



MISKOLCI
EGYETEM
UNIVERSITY OF MISKOLC



MEK
1735
MISKOLCI EGYETEM
MŰSZAKI FÖLD- ÉS
KÖRNYEZETTUDOMÁNYI
KAR



National Talent
Program



KULTURÁLIS ÉS INNOVÁCIÓS
MINISZTERIUM

SHORT COURSE INTRODUCTION TO MACHINE LEARNING APPLICATIONS IN GEOSCIENCES

PROJECT FUNDED BY: NTP-SZKOLL-25-0086 Chemstream

Prepared by:

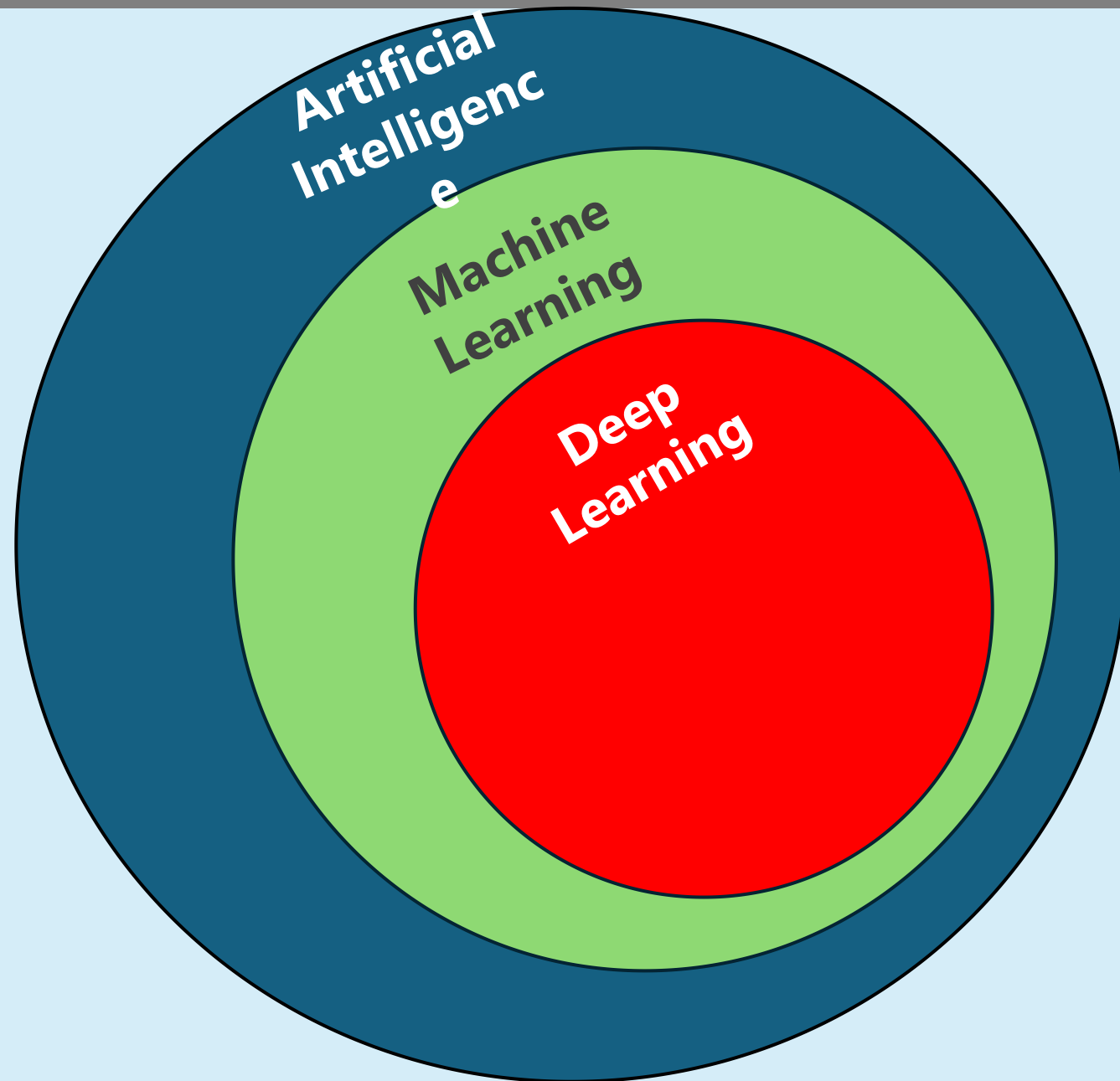
Prof. Dr. Norbert Péter Szabó (NPSZ), *full professor*

Obidiegwu Martins Uchenna (OMU), *Teaching Assistant*

***Institute of Exploration Geosciences, Department of Geophysics,
University of Miskolc, Hungary***

7-8 August 2026

Introduction





What is Machine Learning?

- ❖ Machine learning is the branch of **Artificial Intelligence** that focuses on **developing models** and **algorithms** that let computers learn from data and improve from previous experience without being explicitly programmed for every task.
- ❖ In simple words, **ML teaches** the systems to **think** and **understand** like **humans** by **learning from the data**.

Why do we need Machine Learning?

- ❖ **Traditional programming** requires **exact instructions** and doesn't handle **complex tasks** like understanding images or language well. It can't efficiently process large amounts of data.
- ❖ **Machine Learning** solves these problems by learning from examples and making predictions without fixed rules. Let's see various reasons why it is essential:

1. Solving Complex Business Problems

- ❖ Traditional programming struggles with tasks like language understanding and medical diagnosis. ML learns from data and predicts outcomes easily.

✓ **Examples:**

- ❖ Image and speech recognition in healthcare.
- ❖ Language translation and sentiment analysis.



Why do we need Machine Learning?

2. Handling Large Volumes of Data

❖ The internet generates huge amounts of data every day. Machine Learning processes and analyzes this data quickly by providing valuable insights and real-time predictions.

✓ Examples:

- ❖ Fraud detection in financial transactions.
- ❖ Personalized feed recommendations on Facebook and Instagram from billions of interactions.
- ❖ Geosciences, Medical, and Engineering data.

3. Automate Repetitive Tasks

❖ ML automates time-consuming, repetitive tasks with high accuracy hence reducing manual work and errors.

✓ Examples:

- ❖ Gmail filters spam emails automatically.
- ❖ Chatbots handling order tracking and password resets.
- ❖ Automating large-scale invoice analysis for key insights.



What Makes a Machine "Learn"?

- ❖ A machine **"learns"** by **identifying patterns** in **data** and **improving** its **ability** to **perform** specific tasks without being explicitly programmed for every scenario.
- ❖ This **learning process** helps machines to **make accurate predictions** or **decisions based** on the **information** they receive.
- ❖ **Unlike traditional programming**, where **instructions are fixed**, **ML allows models** to **adapt** and **improve** through experience.



How does the Learning Process Work?

The learning process works:

- ❖ **Data Input:** The Machine needs data like text, images, or numbers to analyze. Good-quality, sufficient data are essential for effective learning.
- ❖ **Algorithms:** Algorithms are mathematical methods that help the machine find patterns in data. Different algorithms are used for various tasks, such as classification or regression.
- ❖ **Model Training:** During training, the machine adjusts its internal settings to predict outcomes better. It learns by reducing the difference between its predictions and actual results (Measured and calculated data).
- ❖ **Feedback Loop:** The Machine compares its predictions with actual outcomes and uses this feedback to correct errors. Techniques like **gradient descent** help it update and improve.
- ❖ **Experience and Iteration:** The Machine repeats training many times with data, which helps in refining its predictions with each pass. More data and iterations improve accuracy.
- ❖ **Evaluation and Generalization:** Model is tested on unseen data to ensure it performs well on real-world tasks.



Critical Aspect of Machine Learning

- ❖ **Data is the Key:** The algorithms that drive machine learning are critical to success. An ML algorithm builds a mathematical model based on sample data, known as “training data”, to make predictions or decisions without being explicitly programmed to do so.
- ❖ **AI is the goal:** ML provides the foundation for AI systems that automate processes and solve data-based business and scientific problems autonomously.



Important of Data in Machine Learning

- ❖ Data is the foundation of machine learning (ML); without quality data, ML models cannot learn, perform, or make accurate predictions.
- ❖ Data provides the examples from which models learn patterns and relationships.
- ❖ High-quality and diverse data improve how well models perform and generalize to new situations.
- ❖ It helps models to understand real-world scenarios and adapt to practical uses. Features extracted from data are essential for effective training.
- ❖ Separate datasets for validation and testing to measure how well the model works on unseen data.
- ❖ Data drives continuous improvements in models through feedback loops.



Types of Machine Learning

There are **several types** of machine learning, each with special characteristics and applications. Some of the main types of machine learning algorithms are as follows:

- ❖ **Supervised Machine Learning**

- ❖ **Unsupervised Machine Learning**

- ❖ **Reinforcement Learning**

- ❖ Additionally, there is a more specific category called **semi-supervised learning**, which combines **elements of both supervised and unsupervised learning**.



Types of Machine Learning

Types of Machine Learning



Types of Machine Learning

1. Supervised Machine Learning

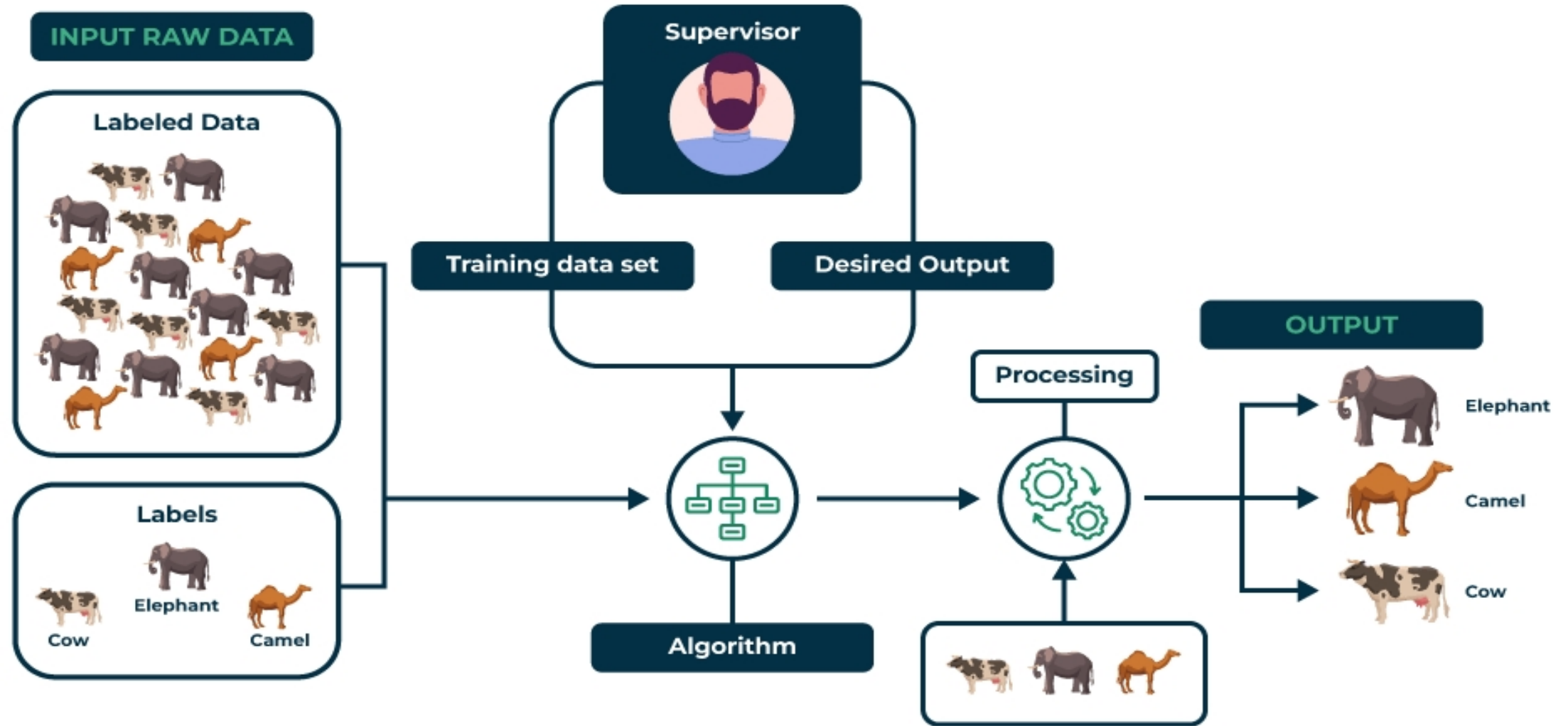
Supervised learning is defined as when a **model** gets **trained** on a **"Labelled Dataset"**.

❖ **Labelled datasets** have both **input** and **output parameters**.

❖ In **Supervised Learning**, **algorithms** learn to **map points** between **inputs** and **correct outputs**. It has both **labelled training** and **validation datasets**.



Supervised Machine Learning Architecture



Supervised Machine Learning Architecture

Let's understand it with an example.

- ❖ Let's consider a **scenario** where you must **build an image classifier** to differentiate between **cats and dogs**.
- ❖ If you **feed** the **datasets** of **dogs** and **cats labelled images** to the **algorithm**, the **machine** will **learn to classify** between a dog and a cat from these labelled images.
- ❖ When we **input new dog** or **cat images** that it has **never seen before**, it will use the **learned algorithms** and **predict whether** it is a **dog** or a **cat**.
- ❖ This is how **supervised learning** works, and this is particularly an **image classification**.
- ❖ There is also a scenario where **numerical data** can be **predicted** from **labelled numerical data**. It is called the **Regression technique**.



Categories of Supervised Learning

Two main categories of supervised learning are mentioned below:

❖ **Classification**

❖ **Regression**



Classification

❖ **Classification** deals with **predicting categorical target variables**, which represent **discrete classes** or **labels**. For instance, classifying **emails** as **spam** or **not spam**, or **predicting whether** a **patient** has a **high risk** of **heart disease**. Classification algorithms learn to map the **input features** to one of the **predefined classes**.

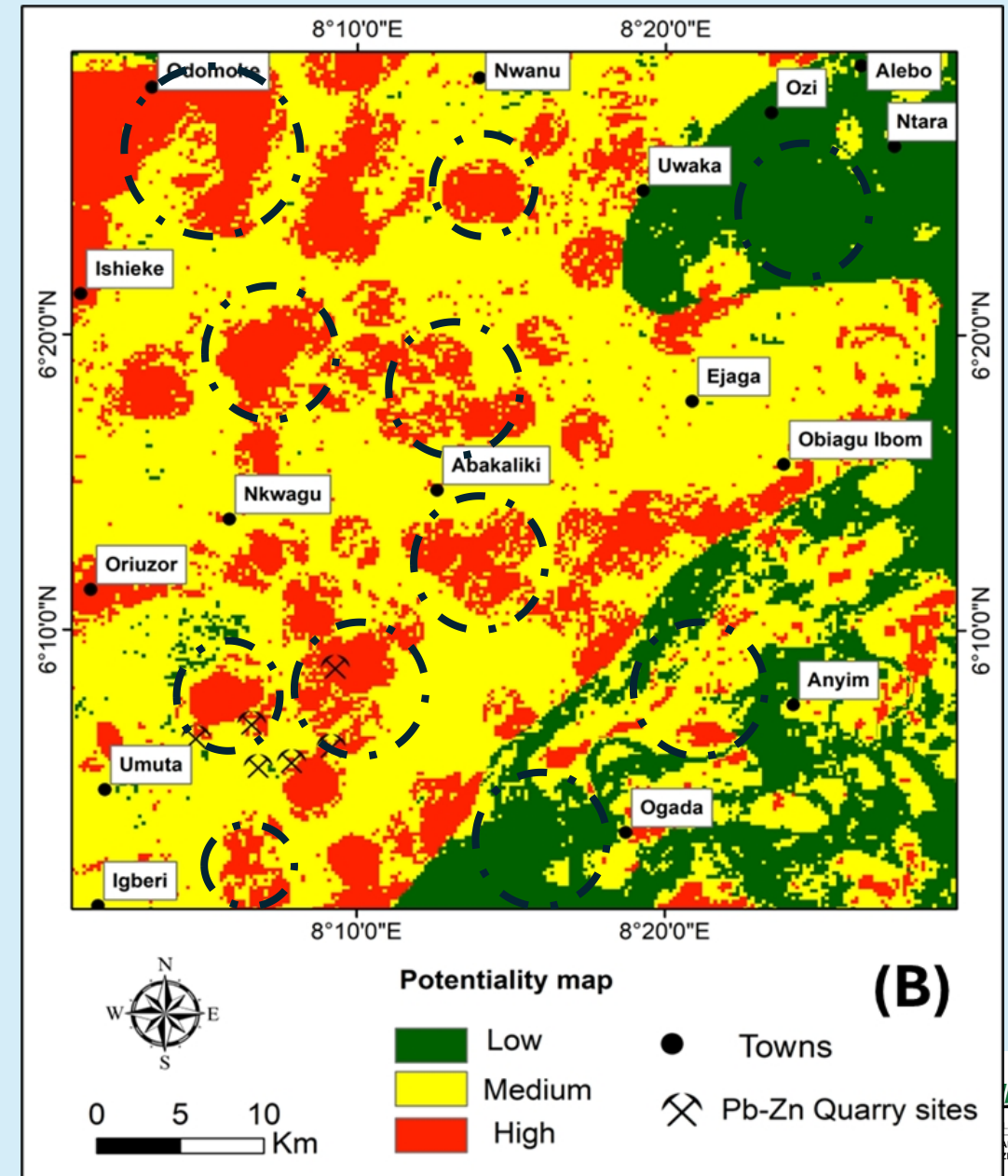
❖ Here are some classification algorithms:

- Logistic Regression
- Support Vector Machine
- Random Forest
- Decision Tree
- K-Nearest Neighbors (KNN)
- Naive Bayes



Classification

Example of a result in which the classification method has been applied (*Martins U.O. TDK Thesis 2025*).



Regression

❖ **Regression**, on the other hand, deals with **predicting continuous target variables**, which represent **numerical values**. For example, **predicting the price of a house based on its size, location, and amenities**, or **forecasting the sales of a product**. Regression algorithms learn to map the **input features** to a **continuous numerical value**.

❖ **Here are some regression algorithms:**

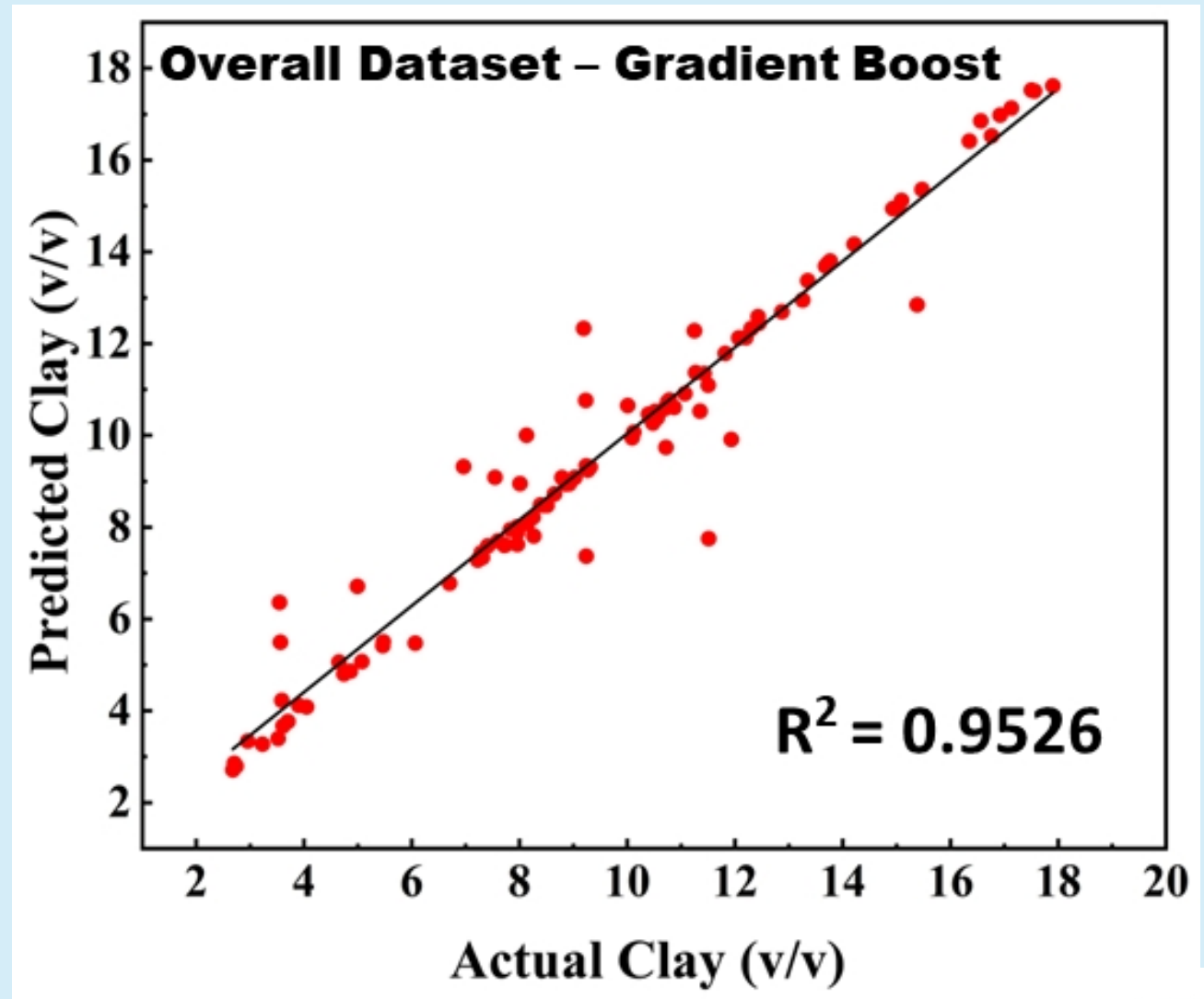
- Linear Regression
- Polynomial Regression
- Ridge Regression
- Lasso Regression
- Decision tree

Random Forest



Regression Analysis

Example of a result in which the Regression Analysis has been applied (*Martins U.O. TDK Thesis 2024*).



Advantages of Supervised Machine Learning

- ❖ **Supervised Learning** models can have high accuracy as they are trained on **labelled data**.
- ❖ The process of **decision-making** in supervised learning models is often interpretable.
- ❖ It can often be used in **pre-trained models**, which saves time and resources when **developing new models from scratch**.



Disadvantages of Supervised Machine Learning

- ❖ It has limitations in knowing patterns and may struggle with unseen or unexpected patterns that are not present in the training data.
- ❖ It can be time-consuming and costly as it relies on **labelled** data only.
- ❖ It may lead to **poor generalizations** based on **new data**.



Applications of Supervised Machine Learning

Supervised learning is used in a wide variety of applications, including:

- ❖ **Geosciences and Engineering:** for predicting geophysical datasets, such as well log data, seismic, production data, etc.
- ❖ **Image classification:** Identify objects, faces, and other features in images.
- ❖ **Natural language processing:** Extract information from text, such as sentiment, entities, and relationships.
- ❖ **Speech recognition:** Convert spoken language into text.
- ❖ **Recommendation systems:** Make personalized recommendations to users.
- ❖ **Predictive analytics:** Predict outcomes, such as sales, customer churn, and stock prices.



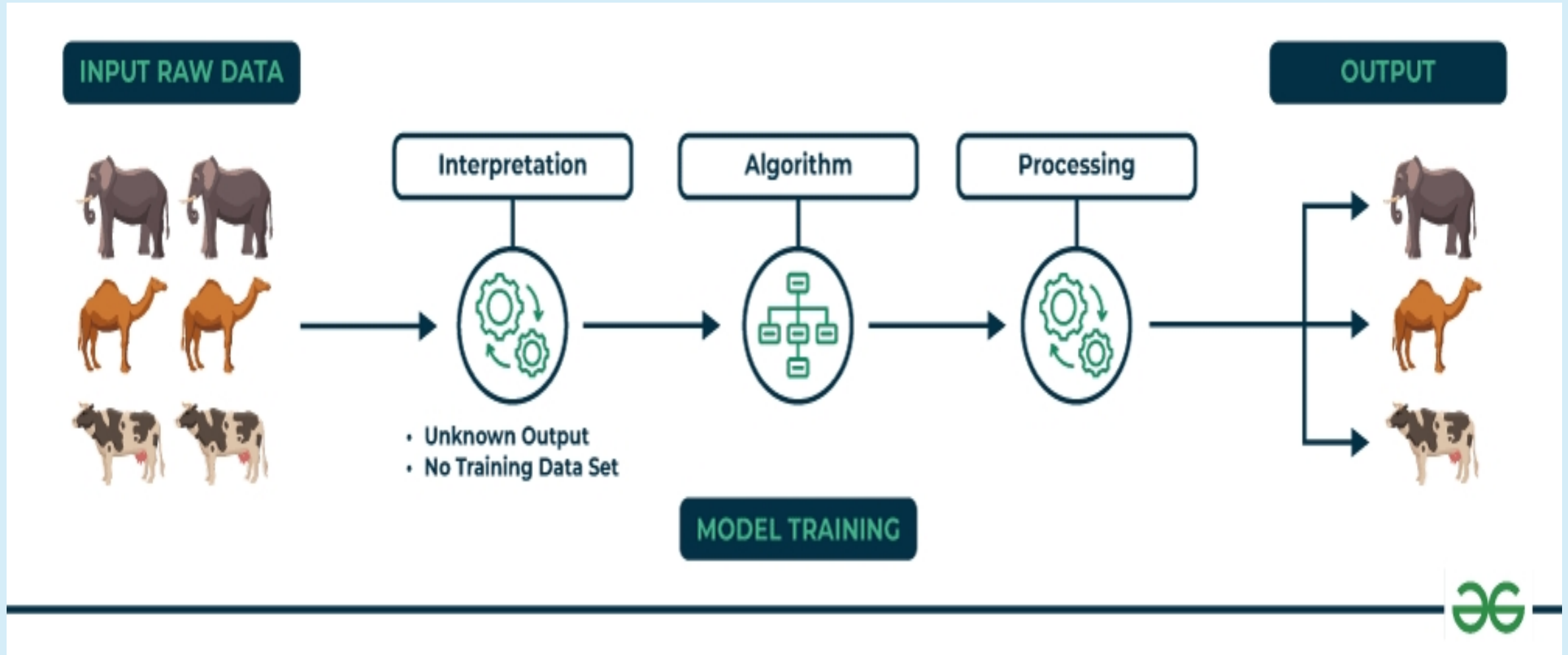
❖ **Medical diagnosis:** Detect diseases and other medical conditions

Unsupervised Machine Learning

- ❖ **Unsupervised Learning** is a type of **machine learning technique** in which an **algorithm discovers patterns and relationships** using **unlabelled data**. Unlike supervised learning, unsupervised learning **doesn't involve** providing the algorithm with **labelled target outputs**.
- ❖ The **primary goal** of Unsupervised learning is often to discover **hidden patterns, similarities, or clusters** within the data, which can then be used for various purposes, such as **data exploration, visualization, dimensionality reduction**, and more.



Unsupervised Machine Learning Architecture



Let's understand it with an example.

- ❖ Consider that you have a dataset that contains well logging information and parameters such as Natural Gamma ray, Resistivity logs, and porosity logs.
- ❖ Through clustering, the algorithm can group parameters having the same properties, which reveals potential parameters without predefined labels.
- ❖ This type of information can help Geophysicists get target well logging parameters as well as identify outliers.



Categories of Unsupervised Learning

Two main categories of **unsupervised learning** are mentioned below:

❖ **Clustering**

❖ **Association**



Clustering

❖ **Clustering** is the process of grouping data points into clusters based on their similarity. This technique helps identify patterns and relationships in data without the need for labelled examples.

Here are some clustering algorithms:

- ❖ K-Means Clustering algorithm
- ❖ Mean-shift algorithm
- ❖ DBSCAN Algorithm
- ❖ Principal Component Analysis
- ❖ Independent Component Analysis (PCA)

Association Rule Learning

❖ **Association rule learning** is a technique for **discovering** relationships between **items** in a **dataset**. It **identifies** rules that **indicate** the presence of one item implies the presence of another item with a specific probability.

Here are some association rule learning algorithms:

- ❖ Apriori Algorithm
- ❖ Eclat
- ❖ FP-growth Algorithm



Advantages of Unsupervised Machine Learning

- ❖ It helps to discover **hidden patterns** and **various relationships between the data**.
- ❖ Used for tasks such as **customer segmentation, anomaly detection, and data exploration**.
- ❖ It **does not** require **labelled data** and **reduces** the effort of data labelling.

Disadvantages of Unsupervised Machine Learning

- ❖ **Without using labels**, it may be **difficult to predict** the quality of the **model's output**.
- ❖ **Cluster Interpretability** may not be clear and may not have **meaningful interpretations**.
- ❖ It has techniques such as **autoencoders** and dimensionality reduction that can be used to **extract meaningful** features from raw data.



Applications of Unsupervised Machine Learning

Here are some typical applications of unsupervised learning:

- ❖ **Clustering:** Group similar data points into clusters.
- ❖ **Anomaly detection:** Identify outliers or anomalies in data.
- ❖ **Dimensionality reduction:** Reduce the dimensionality of data while preserving its essential information.
- ❖ **Recommendation systems:** Suggest products, movies, or content to users based on their historical behavior or preferences.
- ❖ **Topic modeling:** Discover latent topics within a collection of documents.
- ❖ **Density estimation:** Estimate the probability density function of data.



Applications of Unsupervised Machine Learning

- ❖ **Data preprocessing:** Help with data preprocessing tasks such as data cleaning, imputation of missing values, and data scaling.
- ❖ **Market basket analysis:** Discover associations between products.
- ❖ **Genomic data analysis:** Identify patterns or group genes with similar expression profiles.
- ❖ **Image segmentation:** Segment images into meaningful regions.
- ❖ **Community detection in social networks:** Identify communities or groups of individuals with similar interests or connections.
- ❖ **Content recommendation:** Classify and tag content to make it easier to recommend similar items to users.
- ❖ **Exploratory data analysis (EDA):** Explore data and gain insights before defining specific tasks.



Reinforcement Machine Learning

- ❖ **Reinforcement machine learning algorithm** is a **learning method** that **interacts** with the environment by producing actions and discovering **errors**. **Trial, error, and delay** are the most relevant characteristics of reinforcement learning.
- ❖ **In this technique**, the model keeps on **increasing** its **performance** using **Reward Feedback** to learn the **behaviour or pattern**.
- ❖ **These algorithms** are specific to a **particular problem** e.g. Google Self Driving car, AlphaGo where a bot competes with humans and even itself to get better performers in Go Game. Each time we feed in data, they learn and add it to their knowledge, which serves as training data. So, the more it learns, the better it gets trained and hence experienced.



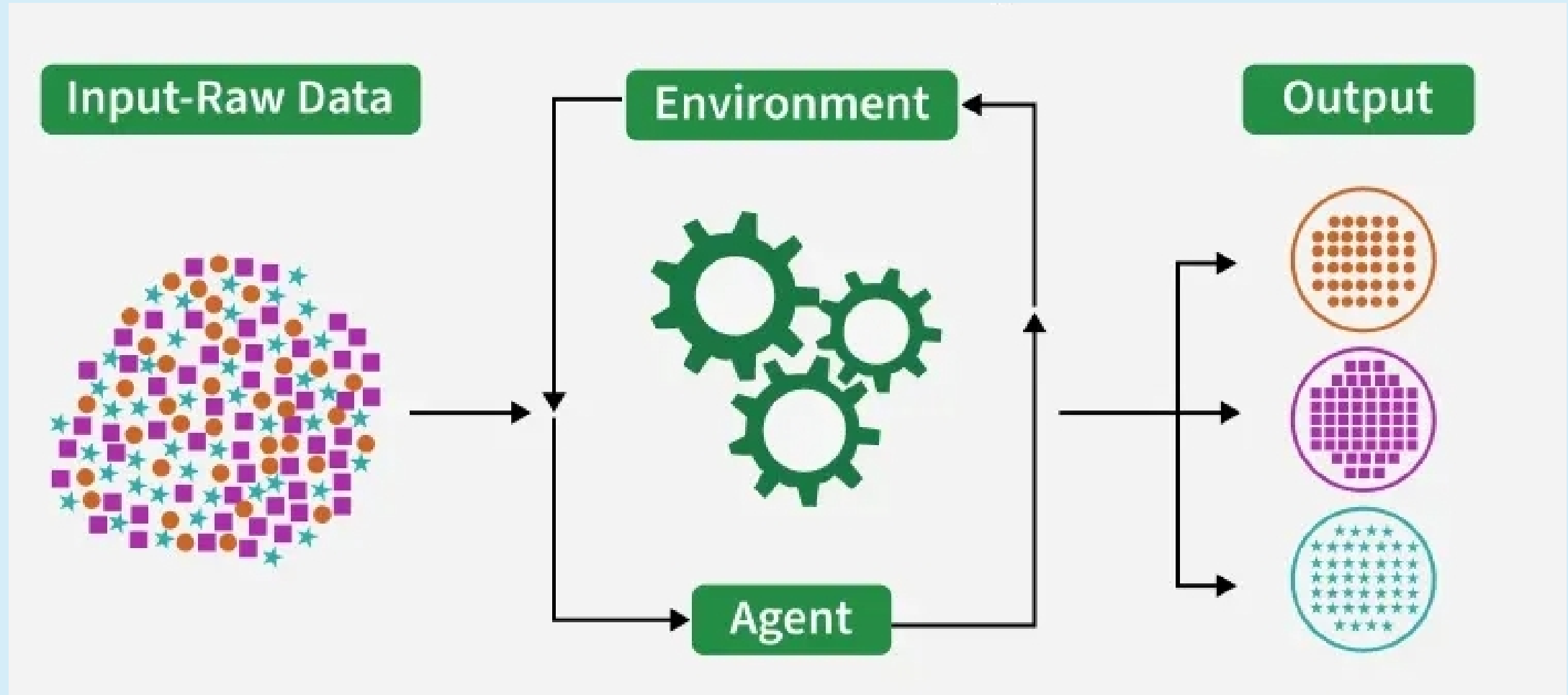
Types of Reinforcement Learning Algorithms

Here are some of the most common reinforcement learning algorithms:

- ❖ **Q-learning:** Q-learning is a **model-free RL algorithm** that learns a **Q-function**, which **maps states to actions**. The **Q-function** estimates the expected reward of taking a particular action in a given state.
- ❖ **SARSA (State-Action-Reward-State-Action):** SARSA is another **model-free RL algorithm** that learns a Q-function. However, **unlike Q-learning**, SARSA updates the Q-function for the action that was **actually taken**, rather than the optimal action.
- ❖ **Deep Q-learning:** Deep Q-learning is a **combination** of **Q-learning** and **deep learning**. Deep Q-learning uses a **neural network** to represent the Q-function, which allows it to **learn complex** relationships between **states and actions**.



Reinforcement Machine Learning Architecture



Reinforcement Learning Algorithms

Let's understand it with an example.

- ❖ Consider that you are training an AI agent to play a game like chess.
- ❖ The agent explores different moves and receives positive or negative feedback based on the outcome.
- ❖ Reinforcement Learning also finds applications in which they learn to perform tasks by interacting with their surroundings.

Types Reinforcement Machine Learning

There are two main types of reinforcement learning:

❖ **Positive reinforcement**

- Rewards the agent for taking a desired action.
- Encourages the agent to repeat the behaviour.
- Examples: Giving a treat to a dog for sitting, providing a point in a game for a correct answer.

❖ **Negative reinforcement**

- Removes an undesirable stimulus to encourage a desired behaviour.
- Discourages the agent from repeating the behaviour.
- Examples: Turning off a loud buzzer when a lever is pressed, avoiding a penalty by completing a task.



Advantages of Reinforcement Machine Learning

- ❖ It has **autonomous decision-making** that is **well-suited** for tasks and that **can learn to make** a sequence of **decisions**, like **robotics** and **game-playing**.
- ❖ This technique is preferred to **achieve long-term results** that are **very difficult to achieve**.
- ❖ It is used to **solve complex problems** that **conventional techniques cannot solve**.



Disadvantages of Reinforcement Machine Learning

- ❖ Training Reinforcement Learning **agents** can be computationally **expensive** and **time-consuming**.
- ❖ Reinforcement learning is **not preferable** to **solving simple problems**.
- ❖ It needs **a lot of data** and a **lot of computation**, which makes it **impractical and costly**.



Applications of Reinforcement Machine Learning

Here are some applications of reinforcement learning:

- ❖ **Game Playing:** RL can teach agents to play games, even complex ones.
- ❖ **Robotics:** RL can teach robots to perform tasks autonomously.
- ❖ **Autonomous Vehicles:** RL can help self-driving cars navigate and make decisions.
- ❖ **Recommendation Systems:** RL can enhance recommendation algorithms by learning user preferences.
- ❖ **Healthcare:** RL can be used to optimize treatment plans and drug discovery.
- ❖ **Natural Language Processing (NLP):** RL can be used in dialogue systems and chatbots.
- ❖ **Finance and Trading:** RL can be used for algorithmic trading.



Applications of Reinforcement Machine Learning

- ❖ **Supply Chain and Inventory Management:** RL can be used to optimize supply chain operations.
- ❖ **Energy Management:** RL can be used to optimize energy consumption.
- ❖ **Game AI:** RL can be used to create more intelligent and adaptive NPCs in video games.
- ❖ **Adaptive Personal Assistants:** RL can be used to improve personal assistants.
- ❖ **Virtual Reality (VR) and Augmented Reality (AR):** RL can be used to create immersive and interactive experiences.
- ❖ **Industrial Control:** RL can be used to optimize industrial processes.
- ❖ **Education:** RL can be used to create adaptive learning systems.
- ❖ **Agriculture:** RL can be used to optimize agricultural operations.



Semi-Supervised Machine Learning (Supervised + Unsupervised)

- ❖ **Semi-supervised learning is a machine learning algorithm that works between** supervised and unsupervised learning, so it uses both **labelled and unlabelled** data. It's particularly useful when obtaining labelled data is costly, time-consuming, or resource-intensive. This approach is practical when the dataset is expensive and time-consuming to obtain. Semi-supervised learning is chosen when labelled data is scarce and requires specialized skills and resources to train on or learn from it.
- ❖ We use these techniques when we are **dealing with data that is a little bit labelled**, and the **rest of the large portion of it is unlabelled**. We can use the unsupervised techniques to predict labels and then feed these labels to supervised techniques. This technique is most applicable to image datasets, where not all images are usually labelled.



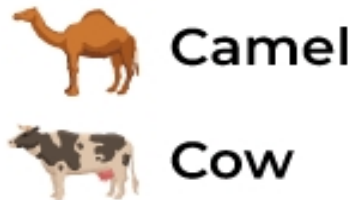
Semi-Supervised Machine Learning Architecture



Input Data



Partial Labels



Machine Learning Model



Prediction

It's an Elephant

Unlabelled Data



Semi-Supervised Machine Learning

Let's understand it with an example.

- ❖ Consider that we are **building a language translation model**, having **labelled translations** for every sentence pair can be resource-intensive.
- ❖ It allows the models to learn from **labelled** and **unlabelled** sentence pairs, making them more accurate.
- ❖ This technique has led to **significant improvements** in the quality of **machine translation services**.

Types of Semi-Supervised Machine Learning Methods

There are many different semi-supervised learning methods, each with its own characteristics. Some of the most common ones include:

- ❖ **Graph-based semi-supervised learning:** This approach uses a graph to represent the relationships between the data points. The graph is then used to propagate labels from the labelled data points to the unlabelled data points.
- ❖ **Label propagation:** This approach iteratively propagates labels from the labelled data points to the unlabelled data points, based on the similarities between the data points.

Types of Semi-Supervised Machine Learning Methods

- ❖ **Co-training:** This approach trains two different machine learning models on different subsets of the unlabelled data. The two models are then used to label each other's predictions.
- ❖ **Self-training:** This approach trains a machine learning model on the labelled data and then uses the model to predict labels for the unlabelled data. The model is then retrained on the labelled data and the predicted labels for the unlabelled data.
- ❖ **Generative adversarial networks (GANs):** GANs are a type of deep learning algorithm that can be used to generate synthetic data. **GANs** can be used to **generate unlabelled data** for **semi-supervised learning** by training **two neural networks**, a **generator** and a **discriminator**.



Advantages of Semi-Supervised Machine Learning Methods

- ❖ It leads to better generalization as compared to **supervised learning**, as it takes both **labelled** and **unlabelled data**.
- ❖ Can be applied to a **wide range of data**.



Disadvantages of Semi-Supervised Machine Learning Methods

- ❖ **Semi-supervised** methods can be more **complex to implement** compared to **other approaches**.
- ❖ It still requires some **labelled data** that might not always be available or easy to obtain.
- ❖ The **unlabelled data** can impact the **model performance** accordingly.

Application of Semi-Supervised Machine Learning Methods

Here are some typical applications of semi-supervised learning:

- ❖ **Image Classification and Object Recognition:** Improve the accuracy of models by combining a small set of labelled images with a larger set of unlabelled images.
- ❖ **Natural Language Processing (NLP):** Enhance the performance of language models and classifiers by combining a small set of labelled text data with a vast amount of unlabelled text.
- ❖ **Speech Recognition:** Improve the accuracy of speech recognition by leveraging a limited amount of transcribed speech data and a more extensive set of unlabelled audio.
- ❖ **Recommendation Systems:** Improve the accuracy of personalized recommendations by supplementing a sparse set of user-item interactions (labelled data) with a wealth of unlabelled user behaviour data.
- ❖ **Healthcare and Medical Imaging:** Enhance medical image analysis by utilizing a small set of labelled medical images alongside a larger set of unlabelled images.



Benefits of Machine Learning

- ❖ **Enhanced Efficiency and Automation:** **ML automates** repetitive tasks, freeing up human resources for more complex work. This leads to faster, smoother processes and **higher productivity**.
- ❖ **Data-Driven Insights:** It can analyze large amounts of data, such as **well logs, production data, and time series data**, to identify patterns and trends that people might miss and help researchers in geosciences, engineering, and businesses make better decisions.
- ❖ **Improved Personalization:** It customizes user experiences by **tailoring recommendations** and **ads based on individual** preferences.
- ❖ **Advanced Automation and Robotics:** It helps **robots** and **machines** to perform complex tasks with **greater accuracy** and **adaptability**. This is transforming industries like **engineering, manufacturing, and logistics**.



Challenges of Machine Learning

- ❖ **Data Bias and Fairness:** ML models learn from **training data, and if the data is biased, the model's decisions can be unfair;** therefore, it's crucial to select and monitor data carefully.
- ❖ **Security and Privacy Concerns:** Since it depends on large amounts of data, there is a risk of sensitive information being exposed, so protecting privacy is essential.
- ❖ **Interpretability and Explainability:** Complex ML models can be **challenging to understand**, which makes it difficult to explain why they make certain decisions. **This can affect trust and accountability.**
- ❖ **Job Displacement and Automation:** Automation may **replace some jobs**, so, **retraining** and **helping workers** learn new skills are important to adapt to these changes.



Applications of Machine Learning

Machine Learning is used in many industries to solve problems and improve services. Here are some typical real-world applications:

- ❖ **Geosciences:** It helps Geophysicists and Geoscientists in evaluating various datasets such as seismic, well logging, geoelectric, aeromagnetic, radiometric data, and many more datasets.
- ❖ **Healthcare:** It helps doctors to diagnose diseases from medical images like X-rays and MRIs. It also predicts patient outcomes and personalizes treatments, which improves healthcare quality.
- ❖ **Finance:** In finance, it detects fraudulent transactions in real time and supports algorithmic trading. It also helps to assess credit risk, making lending safer and faster.



Applications of Machine Learning

- ❖ **Transportation and Automotive:** **Self-driving cars** rely on ML to navigate and make decisions. It optimizes delivery routes and predicts vehicle maintenance needs, which reduces downtime.
- ❖ **Social Media and Entertainment:** Platforms like **Netflix** and **YouTube** use ML to recommend content we'll enjoy. It enables image and speech recognition for better user interaction.
- ❖ **Manufacturing:** It improves **quality control** by **detecting defects** in products automatically, **predicting machine failures** in advance, and helping in **production processes**.



How Machine Learning Works in Geophysics

There are four basic steps for building a machine learning application (or model)

❖ **Step 1: Select and prepare a training data set**

Training data is a data set representative of the data the machine learning model will ingest to solve the problem it's designed to solve.

❖ **Step 2: Choose an algorithm to run on the training data set**

An algorithm is a set of statistical processing steps. The type of algorithm depends on the type (labelled or unlabelled) and amount of data in the training data set, and on the kind of problem to be solved.

❖ **Step 3: Training the algorithm to create the model**

Training the algorithm is an iterative process

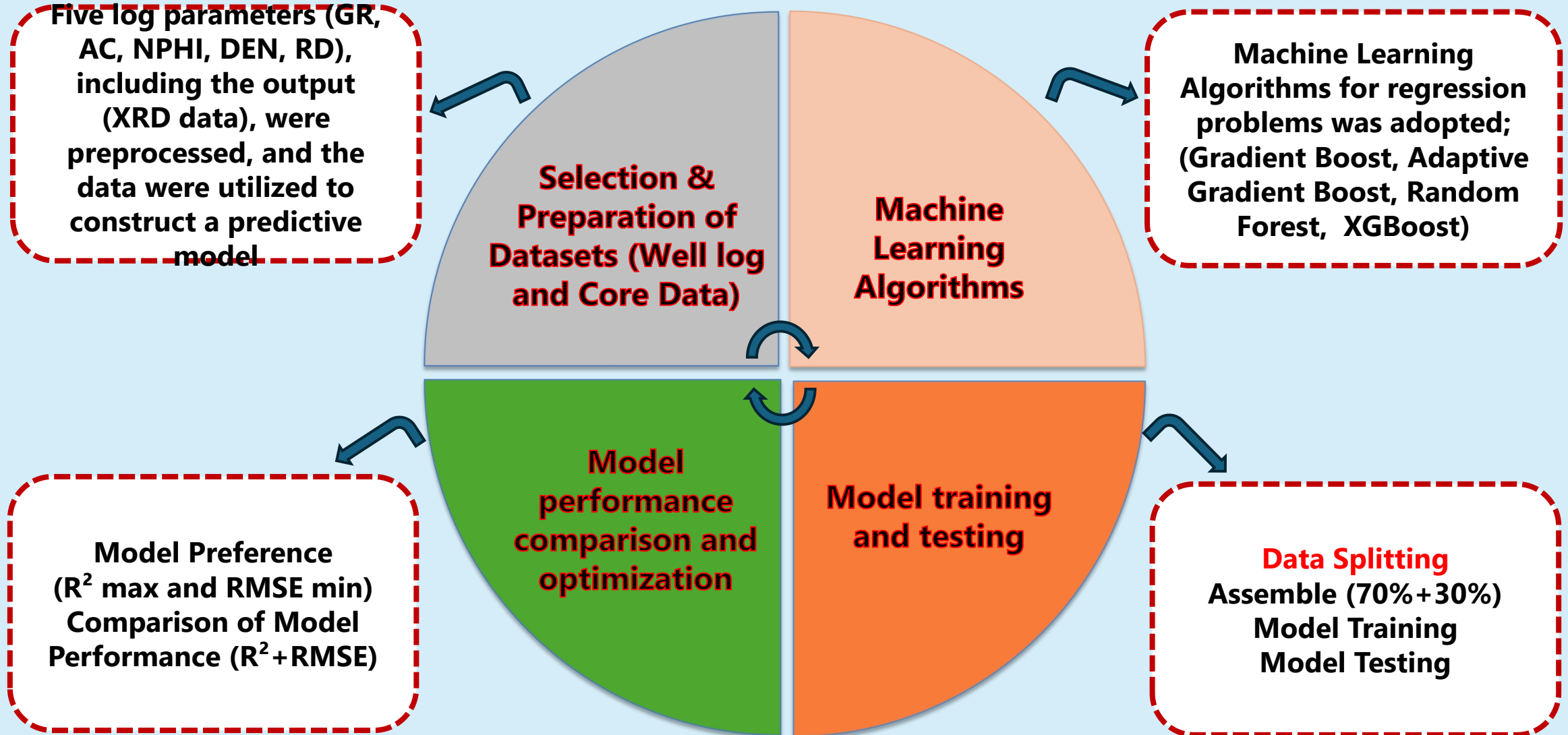
❖ **Step 4: Using and improving the model**

For it to improve in accuracy and effectiveness over time, use the model with new data



How Machine Learning Works

Overall Data Analysis and Result

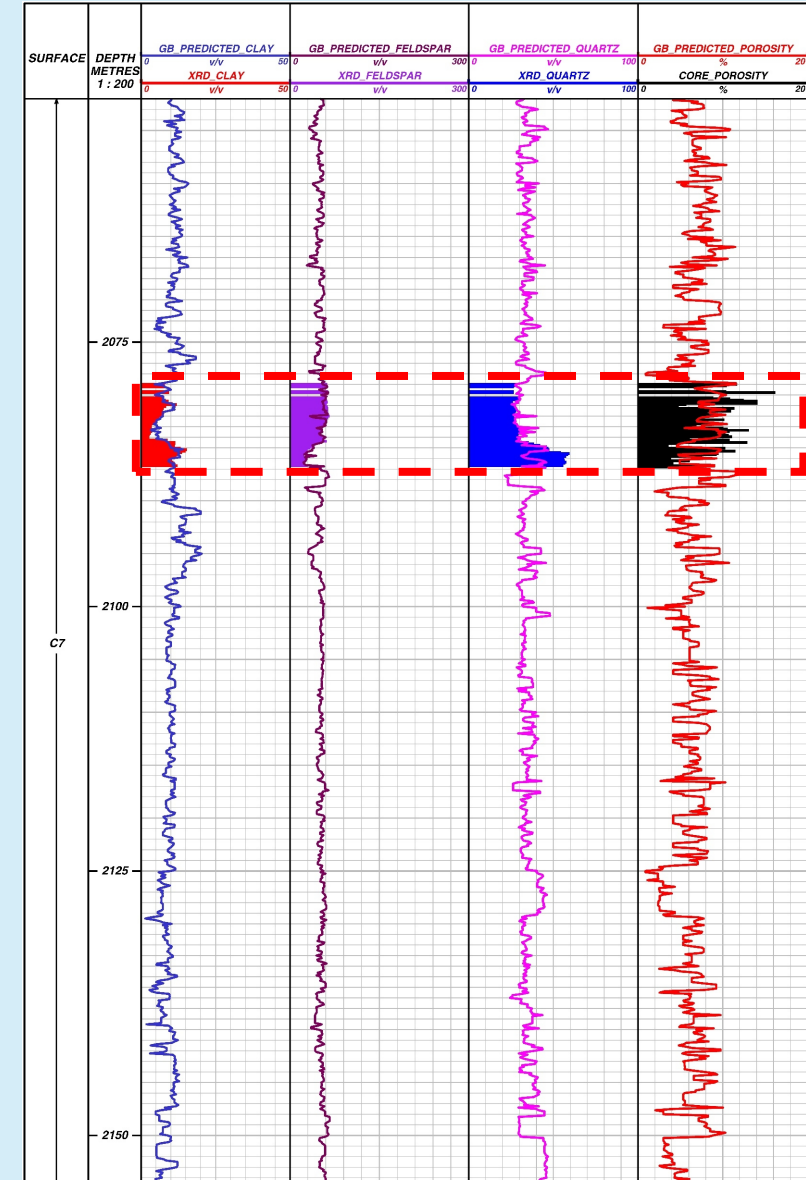
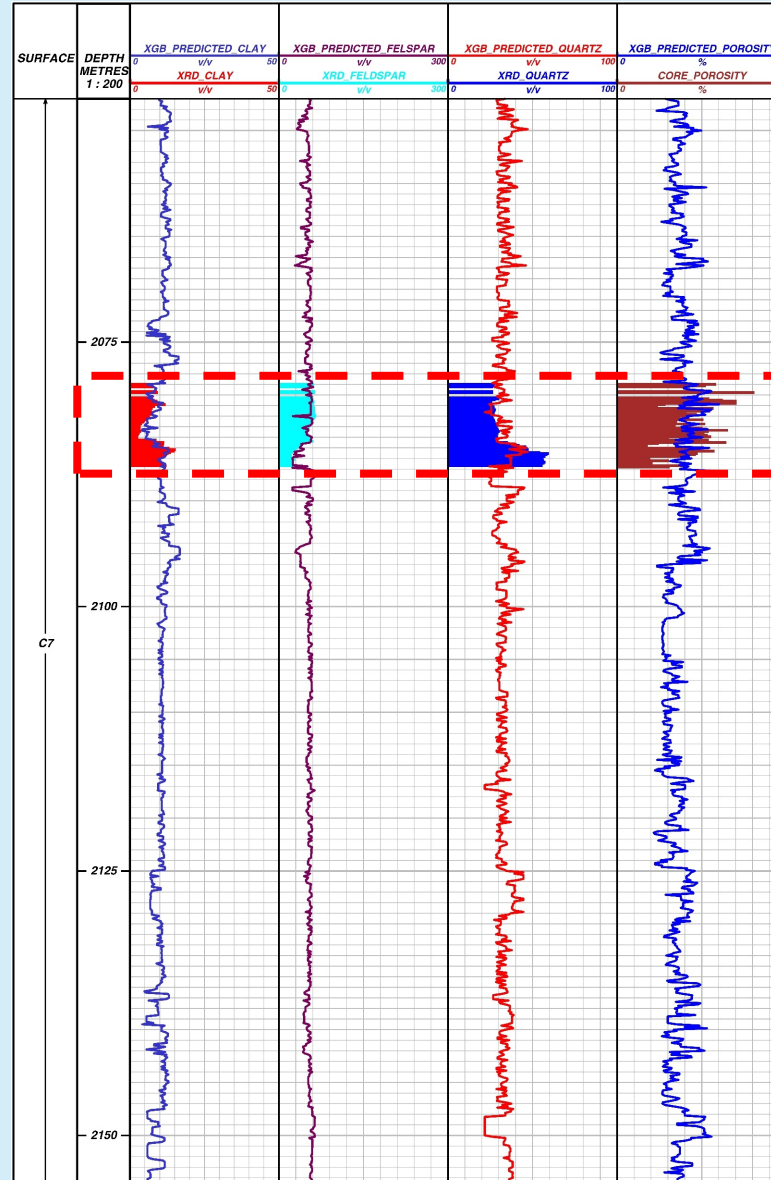
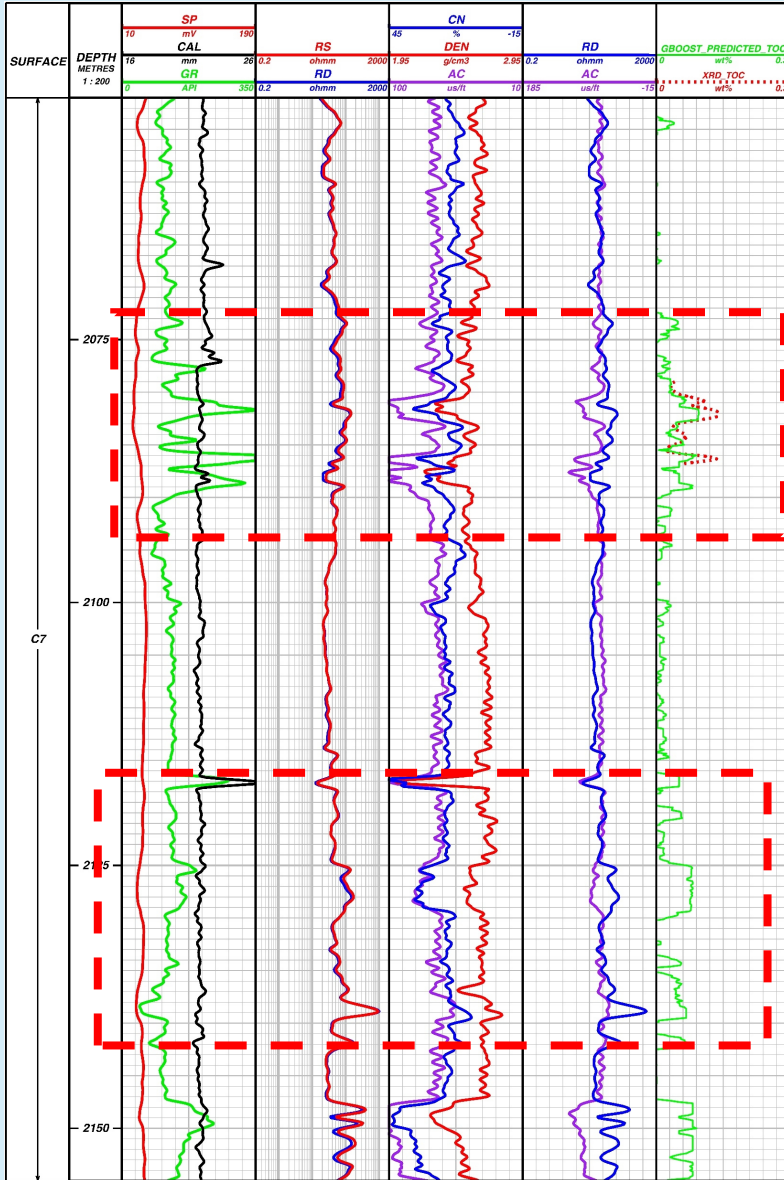


Advantages of Using ML in Petrophysical Analysis

- ❖ ML doesn't require prior knowledge of the petrophysical response equations.
- ❖ ML is self-calibrating. Just give it the data
- ❖ ML avoids the problem of 'Garbage In, Garbage Out',
✓ by ignoring noise and outliers
- ❖ There is very little user intervention
✓ There are no parameters to pick or cross-plots to make
- ❖ ML programs work with an unlimited number of well logs, electrical logs, core, and gas chromatography data, and don't 'fall over' if some of those inputs are missing
- ❖ ML is not a Black Box, as it provides insights into how it makes predictions



Results Obtained from Well Log Evaluation Using ML



Example of Results Obtained from Well Logging through ML Approach (Martins U.O. TDK Thesis)

2024)

ACKNOWLEDGEMENTS



MISKOLCI
EGYETEM
UNIVERSITY OF MISKOLC



THANK YOU FOR LISTENING

QUESTIONS!!!

