NTNU | Norwegian University of Science and Technology

TGB4155
Computer Methods in Engineering
Machine Learning

Introduction to machine learning

Hakan Basarir

Introduction to Data-Driven Modeling in Georesources

Content

Lecture #1: Introduction to Data-Driven Modeling in Georesources

Part 1:

- Introduction to Data-Driven Modeling
  - Fundamentals: Definition of data-driven modeling and its role in geotechnical engineering.
  - Differences from traditional empirical approaches in engineering geology and mining.
  - Applications & Relevance: Predicting rock properties, stability analysis, and tunnel support design.
  - Discussion on the challenges and benefits of using data-driven methods.
- Data Types & Sources in geotechnical engineering
  - Overview of data used in geotechnical contexts (e.g., geological surveys, borehole logs, MWD data).
- Data presented in used cases

Part 2:

- Machine Learning Basics
  - Types of learning: Supervised, unsupervised, and reinforcement learning.
  - Focus on Supervised Learning: Why it is particularly suitable for geotechnical applications.
- Modeling Techniques
  - Regression Modeling: Linear regression, multiple regression, and their applications in predicting rock properties.
  - Classification Modeling: Decision trees, Random Forests, and applications in rock mass classification.
    - Brief overview of clustering techniques for pattern recognition.
- Performance Evaluation
  - Metrics for regression (RMSE, R^2) and classification models (accuracy, precision, recall).
  - Introduction to Explainable AI (XAI): Importance of interpretability in geotechnical projects.
  - Techniques like SHAP, LIME, and feature importance to enhance model interpretability.

Introduction to Data-Driven Modeling in Geotechnical Engineering

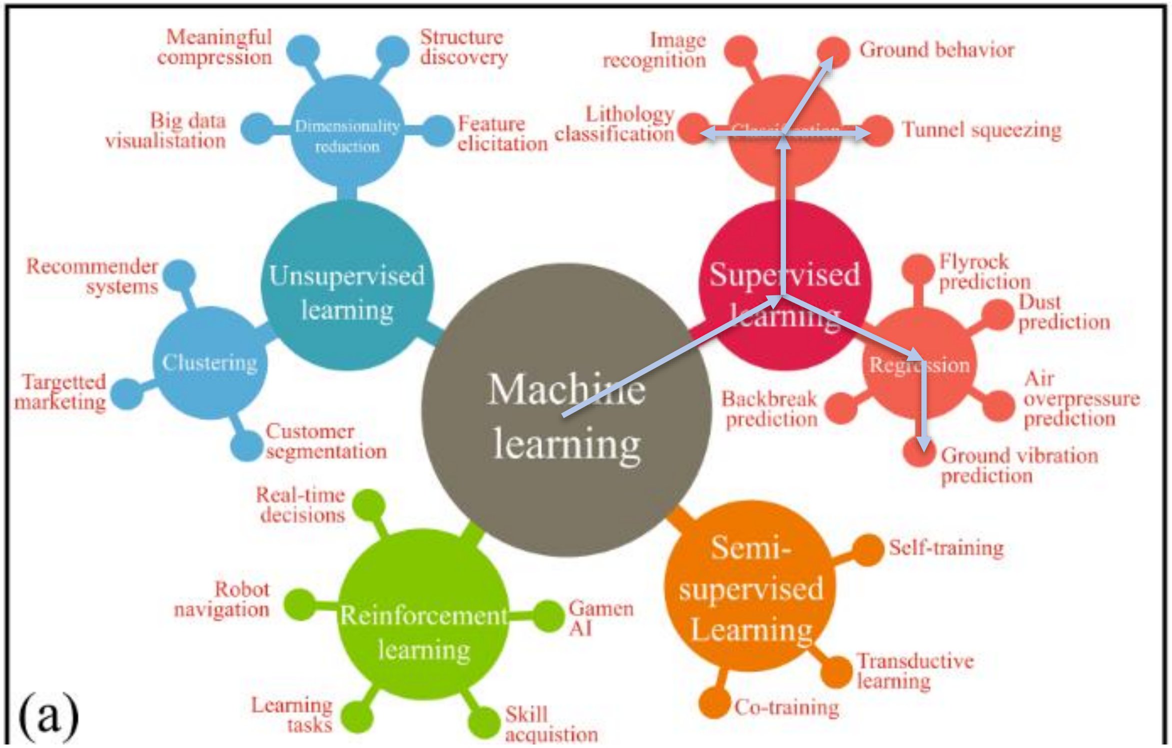
Objective: To understand the role of data-driven modeling in addressing complex geotechnical challenges.

Key Topics:
Fundamentals of data-driven modeling
Differences from traditional empirical methods
Applications in engineering geology and mining
Discussion on the challenges and benefits of using data-driven methods

Introduction to Data-Driven Modeling in Geotechnical Engineering

Data-driven modeling is an approach that uses large volumes of data to build predictive models by identifying patterns, relationships, and trends directly from the data using statistical methods, machine learning, or artificial intelligence, instead of relying on predefined theoretical equations! (can also use theoretical equations to get the data.

This approach is particularly useful in fields where complex, non-linear relationships exist, allowing for more adaptive and accurate solutions compared to traditional, theory-based models.



Introduction to Data-Driven Modeling in Geotechnical Engineering

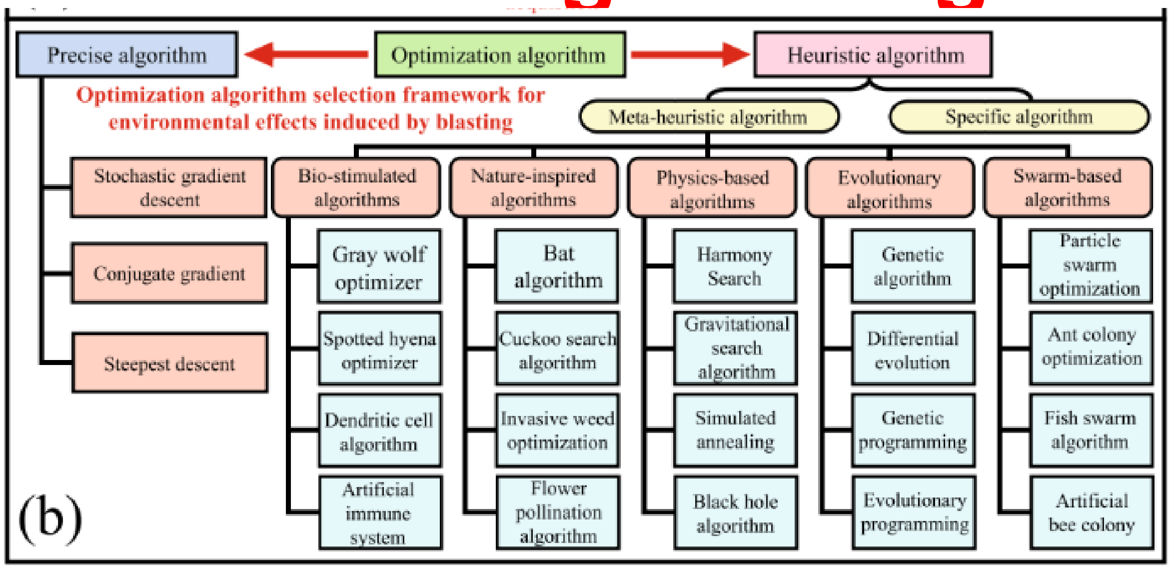


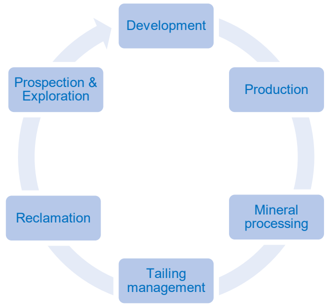
Fig. 9 The main classification of ML algorithms and the selection of OA algorithms. **a** Classification and application of ML algorithms; **b** The category of OA algorithms

What to do ?

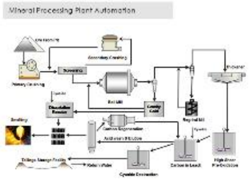
AI based optimization algorithms using machine learning models optimize operations to improve processes like scheduling, maintenance, and resource allocation.

As ML models adapt in real-time, reducing downtime, increasing efficiency, and cutting costs. By automating decision-making, they ensure operations are both efficient and responsive to changing conditions, enhancing productivity and safety.

Why ML, yesyerday ?

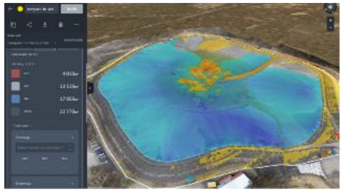


Why ML, today ?



How to link ?


Atlas Copco autonomous drills at BHPB Yandi. (Credit: Atlas Copco)

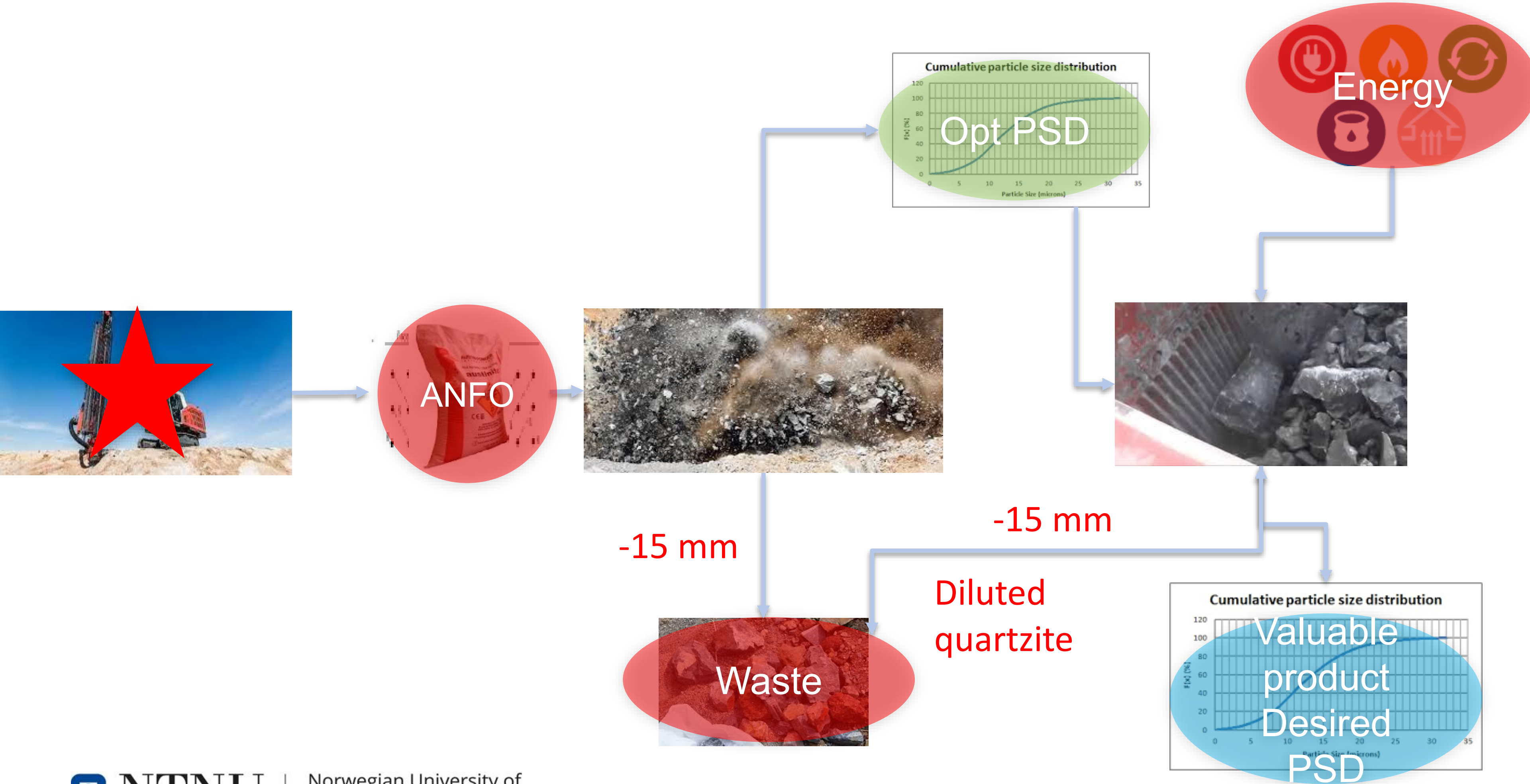


How to optimise ?

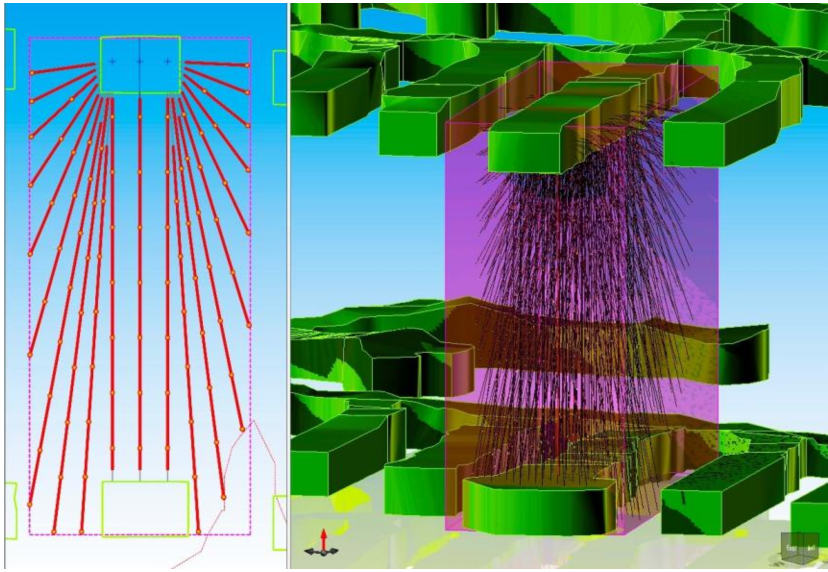
NTNU | Norwegian University of Science and Technology

An example flow, Drilling and blasting

Drilling = f(rock properties, drilling machine props)
Blasting = f(rock properties, blasting agent, blast design parameters, environmental factors)
Crushing = f(rock properties, blasting output, crusher variables, customer demand)



An example flow, Underground backfill

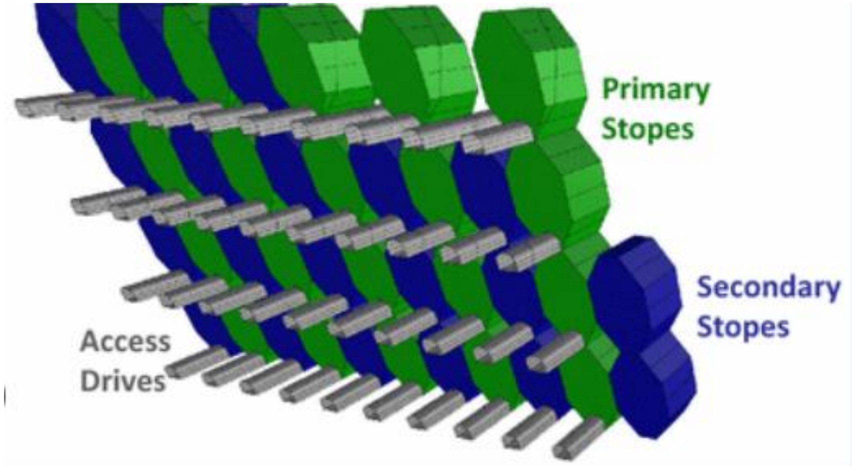


Mixture of water, cement, rock or tailing.
Used to fill the void or produced stope.

Local and global safety.

Maximise recovery.

Reduce environmental impact.



An example flow, Underground backfill

Mixture of water, cement, rock or tailing. Used to fill the void or produced stope.

O1: Strength: UCS
O2: Workability: Shear strength, Slump
O3: Cost!

Key design parameters:
Cement percentage (C):
Solid contents (SD):
Water content (W):

$$C = \frac{Mc}{(Mc + Mt)} * 100$$

$$SC = \frac{(Mt + Mc)}{(Mc + Mt + Mw)} * 100$$

$$W = \frac{Mw}{(Mc + Mt)} * 100$$

An example flow, Underground backfill

Generate model to be used as objective function/s

TABLE 8 Constructed multiple regression models for the prediction of UCS, yield stress, and cost.

Predictive model	Adjusted R^2
Cost = −15.74 + 2.33CD + 0.203SD	99.97
Yield stress = −9138.48–15.94CD + 124.17SD	86.81
UCS = EXP(0.185CD + 0.153SD + 0.00388T-6.7415)	79.91

An example flow, Underground backfill

We used multi-objective particle swarm optimization (MOPSO) algorithm in their work for finding solutions that were a mixture of designs of CPB meeting the considered objectives. It is not possible to find one optimal solution which can meet all objective functions, instead a set of optimal solutions, named Pareto, can be obtained.

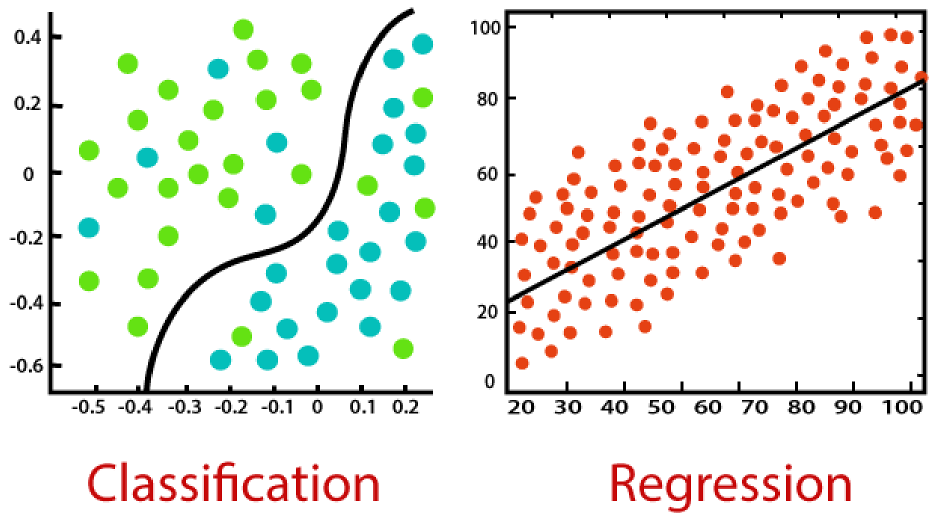
Batch No.	CD, %	SD, %	28 days UCS, kPa	Yield stress, Pa	Cement cost, AUD/m^3
C1_1	7.01	78.57	799.54	505.18	16.54
C1_2	6.70	78.91	796.16	552.68	15.89
C1_3	6.83	78.68	788.29	522.56	16.16
C1_4	6.53	78.99	780.64	565.18	15.51
C1_5	6.69	78.78	778.49	536.78	15.831
C1_6	6.81	78.62	777.83	514.97	16.10
C1_7	6.46	78.99	771.92	566.95	15.36
C1_8	6.71	78.64	765.55	519.64	15.86
C1_9	6.32	79.00	751.83	570.22	15.01

CPB should have UCS between 750 and 800kPa in 28days, yield stress should range from 500 to 800Pa, and maximum cement cost should be 17 AUD/m^3.

Batch No.	CD, %	SD, %	28 days UCS, kPa	Yield stress, Pa	Cement cost, AUD/m^3
C2_1	7.43	78.03	754.44	431.48	17.42
C2_2	7.48	78.25	788.11	459.02	17.58
C2_3	7.66	77.92	774.61	414.51	17.94
C2_4	7.34	78.21	763.66	456.17	17.25
C2_5	7.43	78.03	754.93	432.07	17.42
C2_6	7.73	77.87	778.15	406.88	18.08
C2_7	7.84	77.90	798.60	409.84	18.35

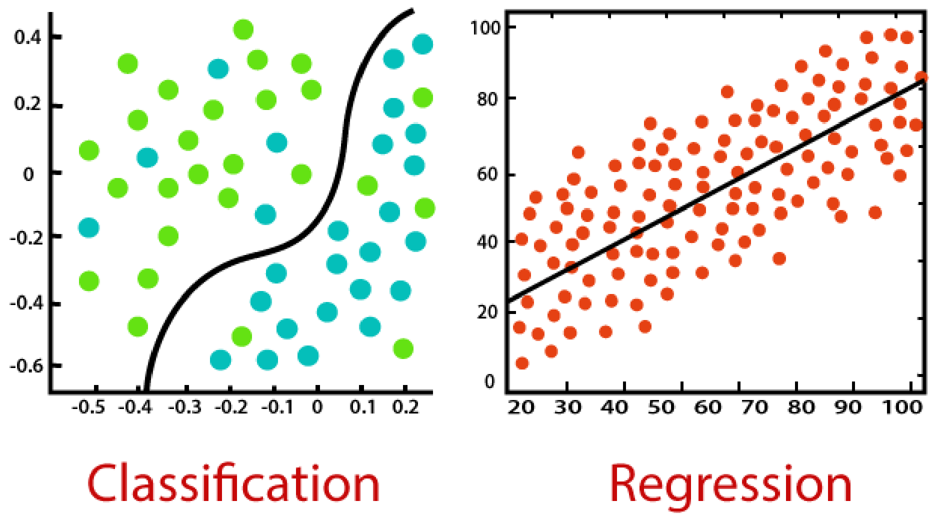
the production engineer was looking for the same UCS (750–800kPa) in 14 day shorter period than the first case and acceptable yield stress ranging from 400 to 500Pa.

Types of ML models



Classification involves the quantization of a set of input vectors into a predefined set of classes. Regression can be defined as a classification with a potentially vast number of classes, even theoretically infinite (Friedland 2024). The learning algorithm in this type of machine learning is trained on labeled data. The data is termed "labeled" because it comprises pairs—one part being the input, represented as a vector, and the other part being the desired output, which serves as a supervisory signal.
Examples are logistic regression, decision trees, SVM, KNN, RF.

Types of ML models



Regression modeling is a machine learning technique used to predict a continuous numeric outcome based on input variables. The model learns the relationship between these variables from historical data to make accurate predictions on new data.
Popular regression algorithms include linear regression, polynomial regression, and neural networks, RF, SVM.

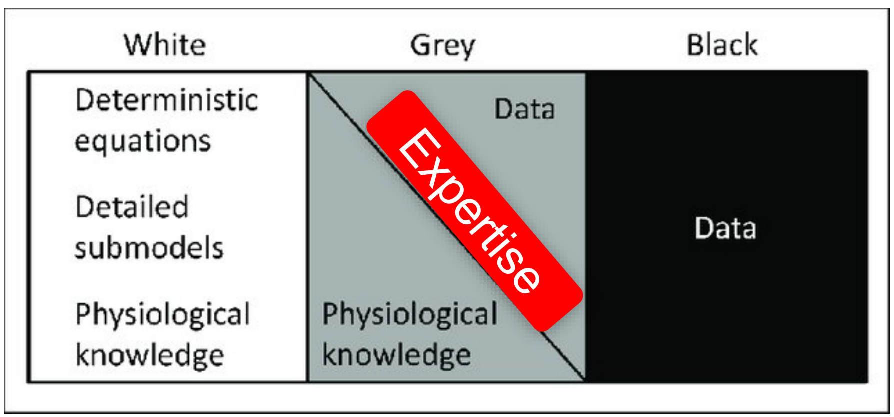
Types of ML models

Machine learning technique comprises models with varying structures, categorized as white-box, black-box and grey-box models.

Which box ?

modeling is a machine lear

Results easy to understand



Models hard to explain from a mathematical point of view (but often "better")

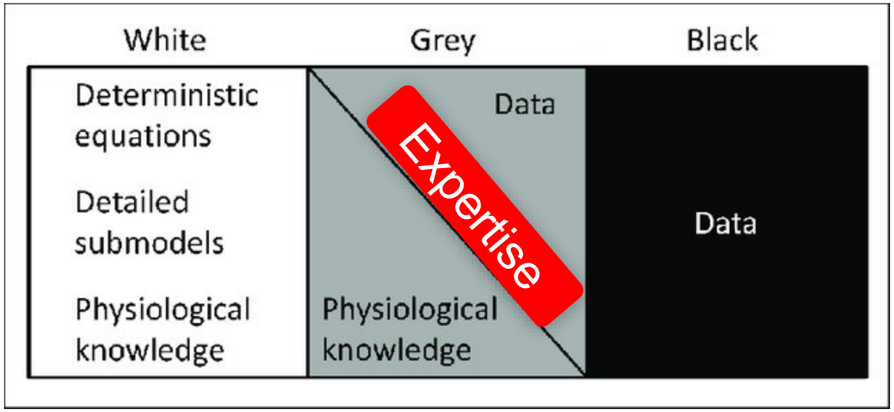
Types of ML models

Machine learning technique comprises models with varying structures, categorized as white-box, black-box and grey-box models.

Which box ?

modeling is a machine lear

Results easy to understand



Models hard to explain from a mathematical point of view (but often “better”)

White-box models pertain to transparent machine learning processes that generate accountable predictions. These models emphasize processes, enabling users to interpret the models and determine the relative importance of the corresponding predictor variables. The transparency allows users to identify potential errors when the model coefficients contradict established physics-based knowledge (Fung et al. 2021). Examples of white-box models include linear regression and decision trees algorithms.

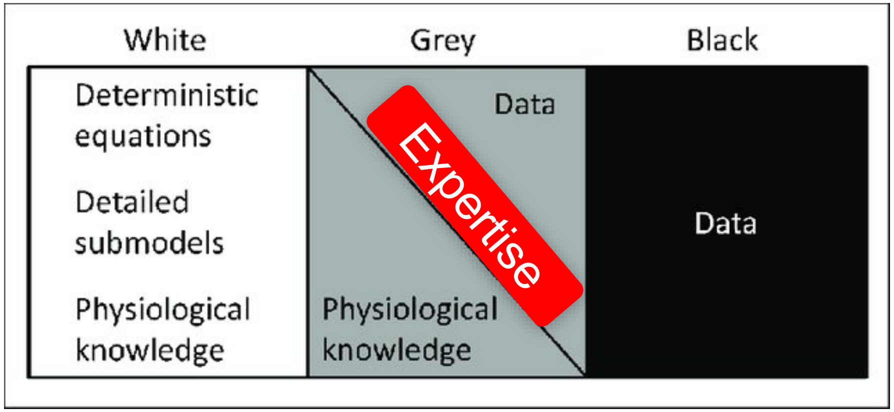
Types of ML models

Machine learning technique comprises models with varying structures, categorized as white-box, black-box and grey-box models.

Which box ?

modeling is a machine lear

Results easy to understand



Models hard to explain from a mathematical point of view (but often “better”)

Black-box models are statistical or machine learning models where the parameters lack physical meaning. These models are often appealing due to their flexibility and ease of construction. They utilize operational data to train their internal parameters, but as the name implies, black-box models typically act as observers since their estimated parameters may have minimal relevance to the actual physical process (Katipamula and Brambley 2005). Black-box models are exemplified by artificial neural networks and XGBOOST algorithm.

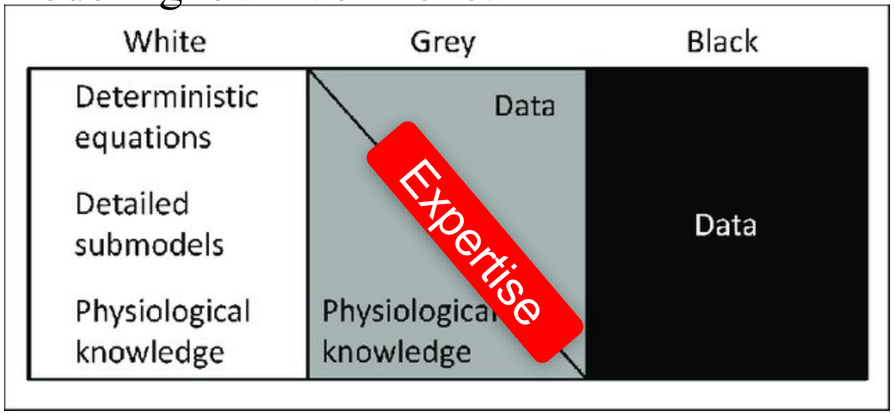
Types of ML models

Machine learning technique comprises models with varying structures, categorized as white-box, black-box and grey-box models.

Which box ?

Results easy to understand

modeling is a machine lear



Models hard to explain from a mathematical point of view (but often "better")

Grey-box models are analytical models that are somewhat based on principles, with parameters that can still be linked to the physical response of the process. Compared to white-box models, they are faster to compute and easier to calibrate and construct. Compared to black-box models, they are more robust and suitable for parameter estimation. However, creating grey-box models requires expert knowledge and extensive measured data to train their parameters. Additionally, they may be less accurate than both black-box and white-box models (Katipamula and Brambley 2005). Grey-box models can be represented by adaptive neuro-fuzzy inference system (ANFIS) and random forest algorithms.

What is ML?

Definition: To understand the role of data-driven modeling in addressing complex geotechnical challenges.

Data-driven modeling refers to the process of using data analytics and machine learning techniques to derive insights and predictions from large datasets. Unlike traditional models that rely on theoretical assumptions, data-driven models learn patterns directly from the data.

Conventional ML tools vs ML modelling

FEM

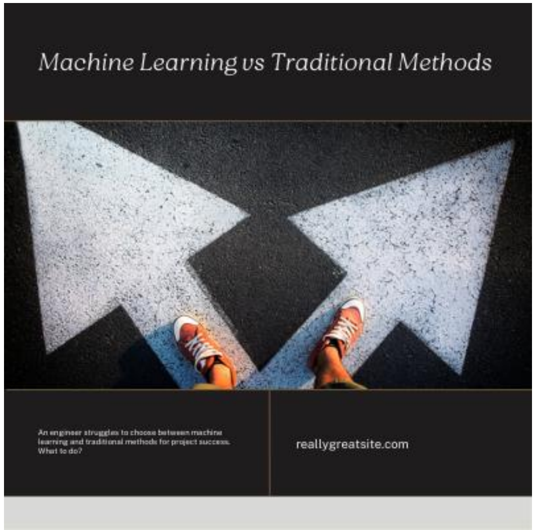
Cons!

- Material properties:
  CHILI vs DIANE?

- Boundary conditions:
  - Realistic enough ?

- Mesh dependency:
  - Good enough quality ?

- Governing equations:
  - Governing equations based on theoretical assumptions. Are they align well with the actual behaviour?



ML

- Cons!
  - Data Dependence: Do you have good enough data?

- Black Box Nature:
  - Can you see inside?

- Static Training:
  - Will the conditions be the same always?

Conventional ML tools vs ML modelling

Mentimeter

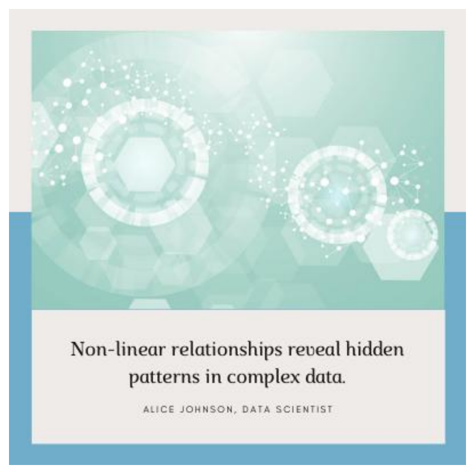
Conventional ML tools vs ML modelling

Both conventional or traditional modelling (CM) and ML have their own sets of assumptions and limitations. However, it is essential to acknowledge that CM's reliability is not inherent; it is based on the validity of its assumptions, which can be as prone to errors as the data-dependency of ML. In fields where data is abundant, or where traditional assumptions about material behavior are not valid/well known, ML can be a powerful alternative or complement to CM.

By recognizing the strengths and limitations of both methods, we can make better-informed decisions about which approach (or combination of approaches) to use for a given problem.

How Data-Driven Models Uncover Hidden Patterns

- **Pattern Recognition**: Machine learning techniques (like neural networks and decision trees) can detect intricate patterns by analyzing large volumes of data.
- **Non-linear Relationships**: Unlike traditional models, data-driven approaches can capture non-linear relationships between variables (e.g., rock strength vs. depth, moisture content, or stress conditions).
- **Multivariate Analysis**: By simultaneously considering multiple input variables, these models can uncover interactions that might be missed by simpler empirical methods.



ML in Georesources

Machine learning is a powerful tool in geosciences because it excels at handling complex, variable, and uncertain data—typical characteristics of geological systems.

Unlike traditional models that rely on fixed equations and assumptions, ML can identify patterns, relationships, and insights directly from large, diverse datasets. This flexibility enables geoscientists to make more accurate predictions, optimize resource exploration, and improve decision-making under uncertainty.

ML can adapt to new information, continuously improving its predictions as more data becomes available, making it especially valuable in dynamic environments like subsurface exploration.

ML in Georesources

Some well known examples are
Rock property assessment
Stability analysis
Seismic data analysis
.
.
.
.
What else..

Here do we get the data?

Data Source	Data Collected	Applications	Data Type	Challenges/Advantages
Borehole and Core Sample Data	Rock/soil samples, core logs, geotechnical properties (e.g., UCS, RQD)	Stratigraphy, rock quality, rock mass classification (e.g., RMR, Q-system)	Qualitative & Quantitative	Limited spatial coverage, high variability, data gaps
Measurement While Drilling (MWD)	Drilling parameters: penetration rate, torque, rotation speed, bit pressure	Estimating rock strength, lithology changes, drilling optimization	Continuous & Quantitative	Continuous data, better subsurface understanding
Geophysical Surveys	Seismic waves, GPR, resistivity, magnetic data	Mapping subsurface structures, fault detection, groundwater exploration	Quantitative & Spatial	Expert interpretation needed, affected by subsurface noise
In-Situ Testing and Monitoring	SPT, CPT, pressuremeter tests, inclinometers, piezometers, strain gauges	Assessing soil strength, groundwater levels, slope stability monitoring	Quantitative & Time-Series	Site-specific data, reflects actual conditions
Laboratory Testing	Soil/rock properties: density, permeability, shear strength, compressibility	Foundation design, slope stability, geotechnical property analysis	Quantitative	Sample disturbance, may not fully capture in-situ conditions
Remote Sensing & Satellite Data	Satellite imagery, LiDAR scans, aerial photos	Monitoring surface deformation, landslide detection, subsidence analysis	Quantitative & Spatial	Large-scale monitoring, remote site accessibility
Historical & Archival Data	Historical records, site investigation reports, case studies	Risk assessment, geotechnical model refinement, predictive accuracy	Qualitative & Historical	Data quality varies, may not align with current methods
Real-Time Monitoring Systems	Sensor readings: deformation, pore pressure, strain	Early warning systems, real-time stability monitoring, construction safety	Continuous & Time-Series	Proactive risk management, real-time decision-making

Here do we get the data?

Mentimeter

ML in mining engineering

Machine learning has become increasingly valuable in mining engineering due to its ability to optimize complex processes, reduce operational costs, and enhance safety.

Mining operations involve handling large volumes of data from various sources, such as drilling, blasting, and equipment sensors.

ML can efficiently analyze this data to identify patterns, optimize resource extraction, predict equipment failures, and improve decision-making.

Unlike traditional methods, ML models can continuously adapt to new information, allowing for real-time adjustments in dynamic mining environments. This leads to better resource management, reduced downtime, and increased productivity..

More examples

4.

9

1

ues for
rative study of
vork and

Erkan Topal

Contents xi

ork (LSTM)

149
151
151
153

Round table

Please prepare and upload a single page document summarizing,

What is your (thesis) topic

Can you define the problem?
What are the variables?
What will be the objective/s?

Do they need to be optimized.

What are input parameters for your problem,

What is next ?

- How to Select the Appropriate Modeling Technique
  - Case studies demonstrating model selection: Regression vs. classification.
  - Discussion: The flexibility of modeling approaches—how it varies from project to project and person to person.
  - Examples of model selection based on my research publications.
- Introduction to Hybrid Modeling
  - Gray Box Modeling: Combining physical models with data-driven approaches.
    - Fuzzy Inference Systems (FIS): Applications in geotechnical parameter estimation.
    - Adaptive Neuro-Fuzzy Inference Systems (ANFIS): Using ANFIS for complex geotechnical predictions with examples from your publications.

What is next ?

- Examples of Black Box Modeling Techniques
  - Random Forests (RF): Applications in classification and regression tasks in geotechnical projects.
  - Gaussian Process Regression (GPR): Understanding its use for uncertainty quantification.
  - ……
  - ……