81	0.11
8240.00	1.40	2.18	52.64	0.25	9.45	0.11
8240.50	1.78	2.27	58.07	0.25	9.47	0.10
8241.00	2.27	2.28	58.19	0.21	9.52	0.10
8241.50	2.93	2.29	57.66	0.21	9.33	0.10
8242.00	2.93	2.29	58.76	0.26	9.12	0.11
8242.50	2.55	2.19	62.53	0.34	9.12	0.10
8243.00	1.81	2.07	70.68	0.42	9.91	0.10
8266.00	6.14	2.41	53.91	0.21	8.86	0.11
8266.50	4.01	2.17	54.62	0.26	8.96	0.11
8267.00	2.71	1.95	56.42	0.31	9.61	0.10
8267.50	1.39	1.90	57.75	0.30	9.79	0.10
8268.00	0.85	1.83	52.65	0.36	9.63	0.11
8268.50	0.74	1.95	46.16	0.20	9.11	0.12
8269.00	0.93	2.28	39.81	0.19	8.91	0.14

Input + Output ››› Rule

Supervised | How

X → Model → Y

Label

Supervision

DEPTH

ILD

RHOB

GR

CNL

CALIPER

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

Core Porosity

Features

Target

Supervised | How

Y

Regression

Y -> Continuous Variable
- Pressure , Rates , Prices

Classification

Y -> Discrete Variable
- Lithology / Facies , Classes

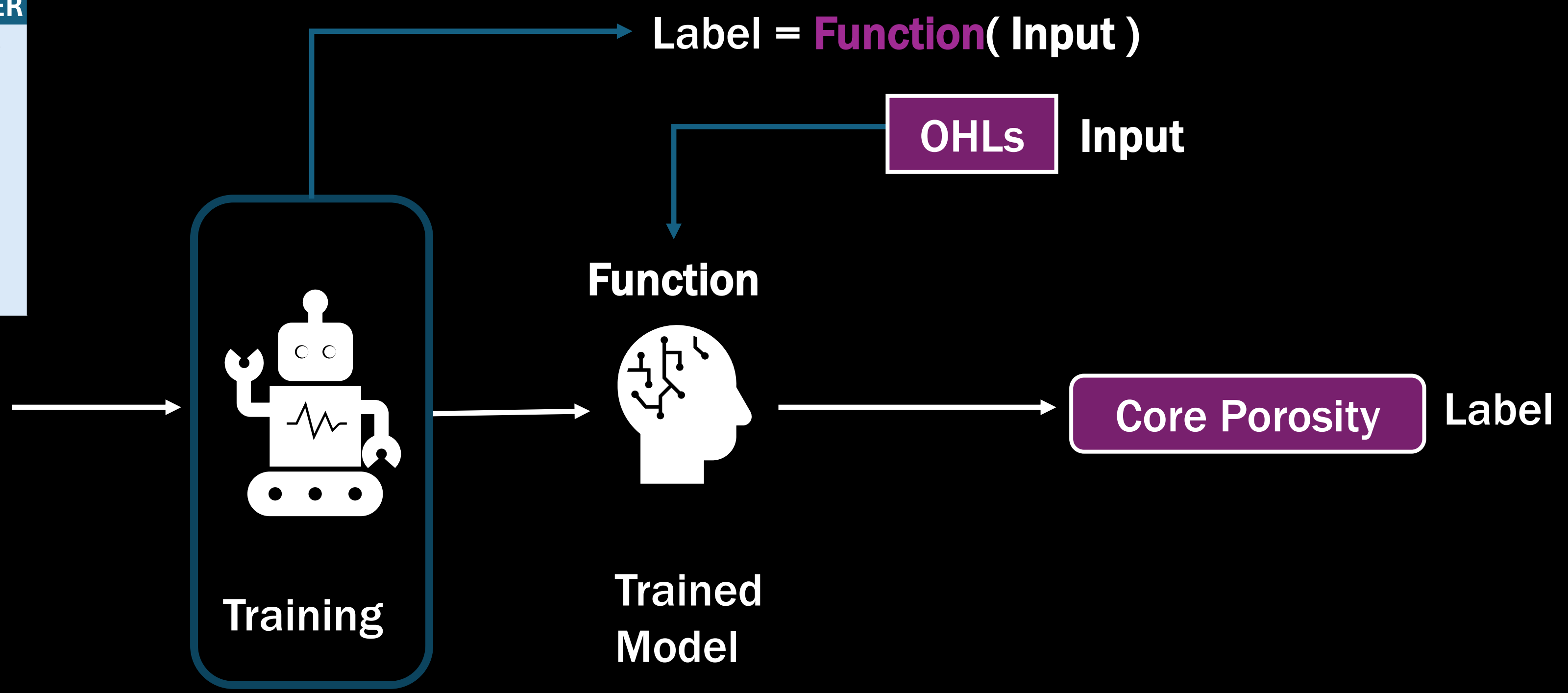
Model Training and Validation

X

DEPTH	ILD	RHOB	GR	CNL	CALIPER
8235.00	0.81	1.89	48.03	0.31	10.56
8235.50	0.80	1.87	47.33	0.35	11.01
8239.50	1.10	1.98	51.09	0.24	9.81
8240.00	1.40	2.18	52.64	0.25	9.45
8240.50	1.78	2.27	58.07	0.25	9.47
8241.00	2.27	2.28	58.19	0.21	9.52
8241.50	2.93	2.29	57.66	0.21	9.33
8242.00	2.93	2.29	58.76	0.26	9.12

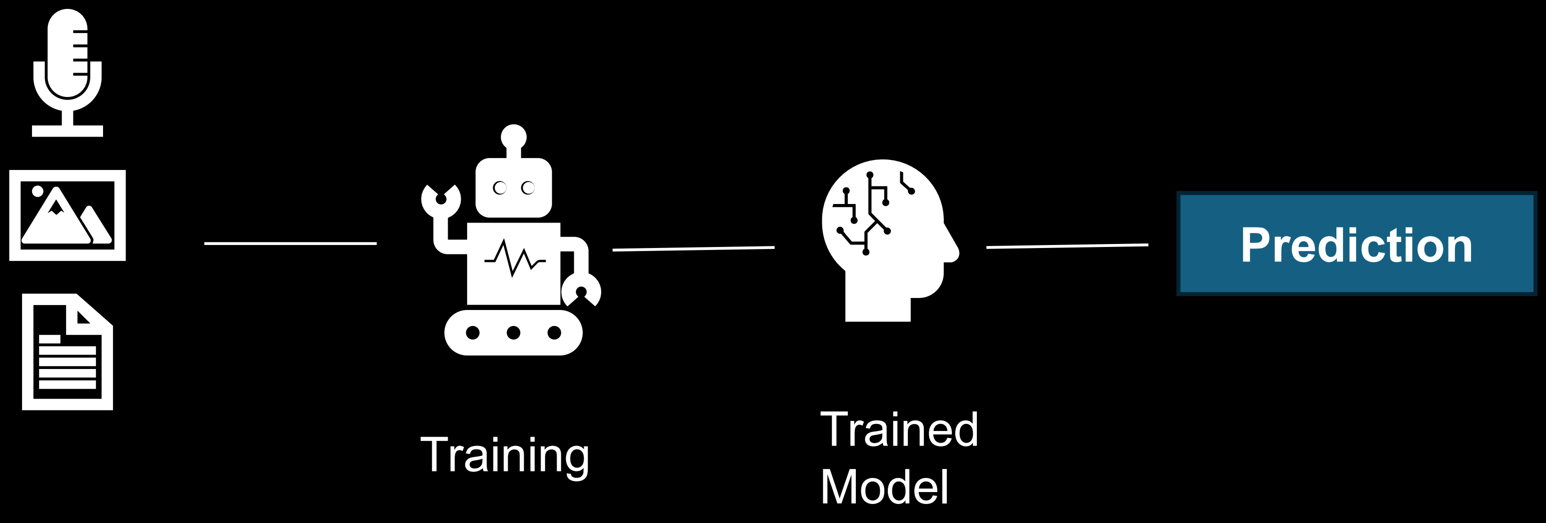
Y

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11



ML Workflow

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)

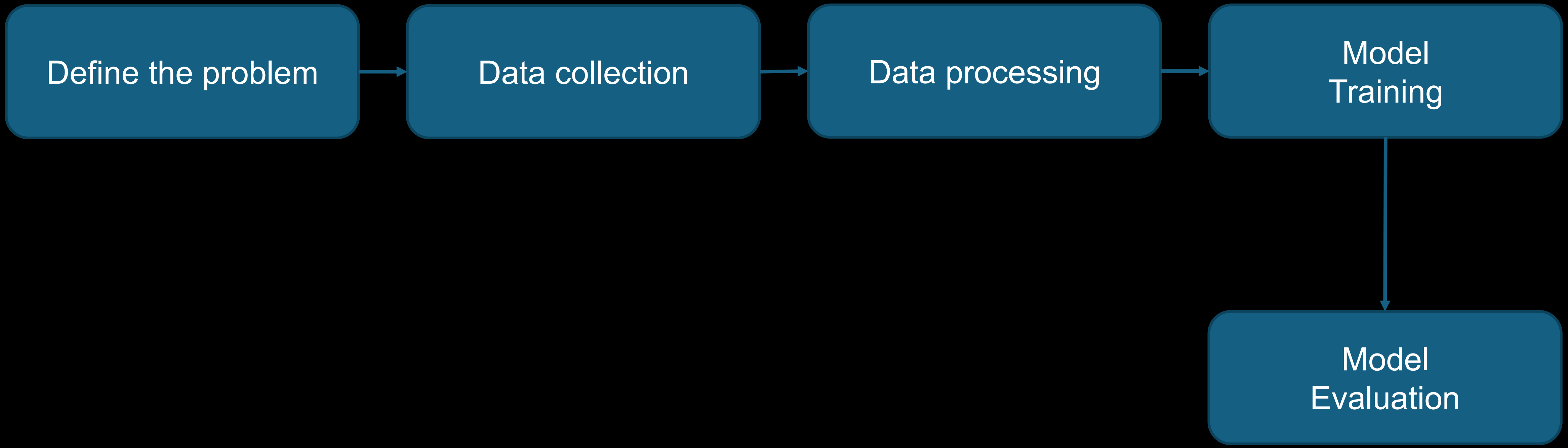
Unstructured data

Images

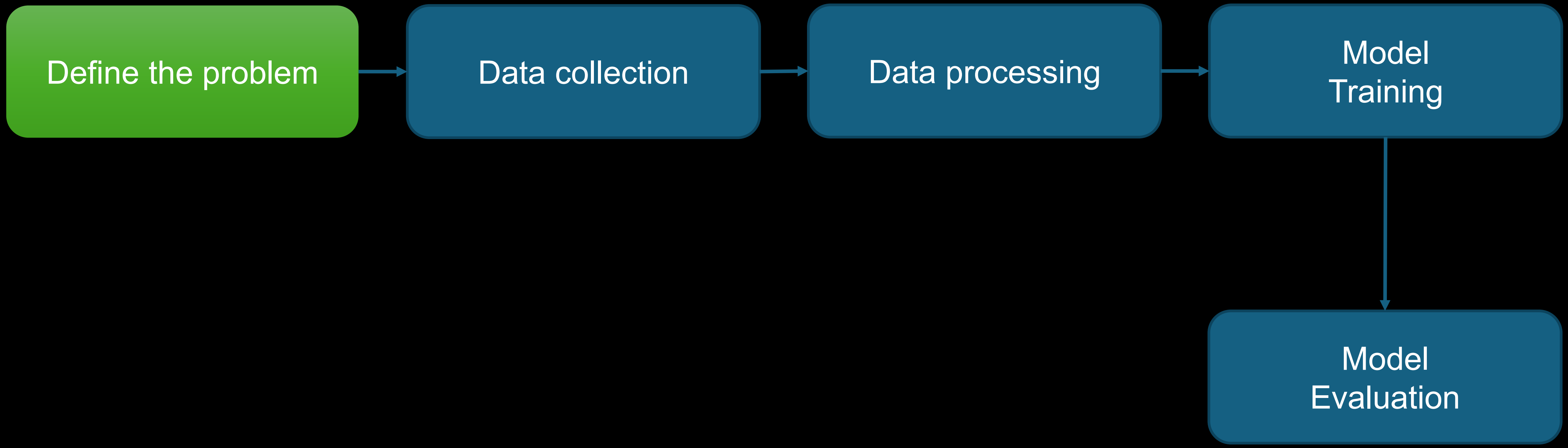
Audio

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

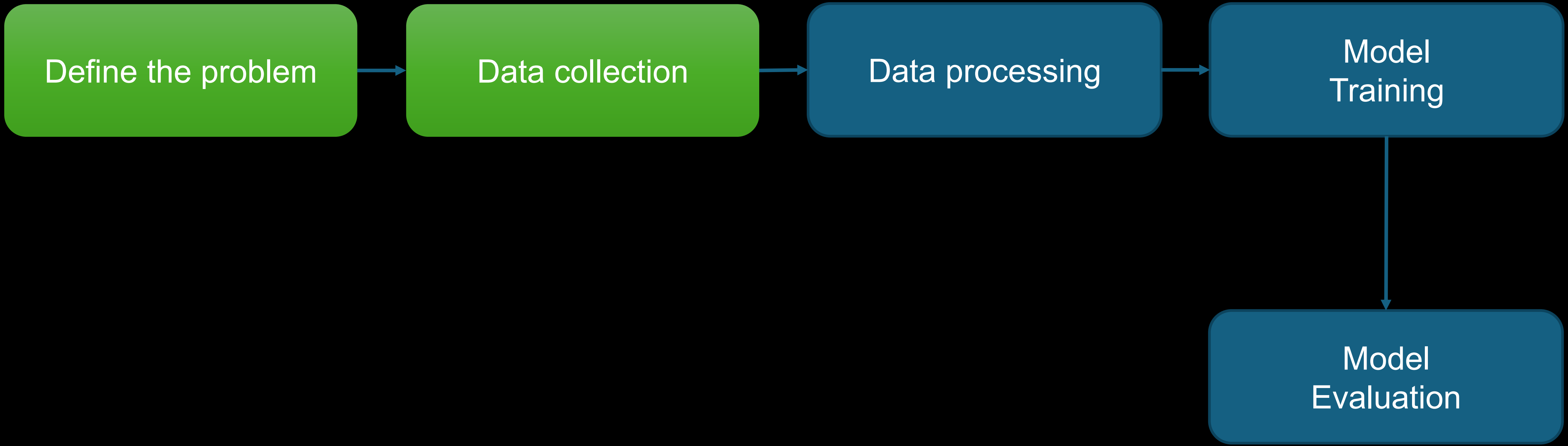
other

Baseline:

Beat the baseline

Successful project

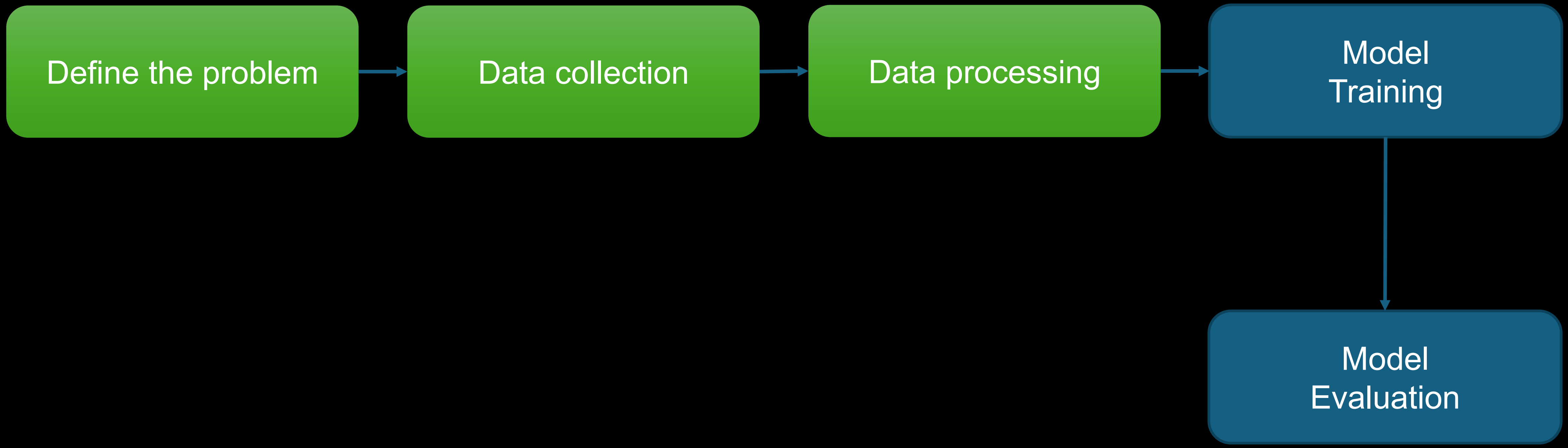
ML Steps


Define the problem
Data collection
Data processing
Model Training
Model Evaluation

Data Collection

DEPTH	ILD	RHOB	GR	CNL	CALIPER	Core_Phi
8235.00	0.81	1.89	48.03	0.31	10.56	0.10
8235.50	0.80	1.87	47.33	0.35	11.01	0.10
8239.50	1.10	1.98	51.09	0.24	9.81	0.11
8240.00	1.40	2.18	52.64	0.25	9.45	0.11
8240.50	1.78	2.27	58.07	0.25	9.47	0.10
8241.00	2.27	2.28	58.19	0.21	9.52	0.10
8241.50	2.93	2.29	57.66	0.21	9.33	0.10
8242.00	2.93	2.29	58.76	0.26	9.12	0.11
8242.50	2.55	2.19	62.53	0.34	9.12	0.10
8243.00	1.81	2.07	70.68	0.42	9.91	0.10
8266.00	6.14	2.41	53.91	0.21	8.86	0.11
8266.50	4.01	2.17	54.62	0.26	8.96	0.11
8267.00	2.71	1.95	56.42	0.31	9.61	0.10
8267.50	1.39	1.90	57.75	0.30	9.79	0.10
8268.00	0.85	1.83	52.65	0.36	9.63	0.11
8268.50	0.74	1.95	46.16	0.20	9.11	0.12
8269.00	0.93	2.28	39.81	0.19	8.91	0.14

ML Steps


Define the problem
Data collection
Data processing
Model Training
Model Evaluation

Data Processing

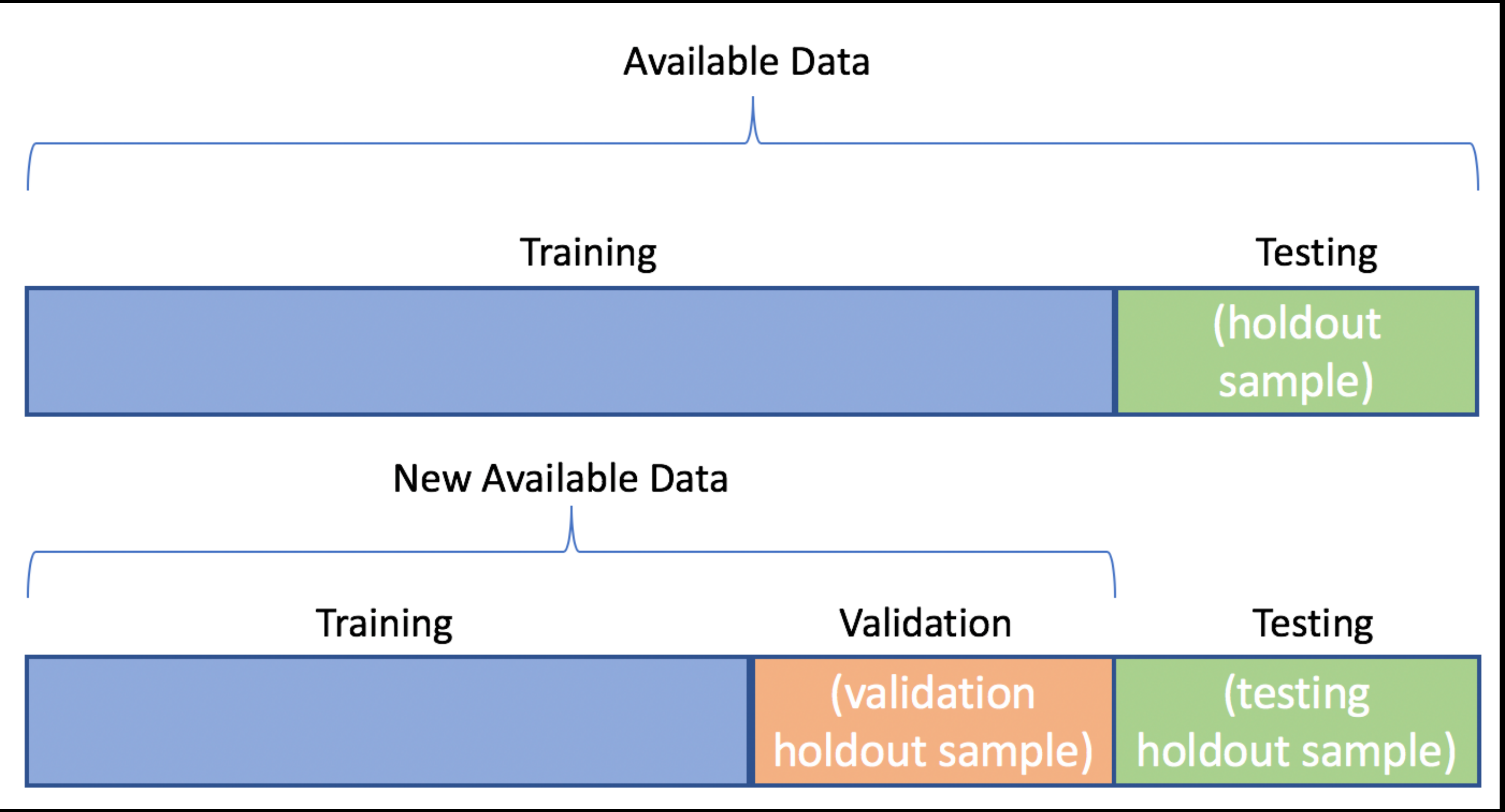
- Dealing with:
  - Missing data
  - Irrelevant data
  - Outliers

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

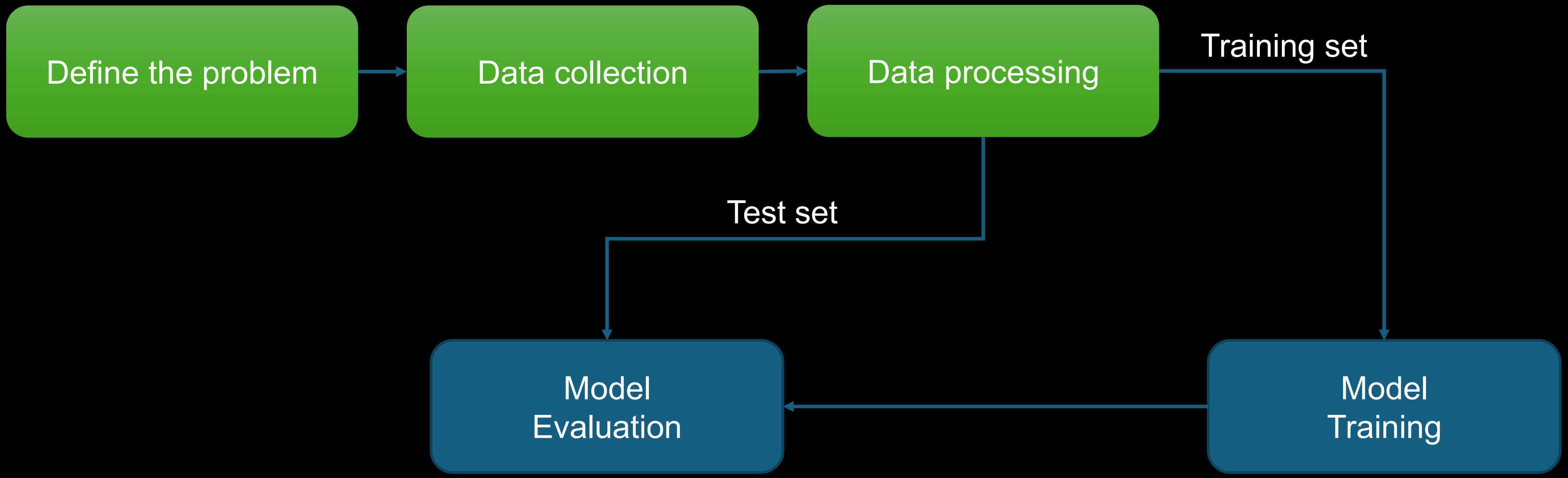
- Data Transformation:
  - normalizing the data
  - converting categorical data into numerical data
  - reducing the dimensionality of the data.

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

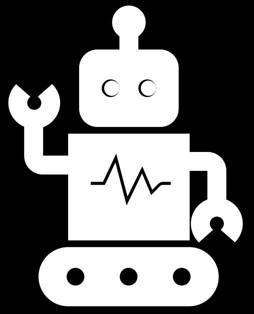
Training set

X

DEPTH	ILD	RHOB	GR	CNL	CALIPER
8235.00	0.81	1.89	48.03	0.31	10.56
8235.50	0.80	1.87	47.33	0.35	11.01
8239.50	1.10	1.98	51.09	0.24	9.81
8240.00	1.40	2.18	52.64	0.25	9.45
8240.50	1.78	2.27	58.07	0.25	9.47
8241.00	2.27	2.28	58.19	0.21	9.52
8241.50	2.93	2.29	57.66	0.21	9.33
8242.00	2.93	2.29	58.76	0.26	9.12

Y

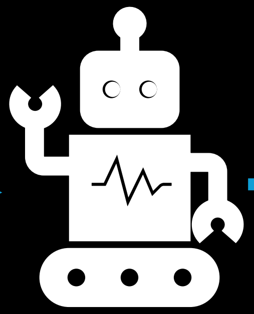
Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11



Test set

X

DEPTH	ILD	RHOB	GR	CNL	CALIPER
8235.00	0.81	1.89	48.03	0.31	10.56
8235.50	0.80	1.87	47.33	0.35	11.01
8239.50	1.10	1.98	51.09	0.24	9.81
8240.00	1.40	2.18	52.64	0.25	9.45
8240.50	1.78	2.27	58.07	0.25	9.47
8241.00	2.27	2.28	58.19	0.21	9.52
8241.50	2.93	2.29	57.66	0.21	9.33
8242.00	2.93	2.29	58.76	0.26	9.12



Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Model Output

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Actual values

Model Evaluation

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Model Output

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Actual values

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

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

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

Model Evaluation

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Model Output

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Actual values

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

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

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

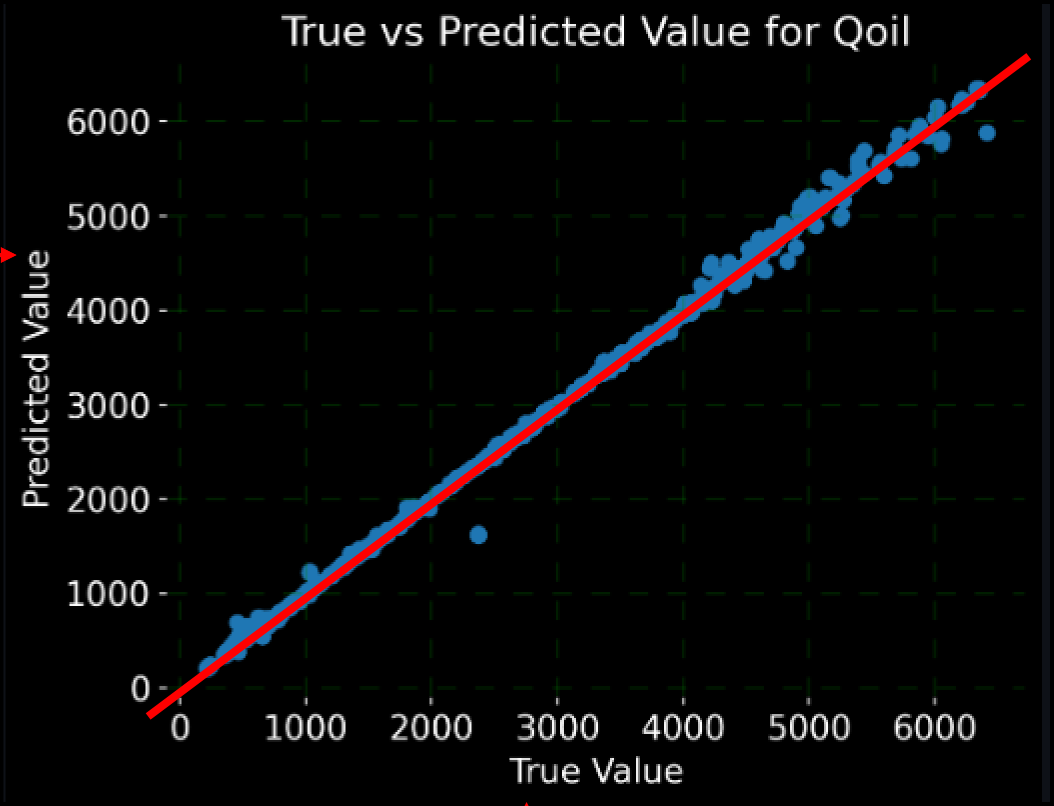
Model Evaluation

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

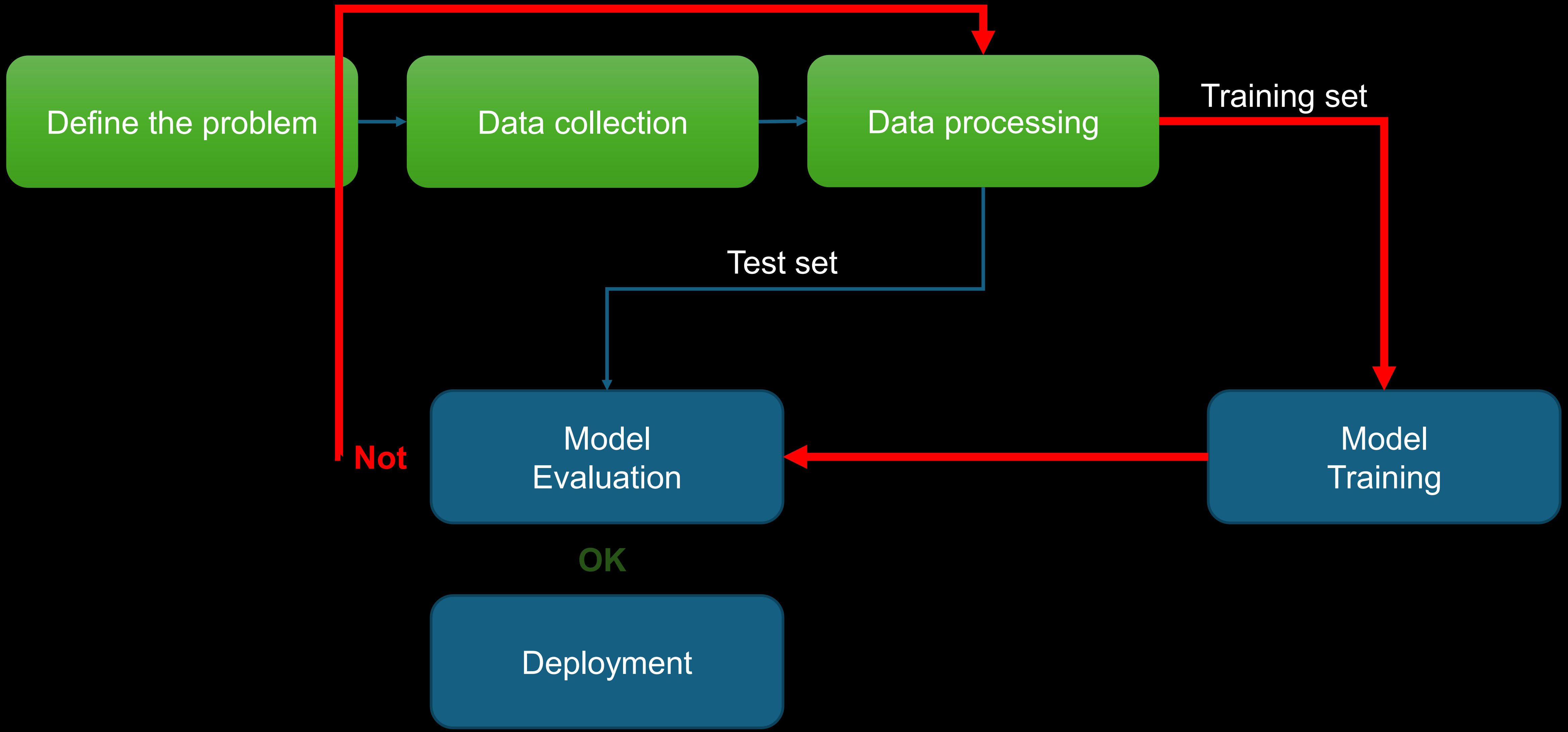
Model Output

Core_Phi
0.10
0.10
0.11
0.11
0.10
0.10
0.10
0.11

Actual values



ML Steps


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

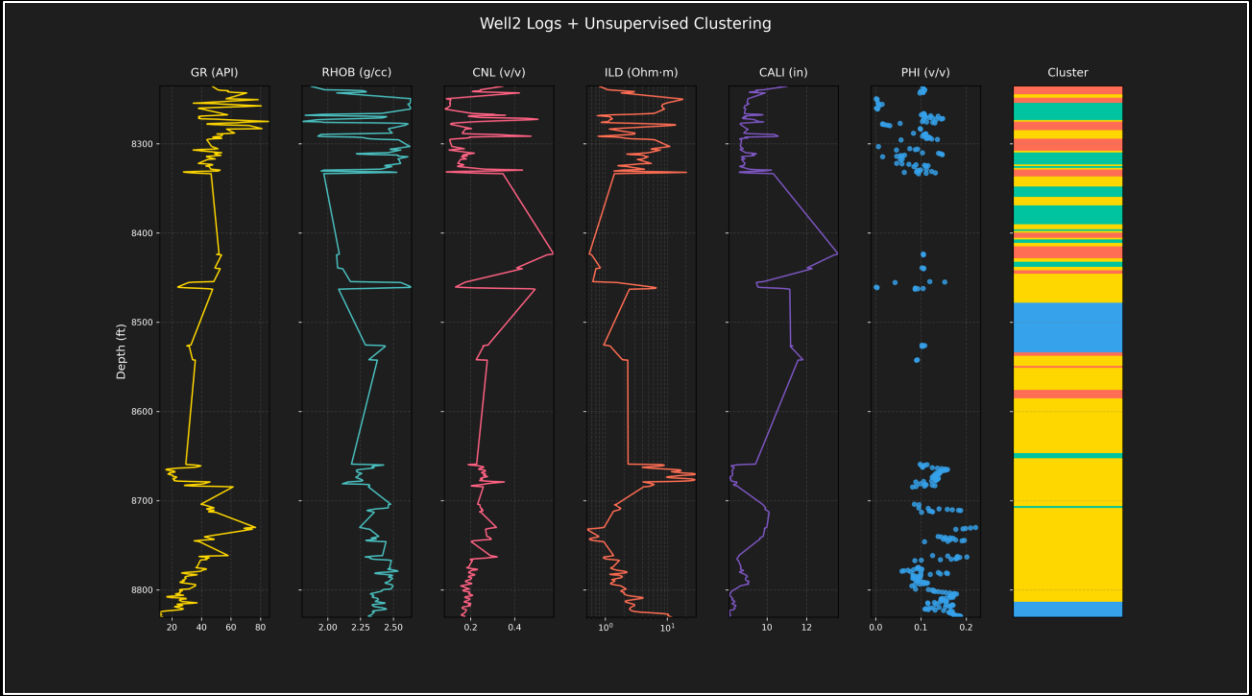
Unsupervised ML

PETRO-ANALYST

Ahmed Abd Elgawad

Unsupervised ML

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

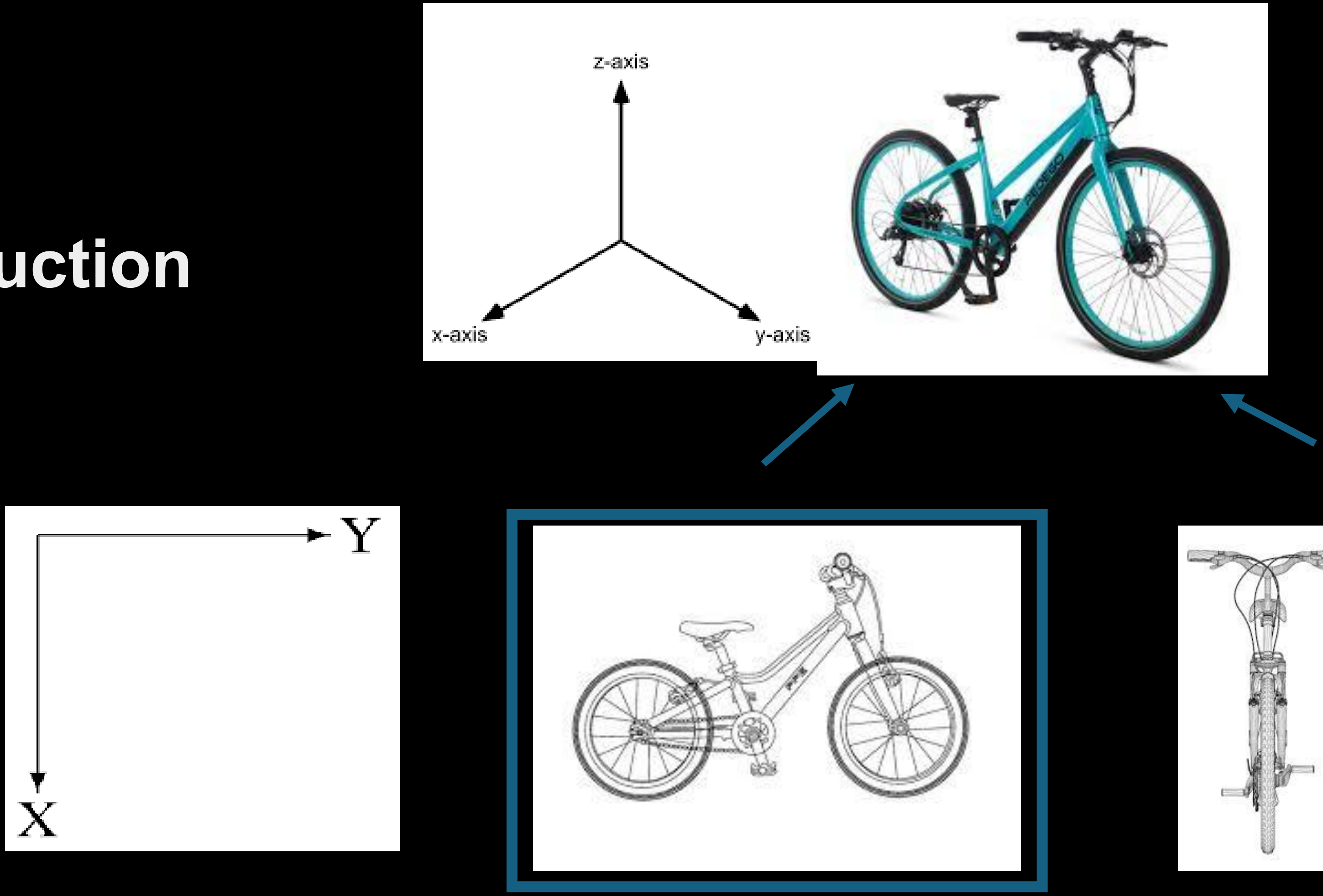


Unsupervised ML

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

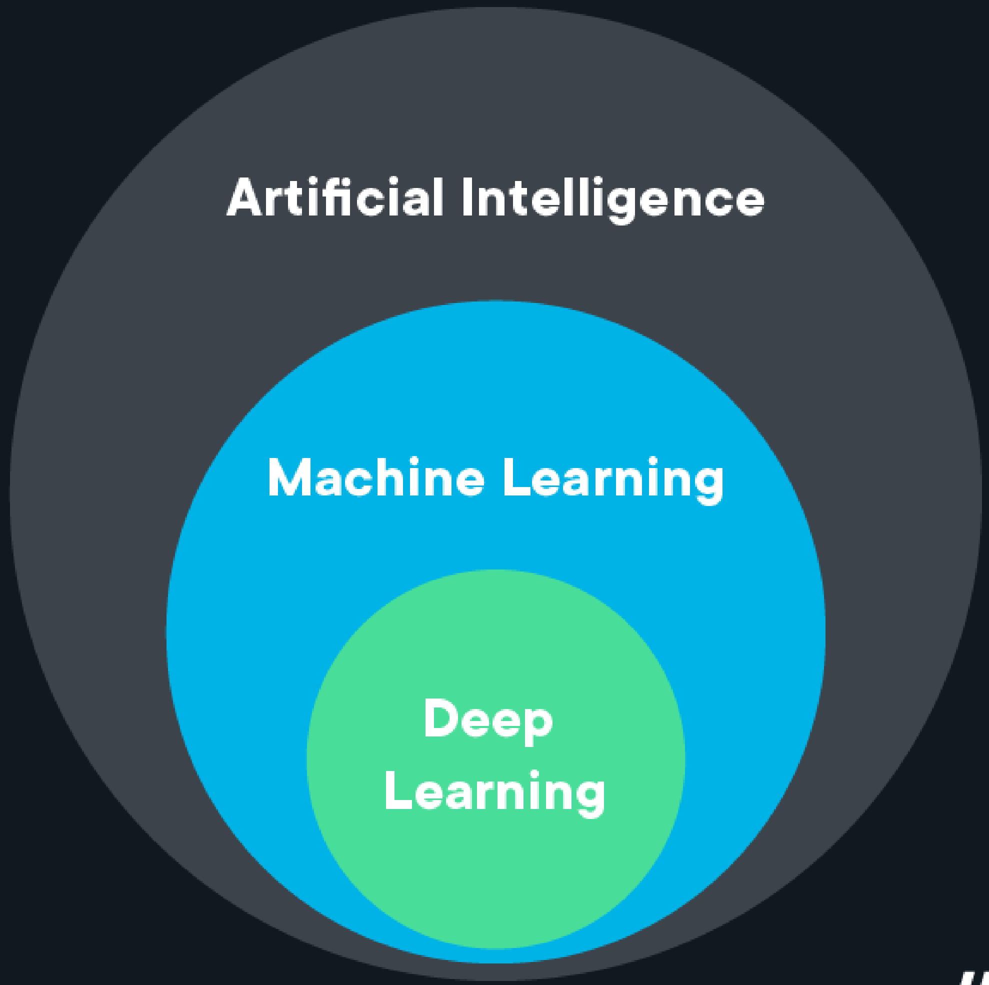
Dimensionality Reduction | PCA

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.



When and When not to do ML

When to do ML

When not to do 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.

When to do ML:

- ***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

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