BAUERBERG KLEIN

TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 01

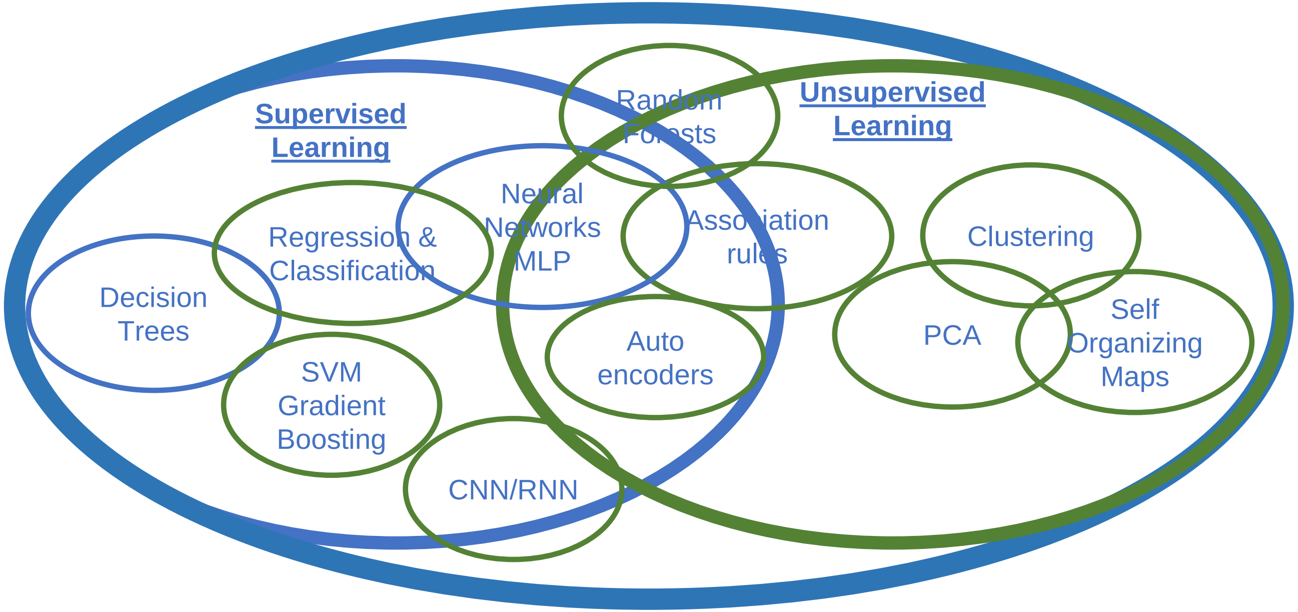
Introduction: Data-driven Geophysical and Petrophysical modeling using AI techniques

LEARNING OBJECTIVES

- GOAL01: Artificial Intelligence Terminology
- GOAL02: Fundamentals of Soft Computing/Statistics
- GOAL03: Machine and Deep Learning Techniques in Upstream Exploration and Production (E&P)

INTRODUCTION

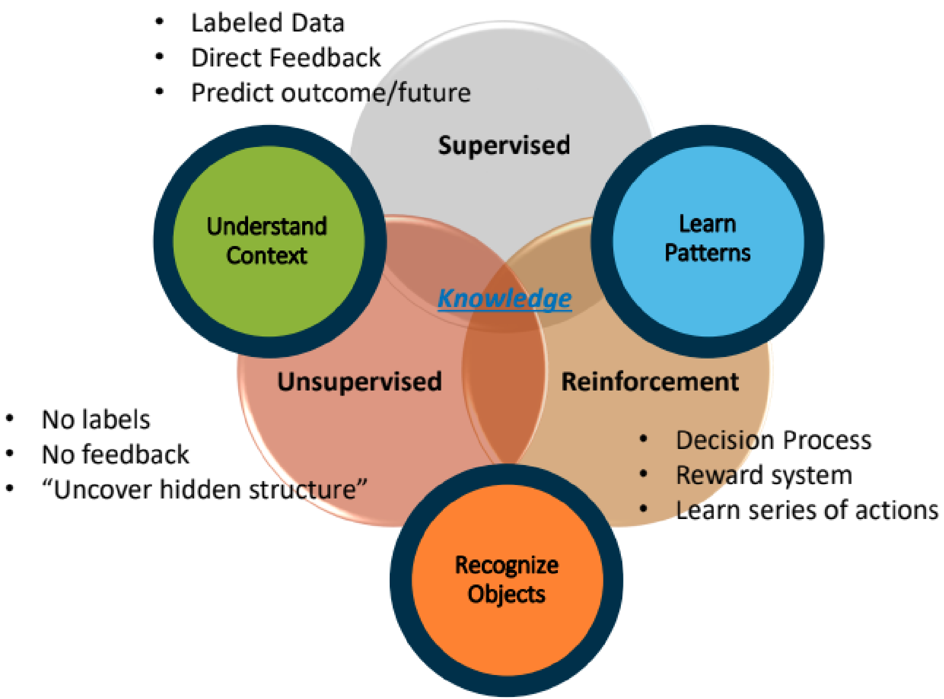
Terms most often heard



Artificial Intelligence Terminology

INTRODUCTION

Artificial Intelligence = Knowledge



INTRODUCTION

Essentials of Statistics

What are statistics?
Make sense of your data
Identify trends and correlations
Find useful patterns
Propose hypotheses worth modeling

Descriptive Statistics:
Descriptive statistics are brief descriptive coefficients that summarize a given data set

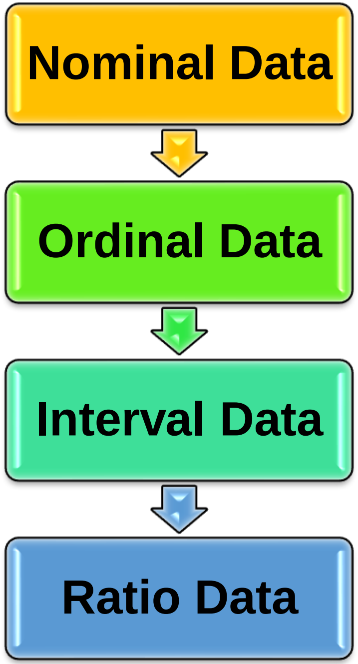
Inferential Statistics:
Help you come to conclusions and make predictions based on your data

INTRODUCTION

Essentials of Statistics

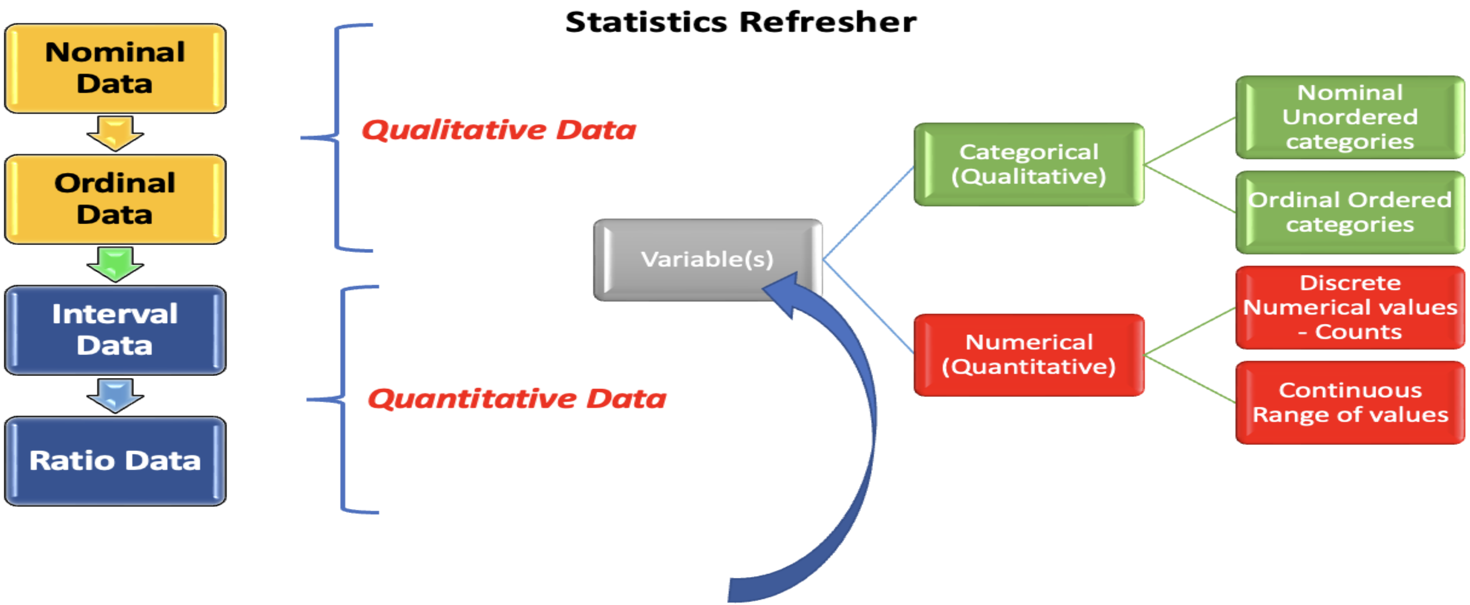
Data is usually classified into one of four different levels of measurement:

- **nominal** (a.k.a. categorical or discrete)
- **ordinal** (rank order)
- **interval** (continuous)
- **ratio**



Qualitative Data	Descriptive/Categorical
Quantitative Data	Numeric Connotation

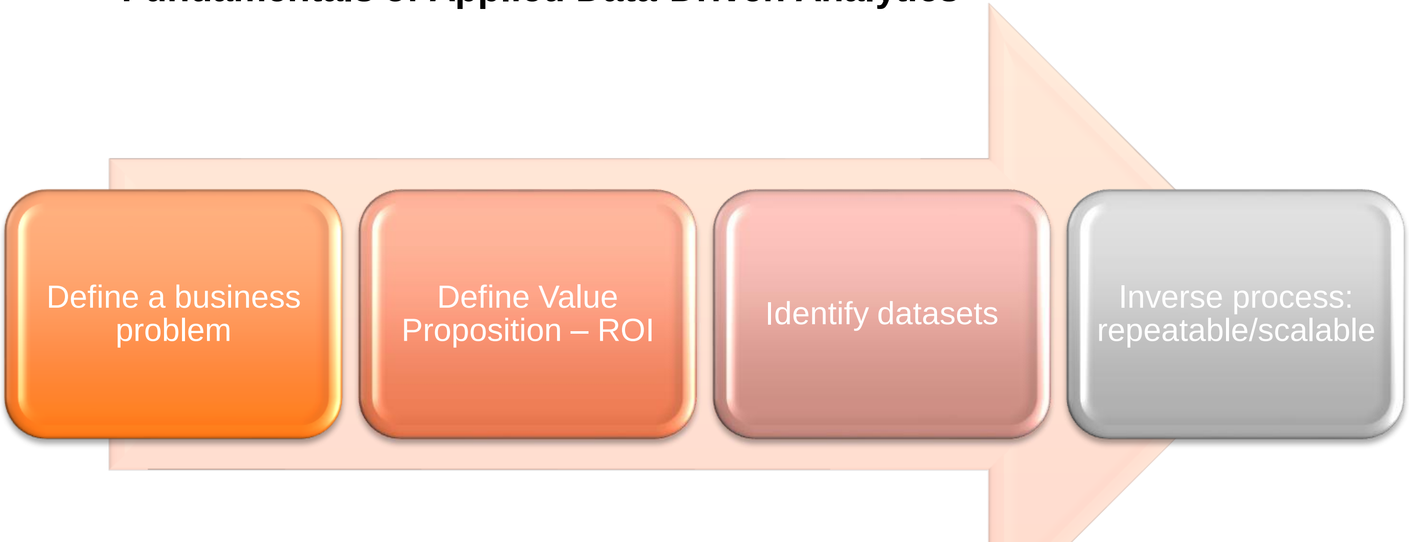
INTRODUCTION



Dependent and **Independent** variables are the components in mathematical modeling, statistical modeling and experimental sciences.

Data Preparation for AI

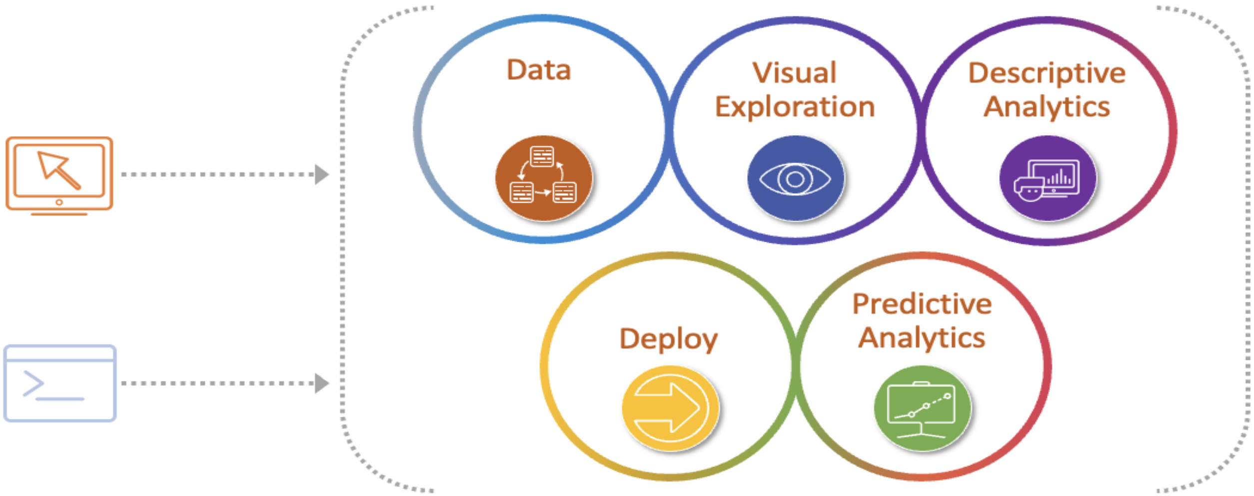
Fundamentals of Applied Data-Driven Analytics



Dependent and **Independent** variables are the components in mathematical modeling, statistical modeling and experimental sciences.

Data Preparation for AI

Fundamentals of Applied Data-Driven Analytics



INTRODUCTION

Machine & Deep Learning Techniques

	Overview	Process	Subtypes	Examples
Supervised Learning	Majority of algorithms. Machine is trained using **well-labeled data**; inputs and outputs are matched.	Mapping function takes inputs and matches to outputs, creating a target function.	Classification, Regression	Linear regression, Random forest, SVM.
Unsupervised Learning	**Unlabeled data** (inputs only) is analyzed. Learning happens without supervision.	Inputs are used to create a model of the data.	Clustering, Association.	PCA, k-Means, Hierarchical clustering.
Semi supervised	**Some data is labeled**, some not. Goal: better results than labeled data alone. Good for real world data.	Combination of above processes.	All the above.	Self training, Mixture models, Semi-supervised SVM

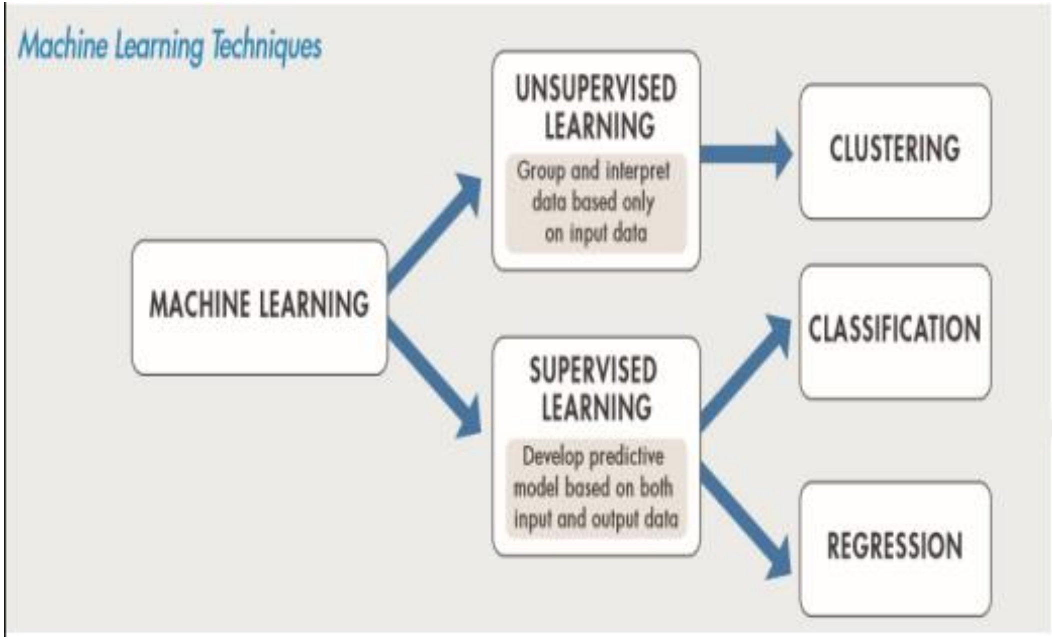
INTRODUCTION

What is Machine Learning?

<u>Supervised Learning</u>
The aim of supervised machine learning is to build a model that makes predictions based on evidence in the presence of uncertainty.

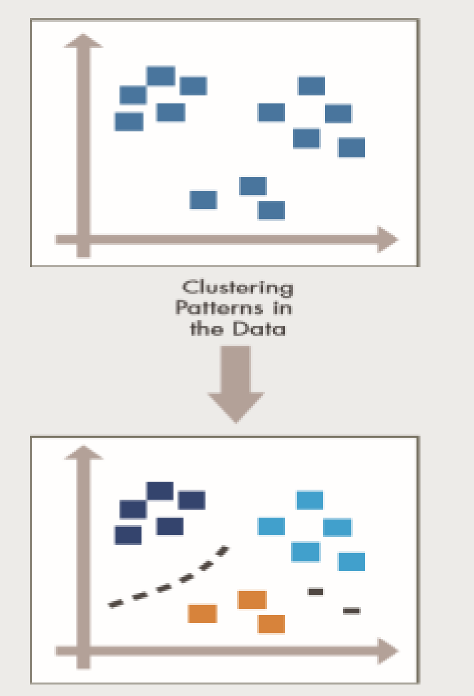
• Classification techniques predict discrete responses, for example, whether a reservoir has bypassed pay. Classification models classify input data. Typical applications include seismic imaging, well-log pattern recognition, and facies classification.

• Regression techniques predict continuous responses, for example, changes in temperature or pressure in a producing well. Typical applications include production forecasting.



INTRODUCTION

What is Machine Learning?



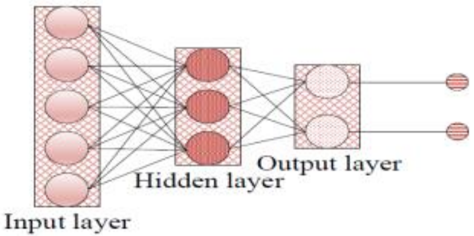
Unsupervised Learning

Finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses.

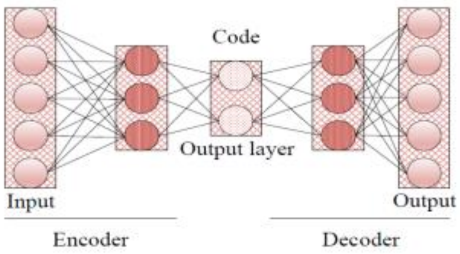
Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Applications for clustering include reservoir characterization, field re-engineering, and DHI object recognition in seismic wavelet data.

INTRODUCTION

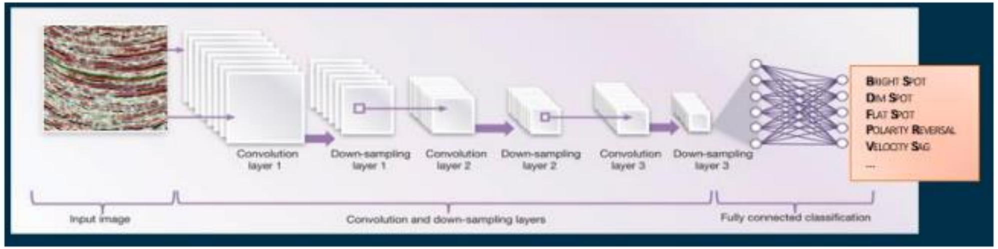
Deep Learning Techniques



Recurrent Neural Network



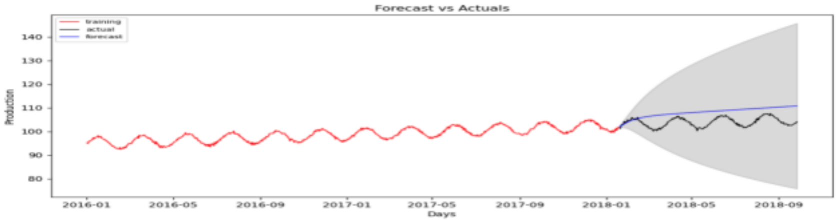
DAE Architecture

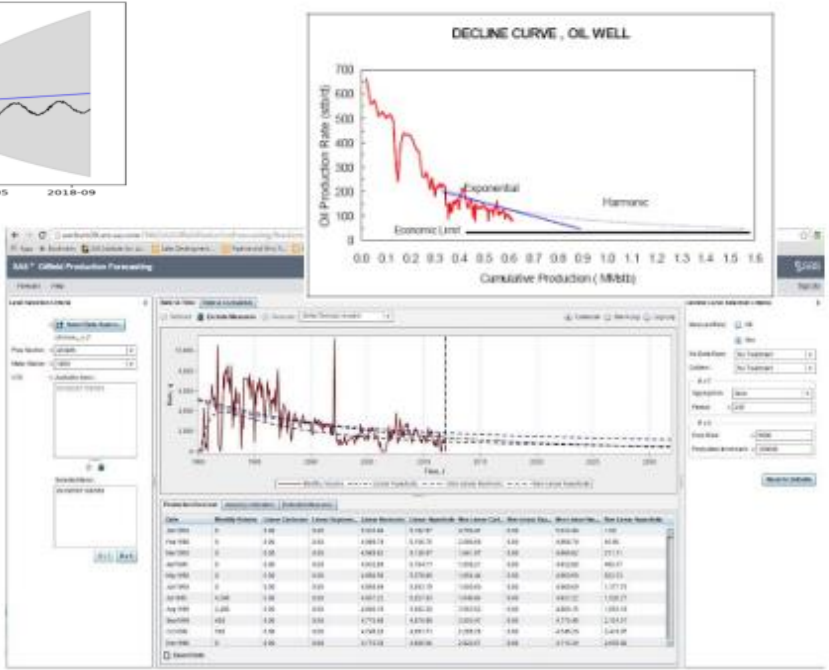


Convolutional Neural Network

INTRODUCTION

Time Series Analysis





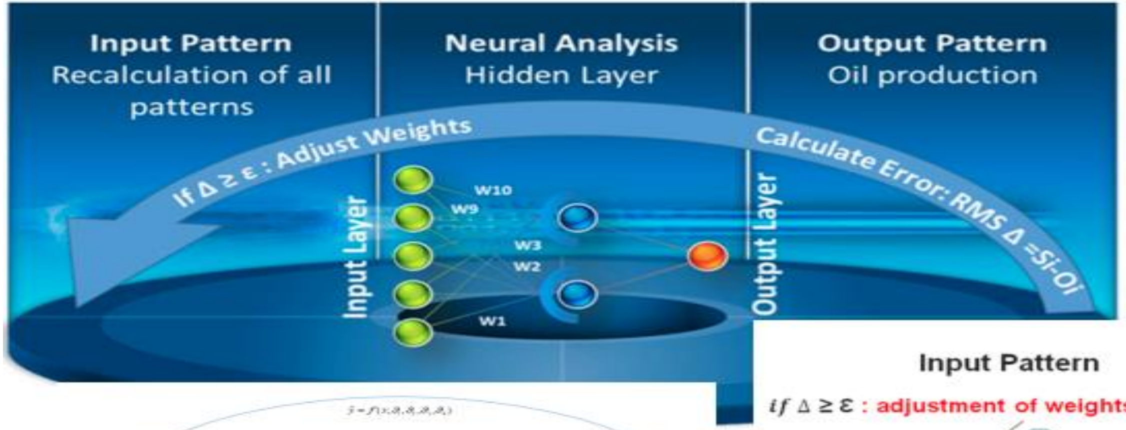
<u>What is Forecasting and Optimization?</u>

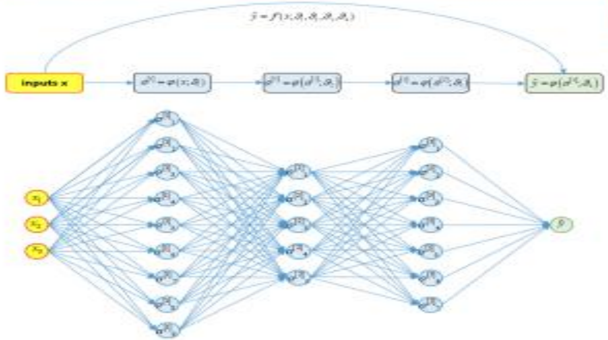
Predicting future needs for a product or service, while **Optimization** is maximizing results within a set of constraints.

INTRODUCTION

Predictive Models

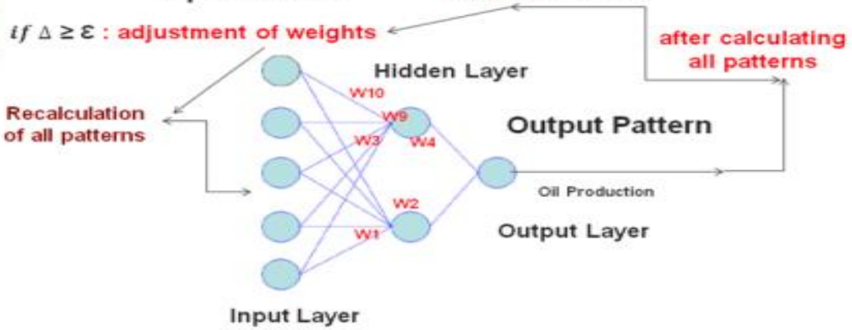
Artificial Neural Networks





Ni is the number of input neurons, *No* the number of output neurons, *Ns* the number of samples in the training data, and **α** represents a scaling factor that is usually between 2 and 10. We can calculate 8 different numbers to feed into our validation procedure and find the optimal model, based on the resulting validation loss.

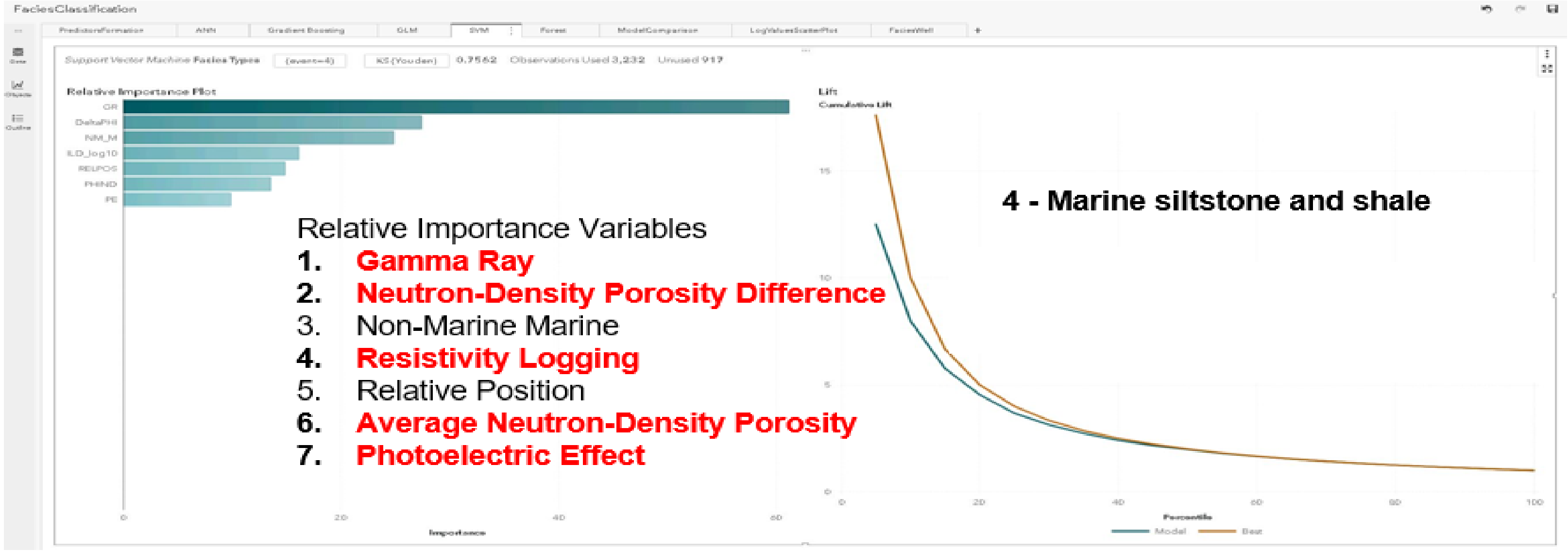
$$N_h = \frac{N_s}{(\alpha * (N_i + N_o))}$$



INTRODUCTION

Predictive Models

<u>Support Vector Machine</u>
The SVM is a supervised machine learning algorithm that constructs a hyperplane or set of hyperplanes to distinguish between instances of different classes.



INTRODUCTION

Predictive Models

Random Forest

A forest is an **ensemble** model that contains a specific number of decision trees. To ensure that a forest does not overfit the data, two key steps are taken. First, each tree in the forest is built on a different sample of the training data. Second, when splitting each node, a set of candidate inputs for the split are selected at random, and the best split is selected from those. Other than these two steps, the trees in a forest are trained like standard trees.

Module 02

Exploratory Data Analysis: Upstream Data Exploration and Explanation

MODULE 02

Exploratory Data Analysis (EDA) is crucial in analyzing upstream Oil and Gas (O&G) data. It involves examining and understanding the data's characteristics, patterns, and relationships before applying formal statistical or machine-learning techniques. EDA helps uncover insights, identify data quality issues, and formulate hypotheses for further analysis. We shall explain the typical steps involved in conducting EDA for upstream O&G data:

1. Data Collection and Data Cleaning
2. Data Visualization
3. Descriptive Statistics
4. Feature Engineering
5. Correlation and Spatial Analysis
6. Hypotheses Generation

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 02

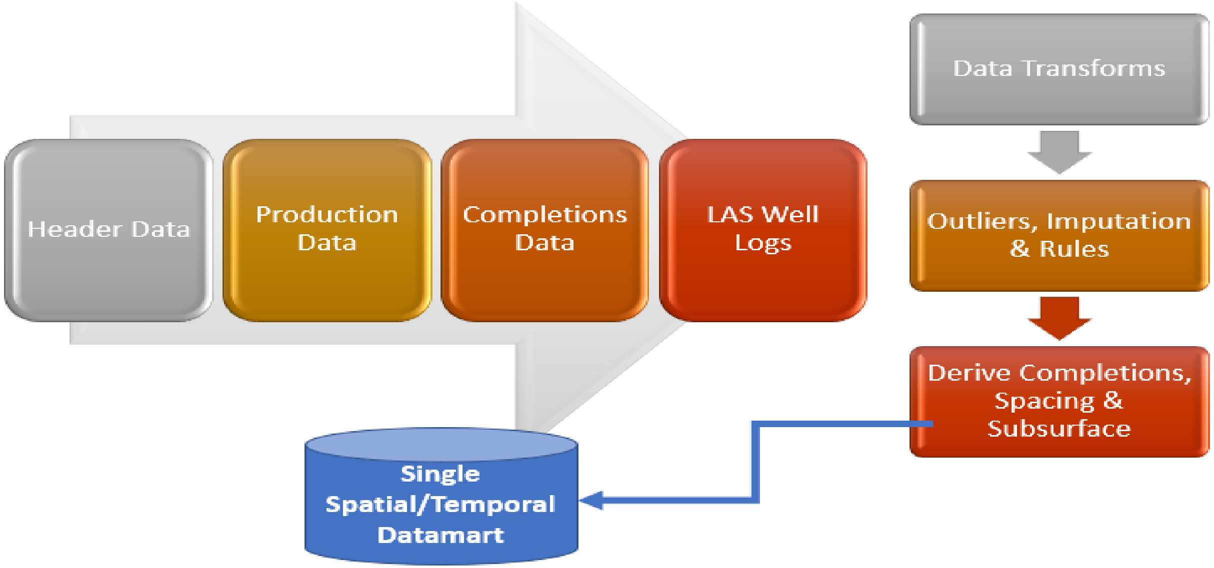
Exploratory Data Analysis: Upstream Data Exploration and Explanation

LEARNING OBJECTIVES

- ➢ GOAL01: Data Management and Data Cleaning Steps
- ➢ GOAL02: Upstream Exploratory Data Analysis (EDA) using Tukey Diagrams
- ➢ GOAL03: Descriptive Modeling in Upstream Exploration and Production (E&P)
- ➢ GOAL04: Process & Methodology for E&P Model Development

Exploratory Data Analysis

UPSTREAM DATA MANAGEMENT



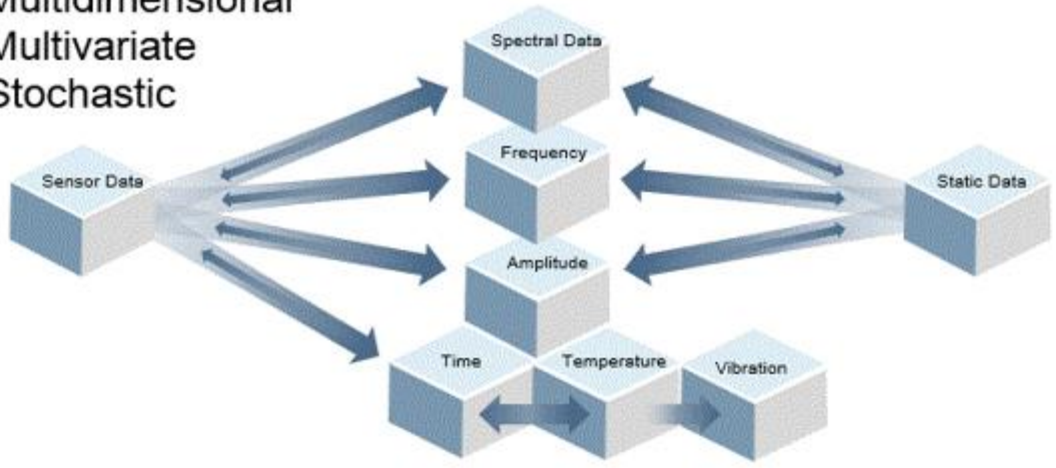
Exploratory Data Analysis

DATA CLEANING STEPS

Removing unwanted observations	• Duplicate/ redundant or irrelevant values deletion .
Missing Data handling	• Fixing issue of unknown missing values
Structural error solving	• Fixing problems with mislabeled classes, typesin names of features, same attribute with different name etc.
Outliers Management	• Unwanted values which are not fiting in datasets.

Exploratory Data Analysis

- Multivariant
- Multidimensional
- Multivariate
- Stochastic



Exploratory Data Analysis

Identify Trends, Correlations and Signatures in Patterns

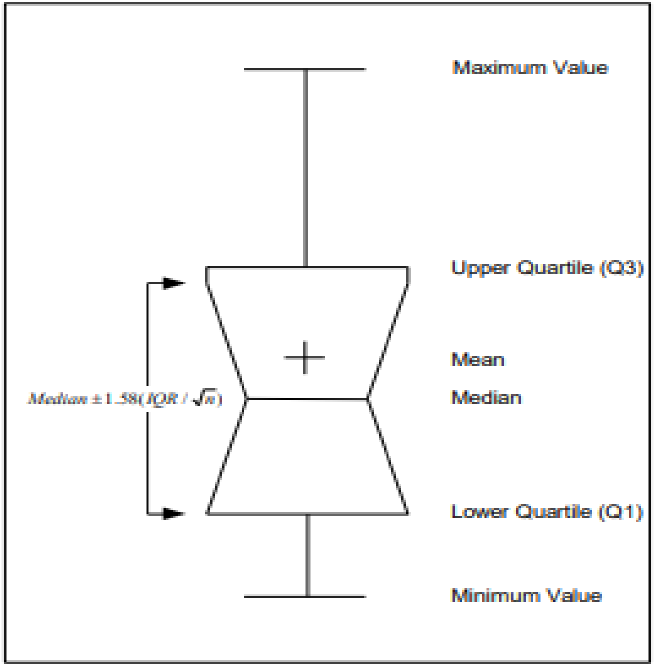
Explore the data to find trends, correlations, and hidden relationships. The goal is to find patterns or signatures in your data to use them to predict future events in a time series or across spatial data.

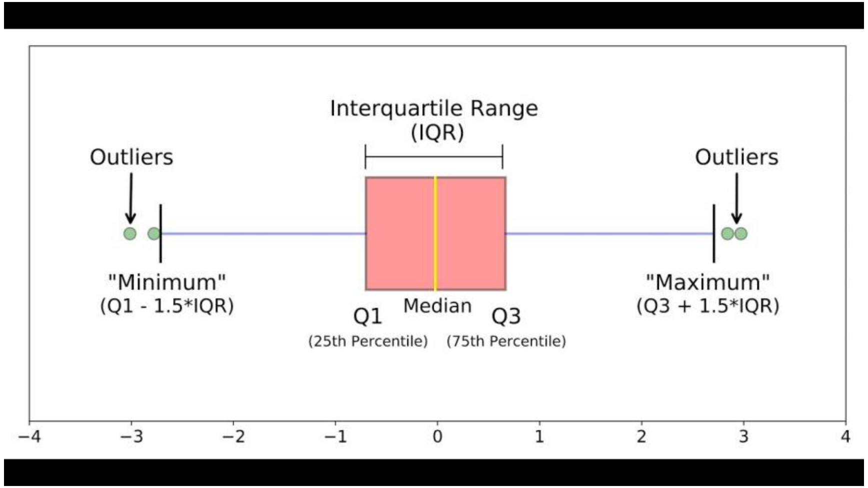
The Tukey diagrams give you a visual appreciation of the data in univariate and bivariate analysis.

Exploratory Data Analysis

Box Whisker Plots

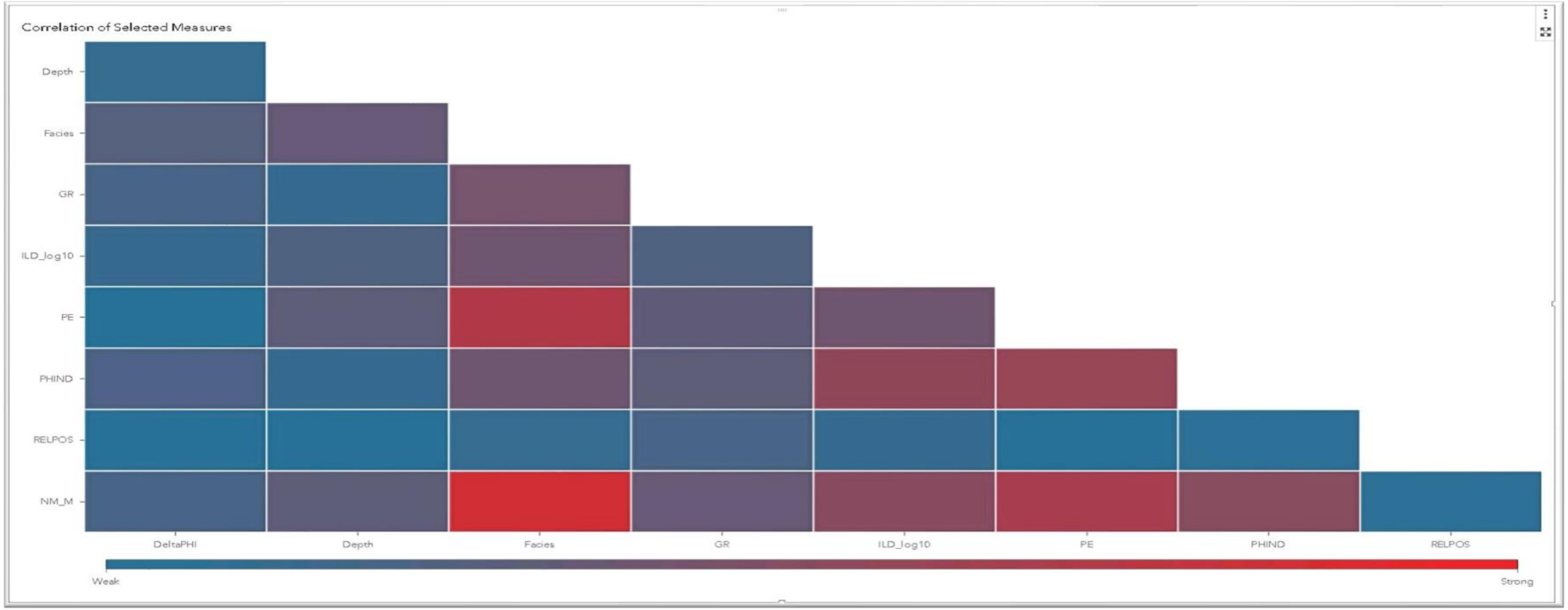
$$\text{Median } \pm 1.58\left(\frac{\text{IQR}}{\sqrt{n}}\right)$$





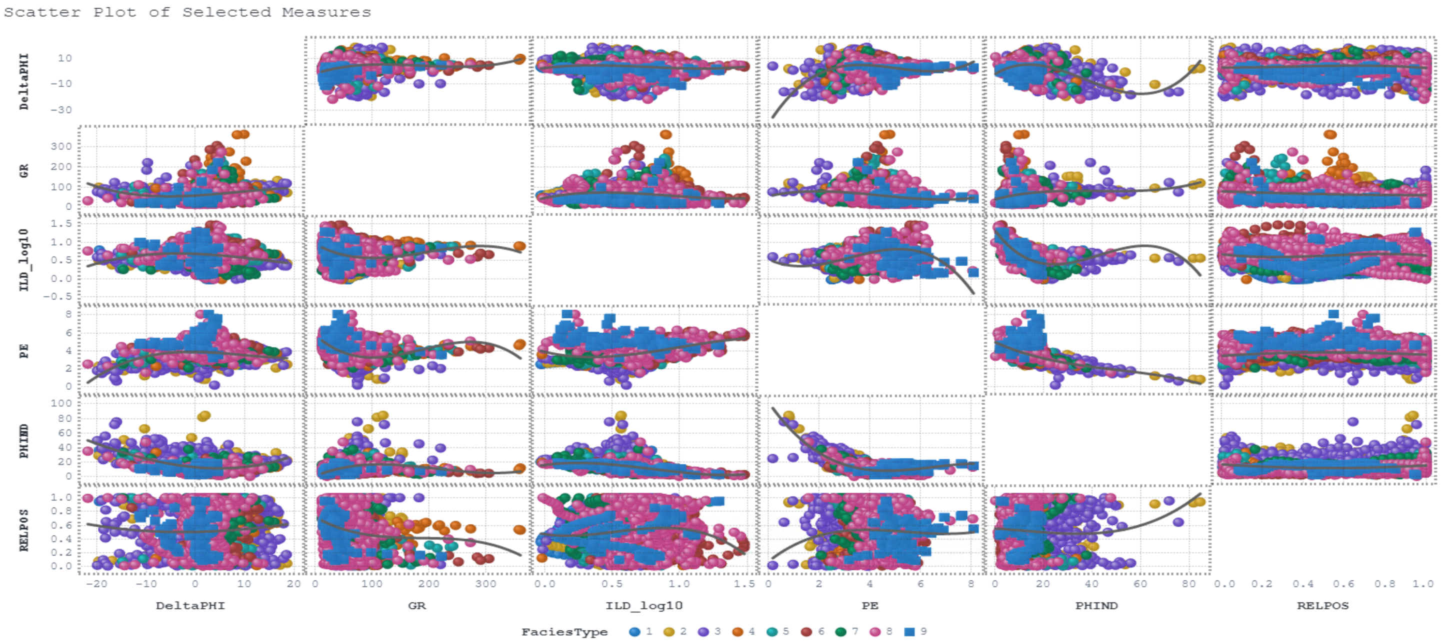
Exploratory Data Analysis

Correlation Matrices



Exploratory Data Analysis

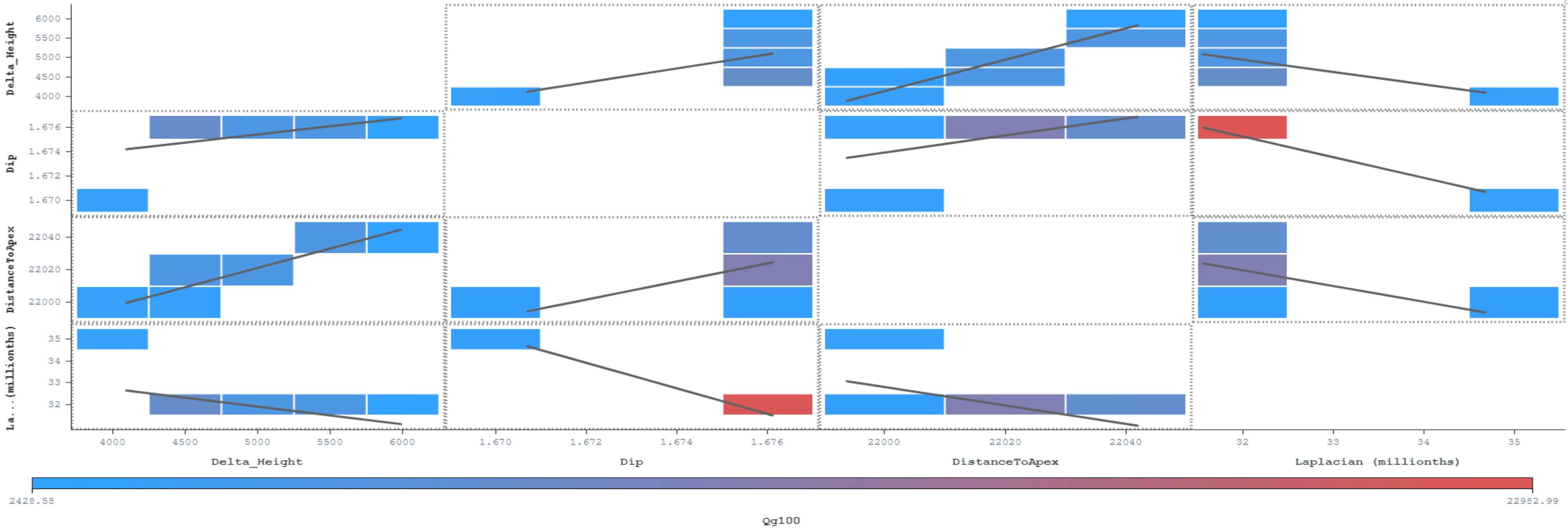
Scatter Plots



Exploratory Data Analysis

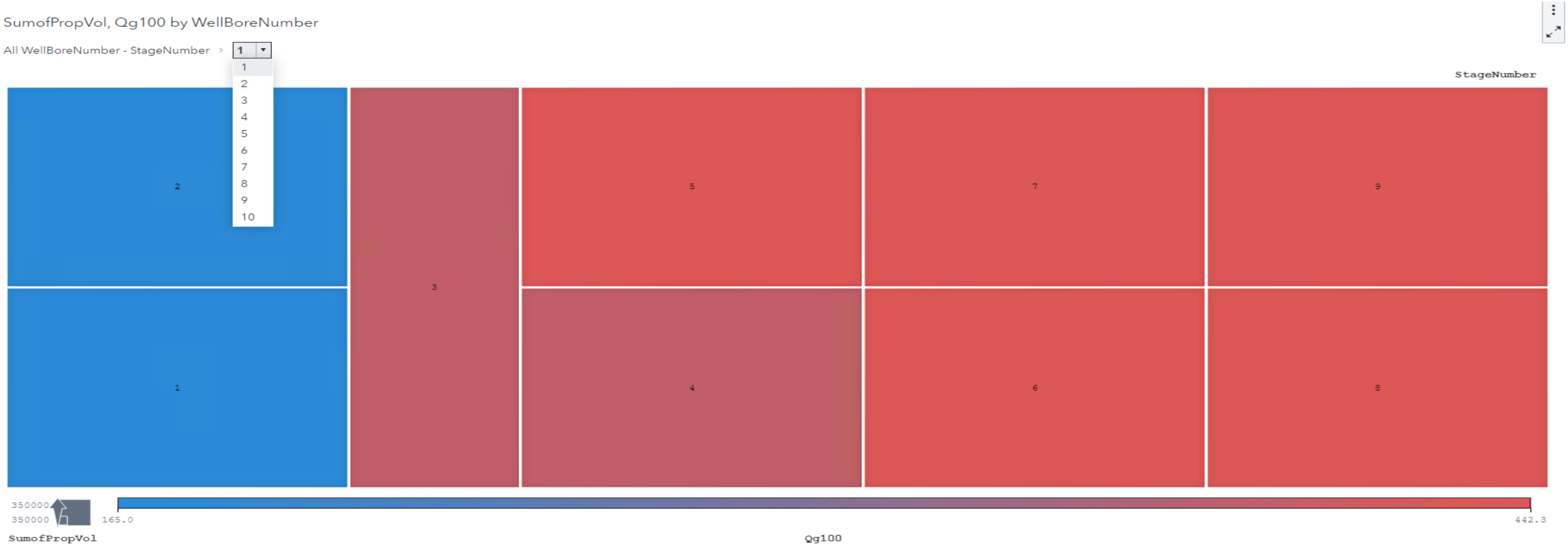
Heat Maps

Qg100 by Delta_Height, Dip, DistanceToApex, Laplacian



Exploratory Data Analysis

Tree Maps



Exploratory Data Analysis

Descriptive Modeling

Descriptive modeling techniques cover two major areas:

1. Clustering
2. Associations
3. Classification

The objective of clustering or segmenting your data is to place objects into groups or clusters suggested by the data such that objects in each cluster tend to be like each other in some sense and objects in different clusters tend to be dissimilar

Clustering Examples:

Upstream: Well Characteristics (Operational – Completion strategies/Petrophysical
Midstream: Pipeline segments, Cathodic Protection Stations, Pressure drops
Downstream: Amine Towers (Good & Bad process efficiency), Pump Lifetimes

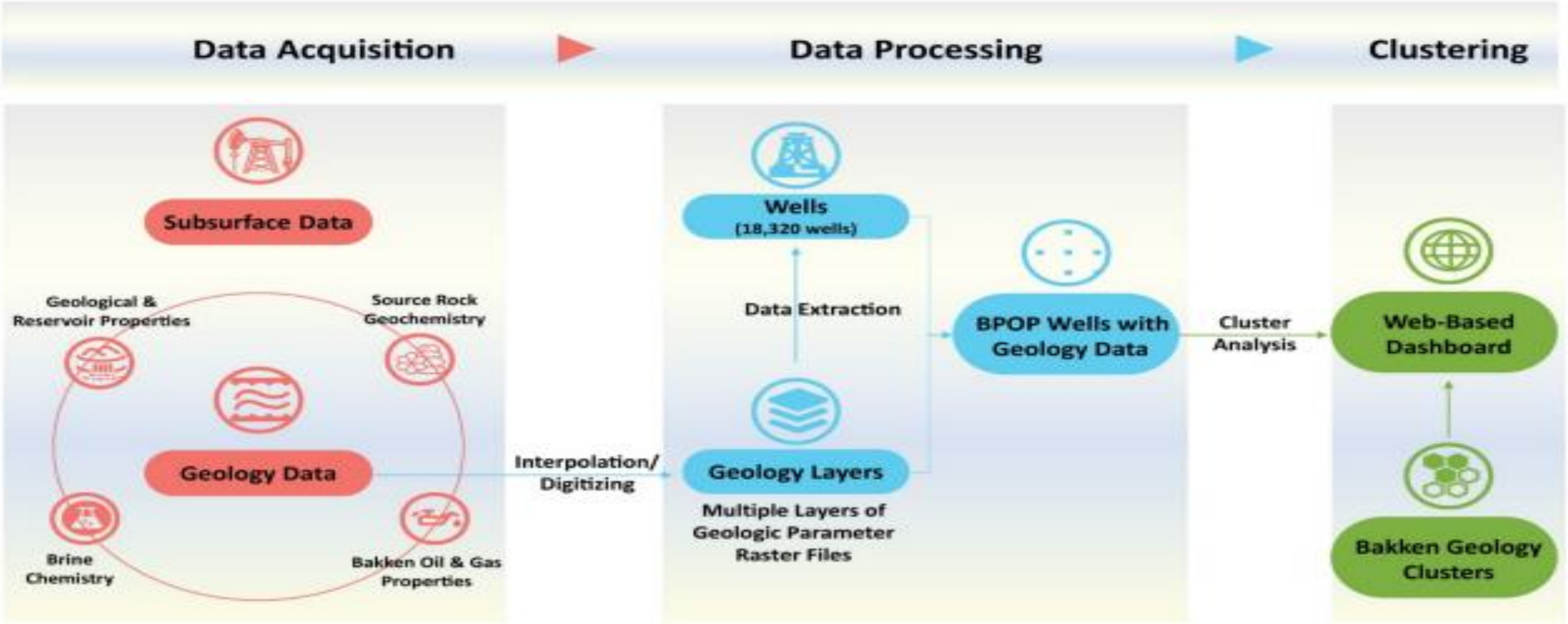
Exploratory Data Analysis

Clustering

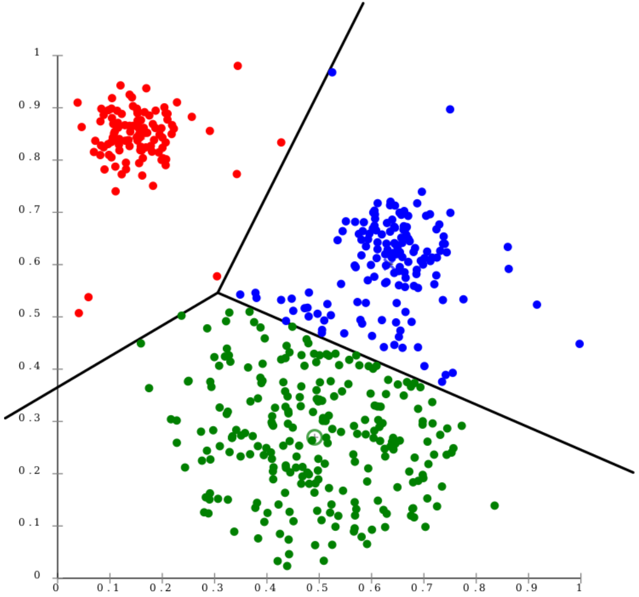
- Cluster_Analysis
- Distance Measures (Metrics)
- Evaluating Clustering
- Number of Clusters
- k-means Algorithm
- Hierarchical Clustering
- Profiling Clusters

Exploratory Data Analysis

Cluster Analysis to Optimize Completion Strategies in an Unconventional Reservoir

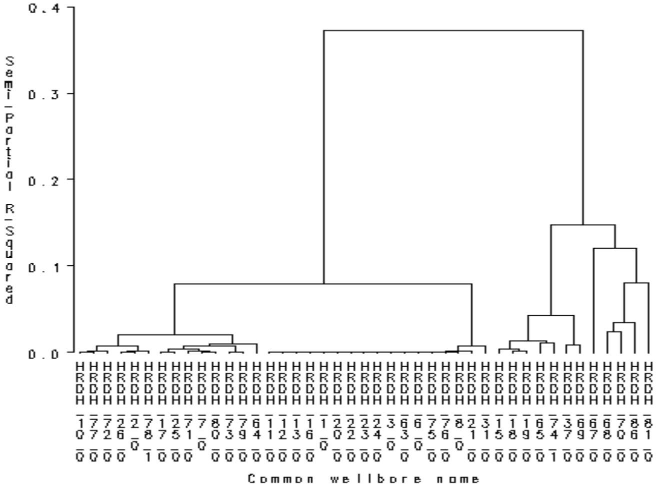


Exploratory Data Analysis



Data Processing

1. K-means Clustering Algorithm
2. Hierarchical Clustering Algorithm
3. Model-Based Clustering Algorithm

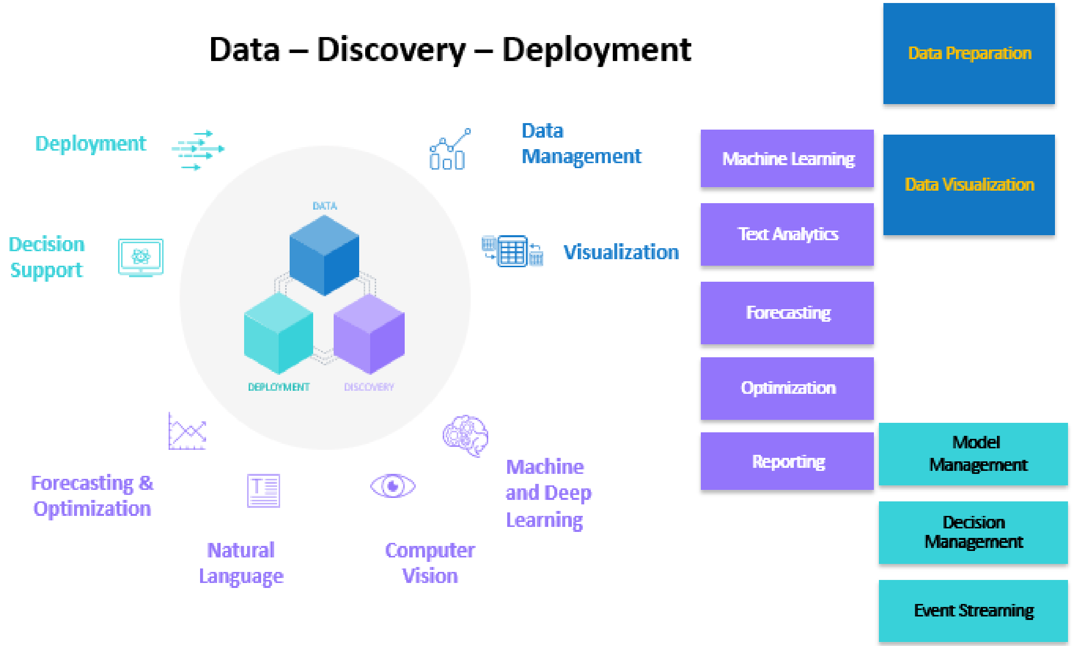


Exploratory Data Analysis

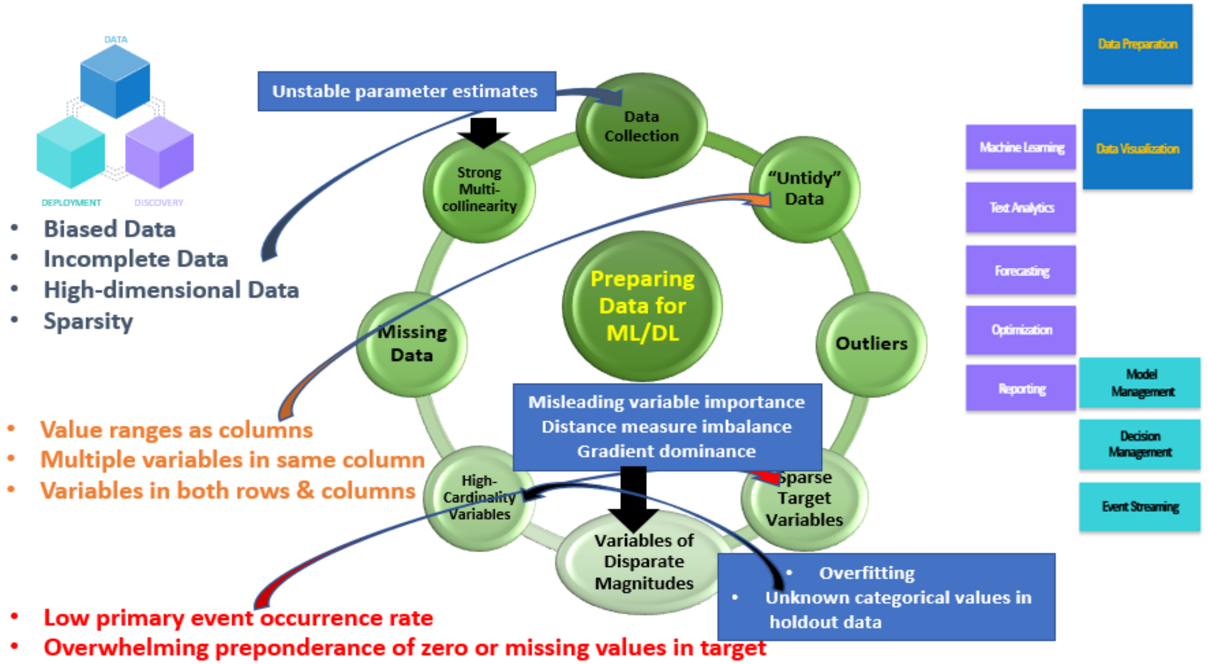
SEMMA Process

- Sample
- Explore
- Modify
- Model
- Assess

Exploratory Data Analysis



Exploratory Data Analysis



Module 03

Data Preparation for AI: Upstream Data Augmentation and Feature Engineering

MODULE 03

Data Preparation:

1.Data Collection: Gather the relevant data from various sources, such as well logs, production data, geological surveys, and reservoir engineering reports.

2.Data Cleaning: Remove or handle missing values, outliers, and inconsistencies in the dataset. This may involve imputation techniques, filtering, or deleting problematic data points.

Data Augmentation (GAI):

1.Synthetic Data Generation: Generate additional data points using techniques like oversampling, undersampling, or SMOTE (Synthetic Minority Over-sampling Technique) to balance imbalanced classes or increase the diversity of the dataset.

2.Time-Series Augmentation: Create variations of the original time-series data by introducing noise, time shifting, or resampling to capture different scenarios or increase the dataset size.

Feature Engineering:

1.Domain Knowledge: Leverage subject matter expertise to identify relevant features based on geological, geophysical, and engineering insights. This may involve extracting key attributes, engineering composite features, or creating derived variables.

2.Dimensionality reduction: Apply techniques such as Principal Component Analysis (PCA) or feature selection algorithms to reduce the number of features while retaining the most informative ones.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 03

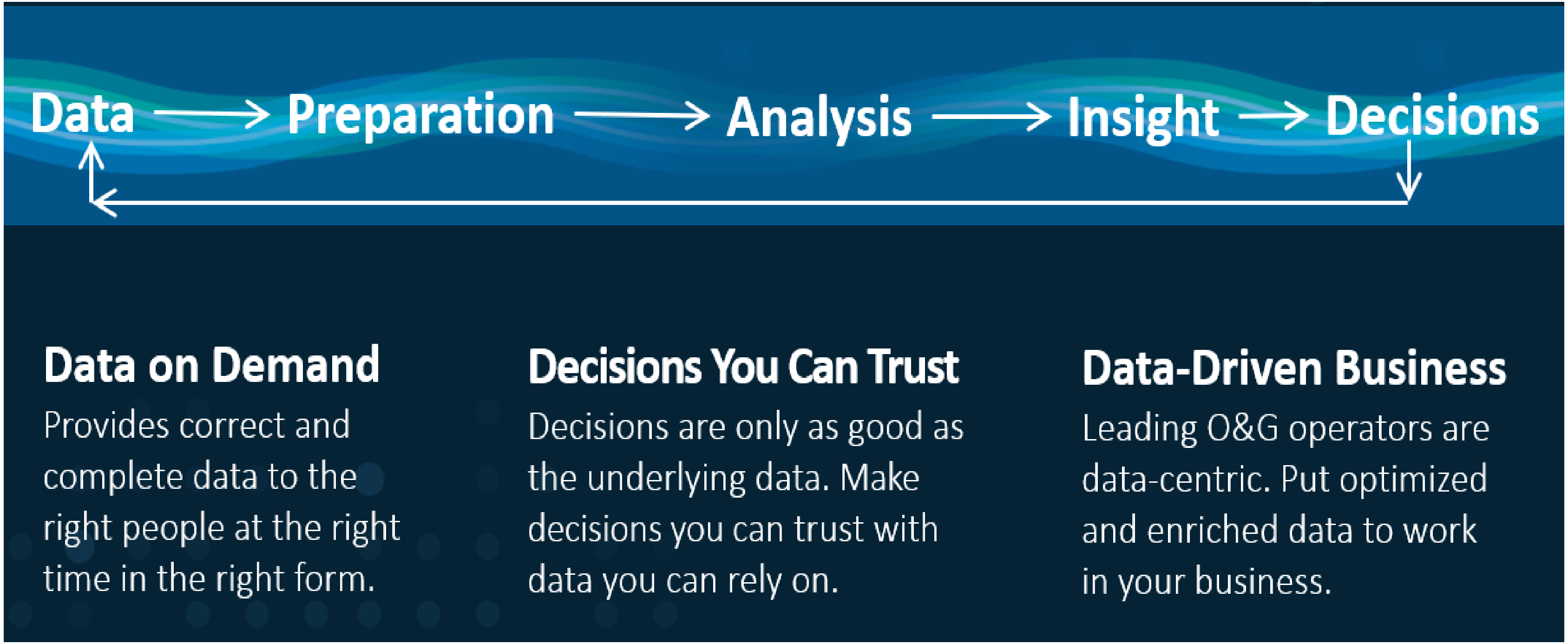
Data Preparation for AI: Upstream Data Augmentation and Feature Engineering

LEARNING OBJECTIVES

- GOAL01: Upstream Data Preparation Techniques for AI Workflows
- GOAL02: Upstream Data Augmentation: GAI
- GOAL03: Feature Engineering

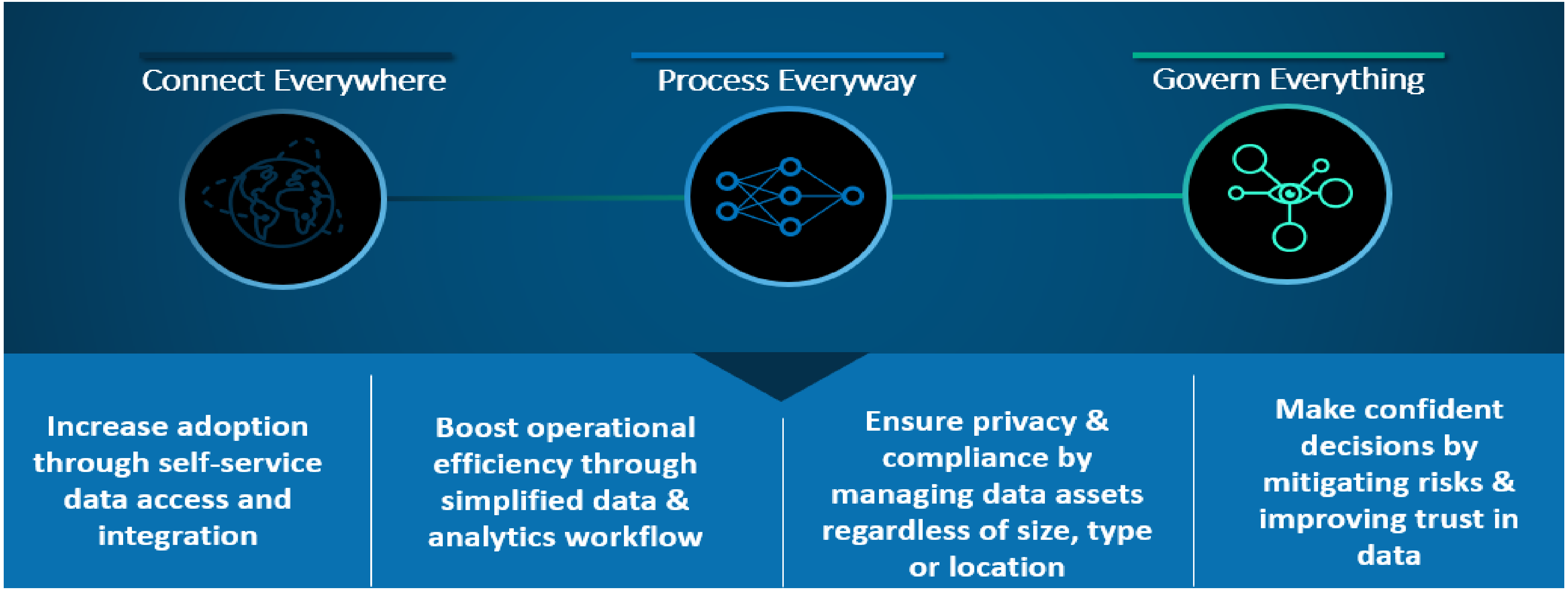
Data Preparation for AI

Upstream Data Preparation Steps

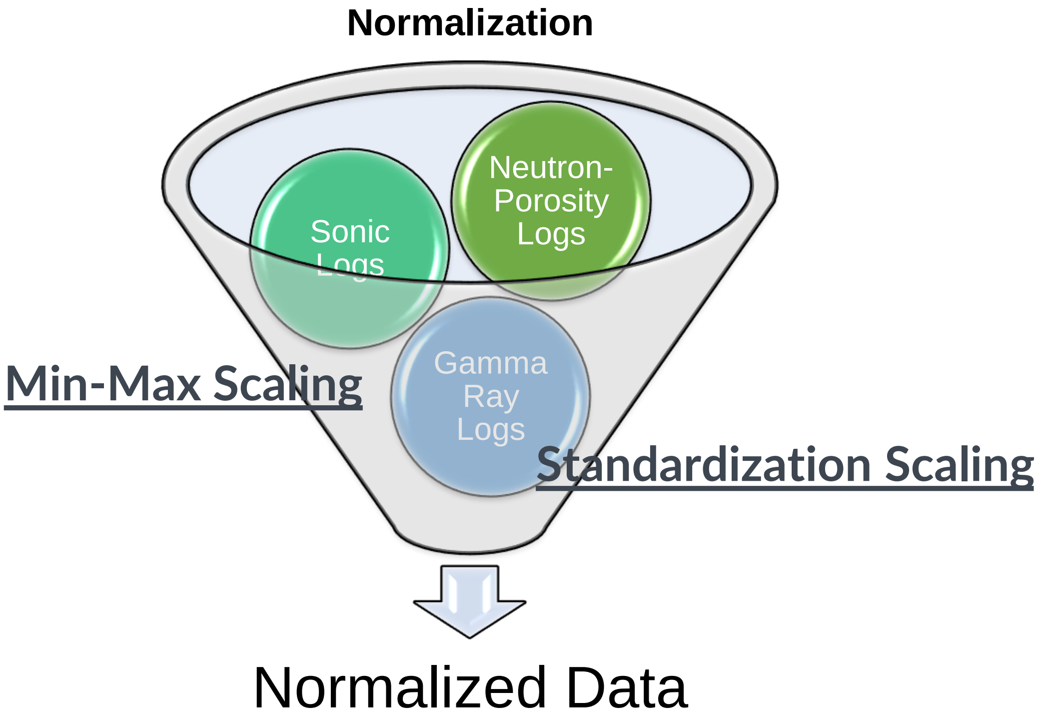


Data Preparation for AI

Upstream Data Preparation Steps

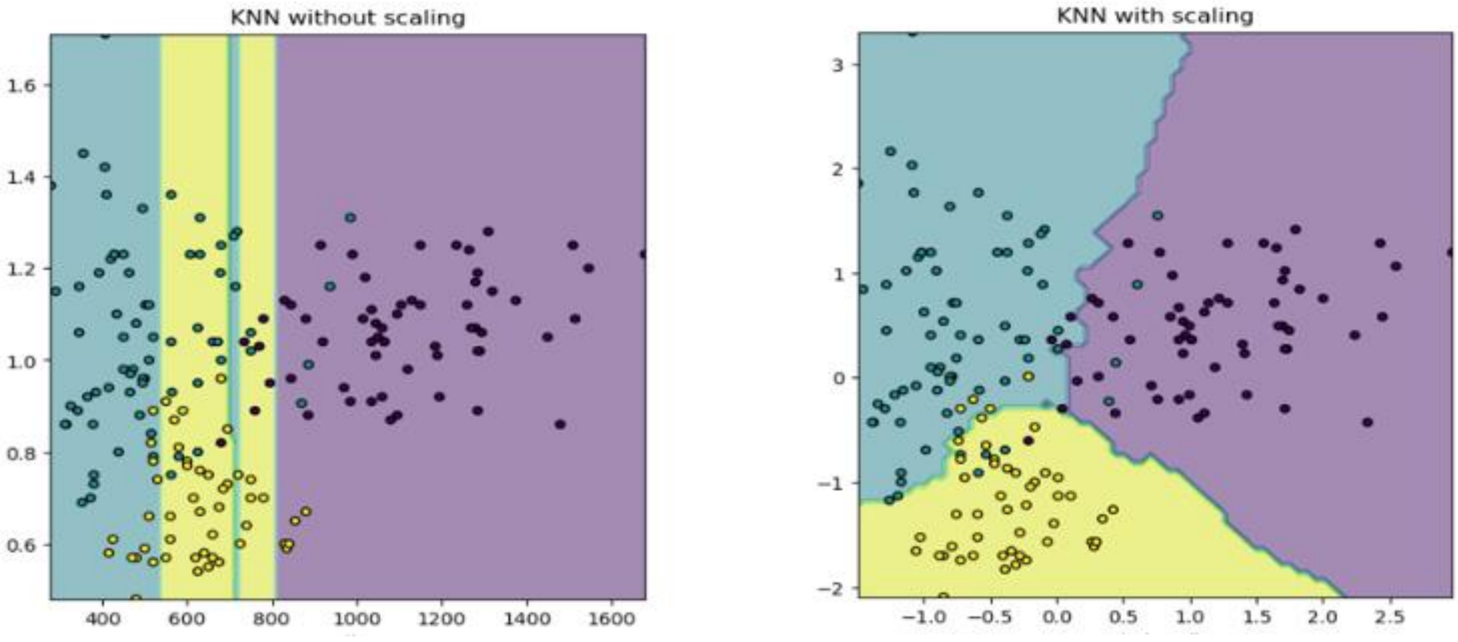


Data Preparation for AI



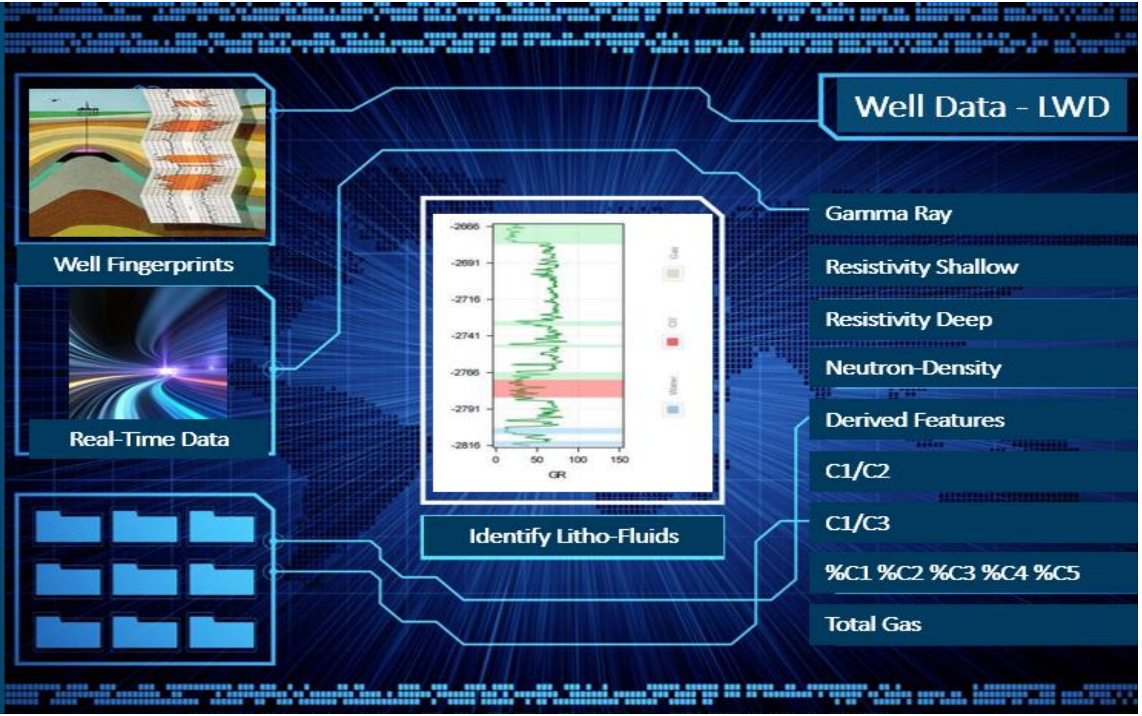
Data Preparation for AI

Scaling



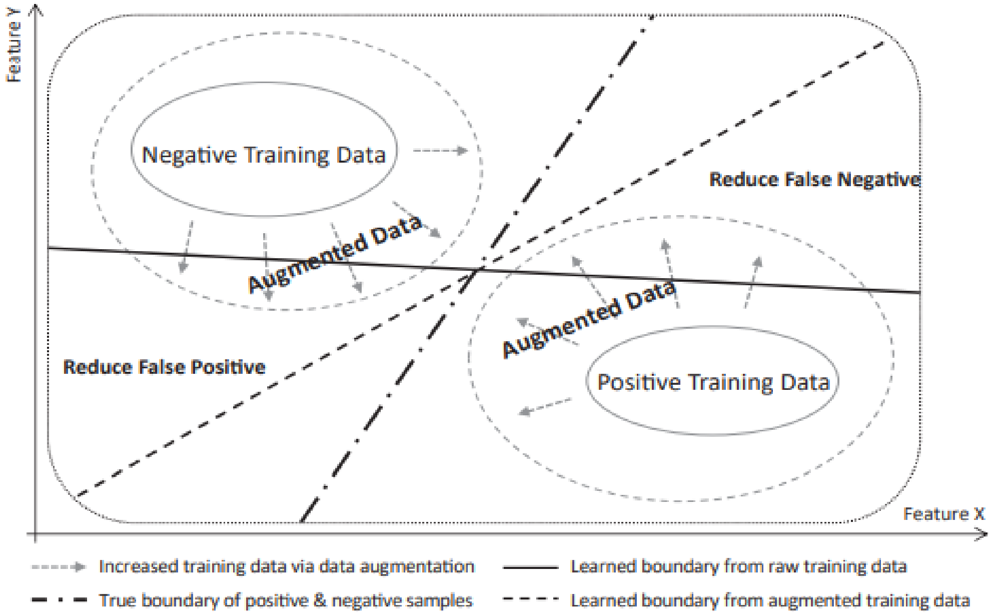
Data Preparation for AI

We have implemented a data preparation set of well log analytics and enrichment workflows to enable an innovative lithology-fluid pattern recognition assistant.



Data Preparation for AI

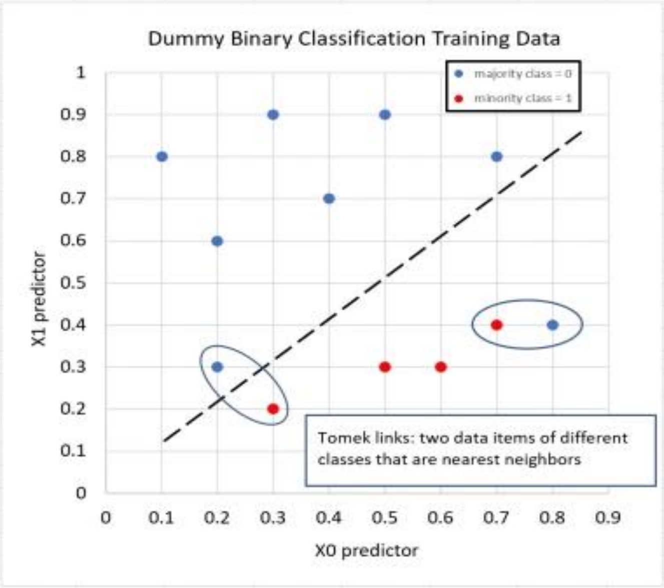
Augmentation



Data Preparation for AI

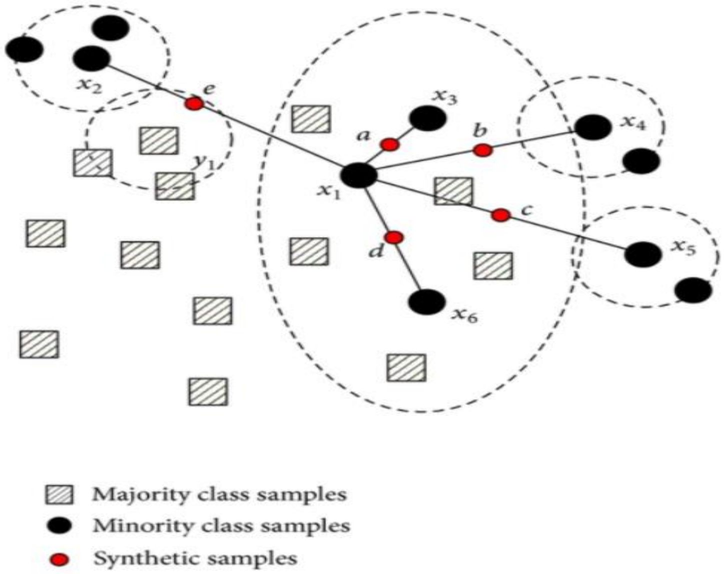
Augmentation: Imbalanced Data – Tomek Link

x0	x1	class
0.1	0.8	0
0.2	0.3	0
0.2	0.6	0
0.3	0.9	0
0.4	0.7	0
0.5	0.9	0
0.7	0.8	0
0.8	0.4	0
0.3	0.2	1
0.5	0.3	1
0.6	0.3	1
0.7	0.4	1



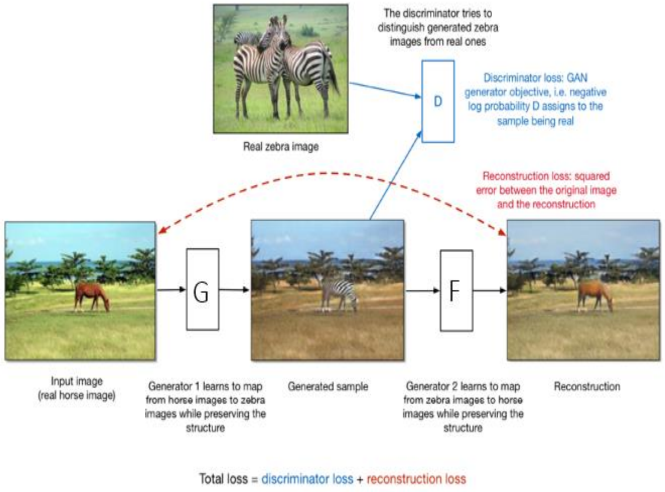
Data Preparation for AI

Augmentation: Imbalanced Data – SMOTE



Data Preparation for AI

Augmentation: Generative Adversarial Networks



- Our goal is to learn a mapping G: X → Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss.
- We couple this mapping with an inverse mapping F: Y → X and introduce a cycle consistency loss to push F(G(X)) ≈ X (and vice versa).

GAN Training: Train the GAN using real data samples. The generator is trained to generate synthetic samples that mimic the characteristics of the real data, while the discriminator is trained to distinguish between real and synthetic samples. The training involves an iterative process of updating the generator and discriminator networks to achieve a competitive equilibrium.
Synthetic Data Generation: The generator network generates synthetic samples once the GAN is trained. The GAN will produce synthetic data samples like the real data distribution.
Data Integration: Combine the real data samples with the generated synthetic samples to form an augmented dataset.

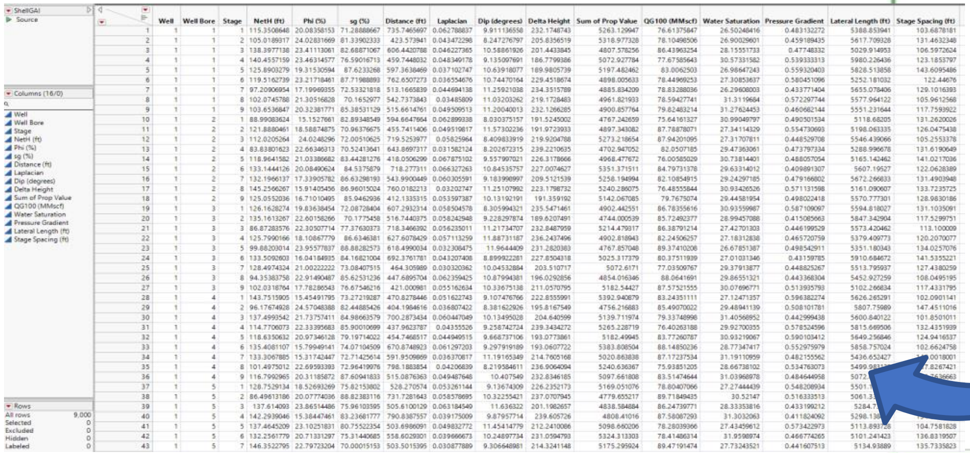
Data Preparation for AI

Generative AI (ChatGPT) in O&G Upstream

ZE

Create a python script for sample data for 50 wells with 10 well bores each and each well bore has 9 stages. Include the following parameters NetH, Phi, Sg, Distance from Peak, Laplacian, Dip, Delta Height, Sum of Prop Vol, Qg100, Water Saturation, Pressure Gradient, EUR, Lateral Length, Stage Spacing. Save the output as a CSV file

```
In [2]: import random
        import pandas

        # Set the numb          oil wells and bores
        num_wells = 5
        num_bores =

        # Set the ra           each parameter
        neth_range =              )
        phi_range = (        25)
        sg_range = (70, 90)
        distance_range = (400, 800)
        dip_range = (8, 12)
        og100_range = (75, 90)
        lateral_length_range = (5000, 6000)
        stage_spacing_range = (100, 150)

        # Initialize an empty list to store the data
        data = []

        # Generate random data for each well bore
        for well in range(1, num_wells + 1):
            for bore in range(1, num_bores + 1):
                neth = random.uniform(*neth_range)
                phi = random.uniform(*phi_range)
                sg = random.uniform(*sg_range)

                distance = random.uniform(*distance_range)
                laplacian = random.uniform(0.9 * neth, neth)
                dip = random.uniform(*dip_range)
                delta_height = random.uniform(0.1 * neth, 0.5 * neth)

                sum_prop_value = random.uniform(0.8 * neth * distance, 1.2 * neth * distance)
                 g100 = random.uniform(*og100_range)
                  ter_saturation = random.uniform(0.5 * sg, 0.8 * sg)
                  ssure_gradient = random.uniform(0.1 * sg, 0.2 * sg)

                   = random.uniform(0.1 * neth * phi * sg * distance, 0.3 * neth * phi * sg * distance)
                  ral_length = random.uniform(*lateral_length_range)
                  _spacing = random.uniform(*stage_spacing_range)

                #      end the generated data to the list
                d      append([well, bore, neth, phi, sg, distance, laplacian, dip, delta_height,
                               sum_prop_value, og100, water_saturation, pressure_gradient, eur,
                               lateral_length, stage_spacing])

        # Create      DataFrame from the generated data
        columns      ['Well', 'Well Bore', 'NetH (ft)', 'Phi (%)', 'sg (%)', 'Distance (ft)', 'Laplacian',
                      'Dip (degrees)', 'Delta Height', 'Sum of Prop Value', 'OG100 (%)',
                      'Water Saturation', 'Pressure Gradient', 'EUR (MMbbl)', 'Lateral Length (ft)',
                      'Stage Spacing (ft)']
              pd.DataFrame(data, columns=columns)

        # Save the DataFrame as a CSV file
        df.to_csv('oil_wells_data.csv', index=False)
```



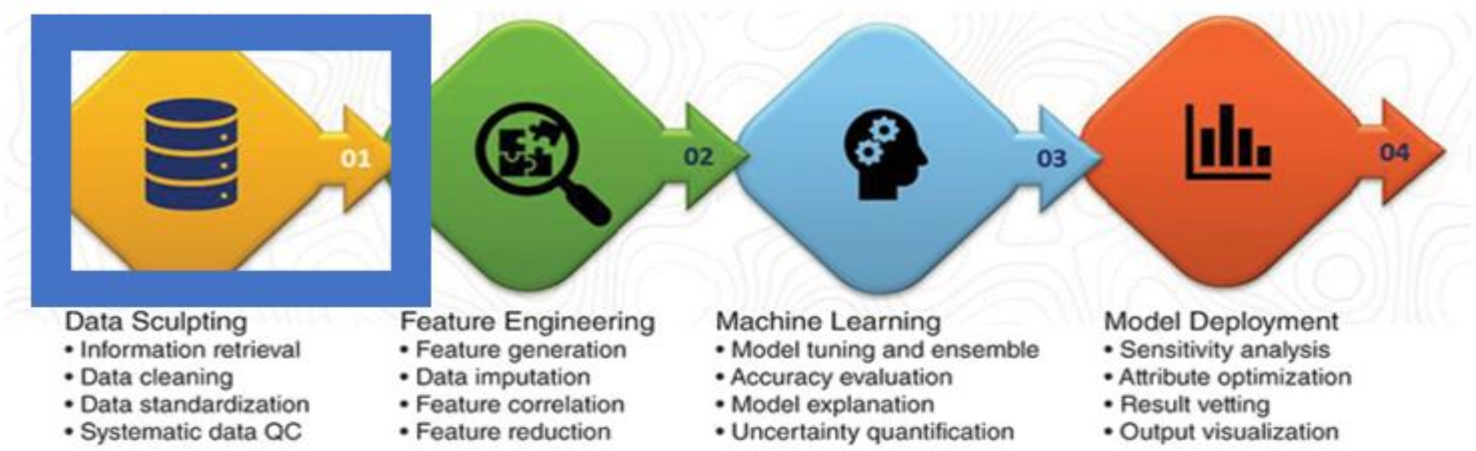
Data Preparation for AI

Feature Engineering

Feature engineering transforms raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.

Data Preparation for AI

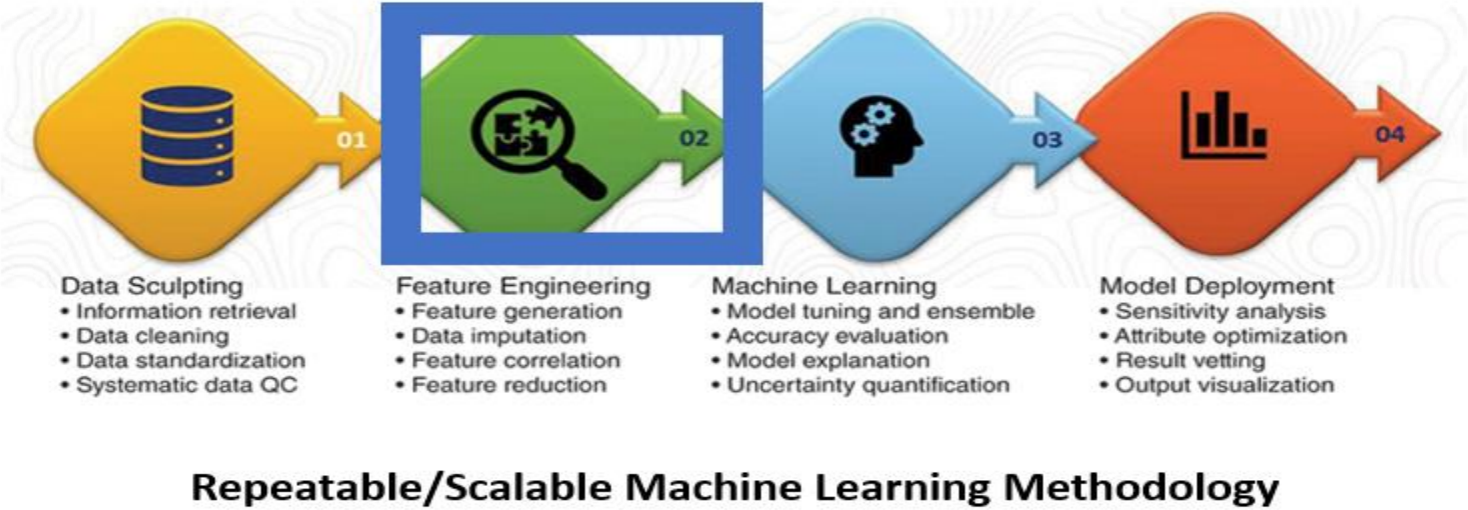
Optimize Well Spacing: Data Sculpting



Repeatable/Scalable Machine Learning Methodology

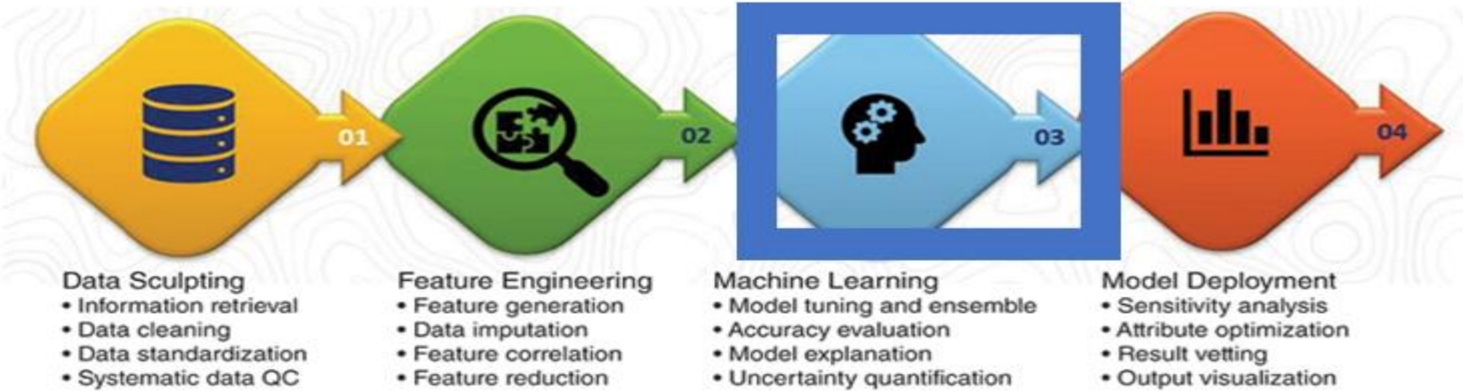
Data Preparation for AI

Optimize Well Spacing: Feature Engineering



Data Preparation for AI

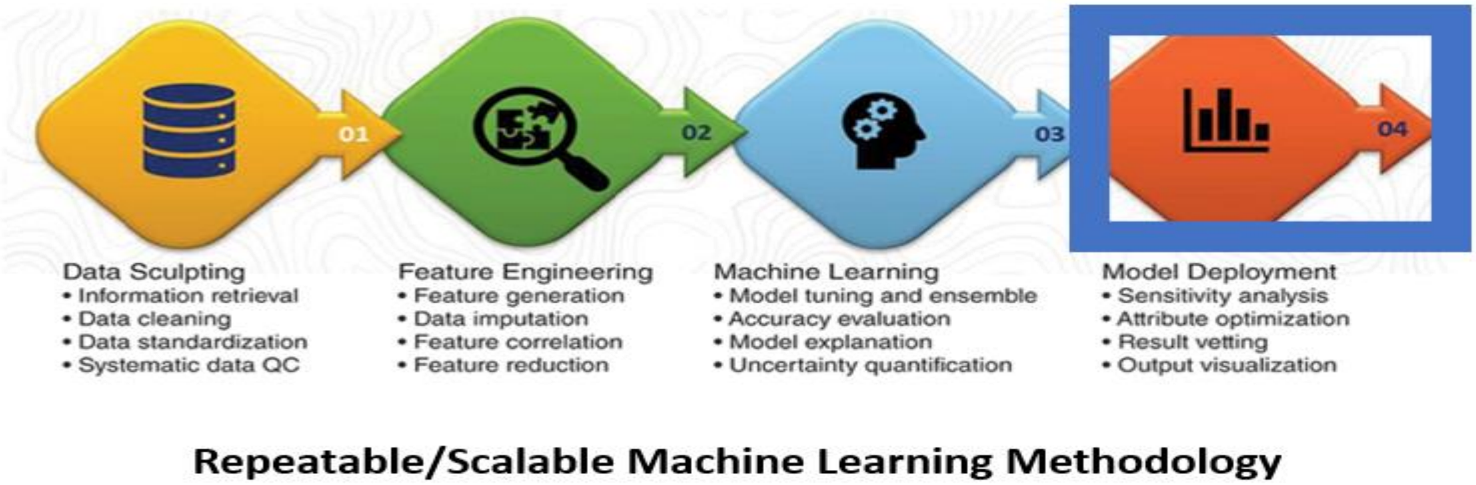
Optimize Well Spacing: Machine Learning Steps



Repeatable/Scalable Machine Learning Methodology

Data Preparation for AI

Optimize Well Spacing: Model Deployment



Module 04

Machine Learning Techniques: Supervised and Unsupervised in E&P

MODULE 04

It's worth noting that a combination of supervised and unsupervised techniques, known as semi-supervised learning, can also be employed in situations with limited labeled data. This allows leveraging labeled and unlabeled data to train models and make predictions.

The choice between supervised and unsupervised techniques depends on the specific objectives of the analysis, the availability of labeled data, and the nature of the exploration and production data. Combining these techniques can often provide comprehensive insights and support decision-making in the oil and gas industry.

Supervised Machine Learning: Machine learning techniques require labeled data, where the input features and corresponding output labels are known. These techniques are commonly used in exploration and production data analysis for prediction, classification, and regression tasks.

Unsupervised Machine Learning: Unsupervised machine learning techniques do not require labeled data and are used to discover patterns, relationships, or structures within the data. These techniques can be useful in exploring and producing data analysis for data exploration, clustering, and dimensionality reduction.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 04

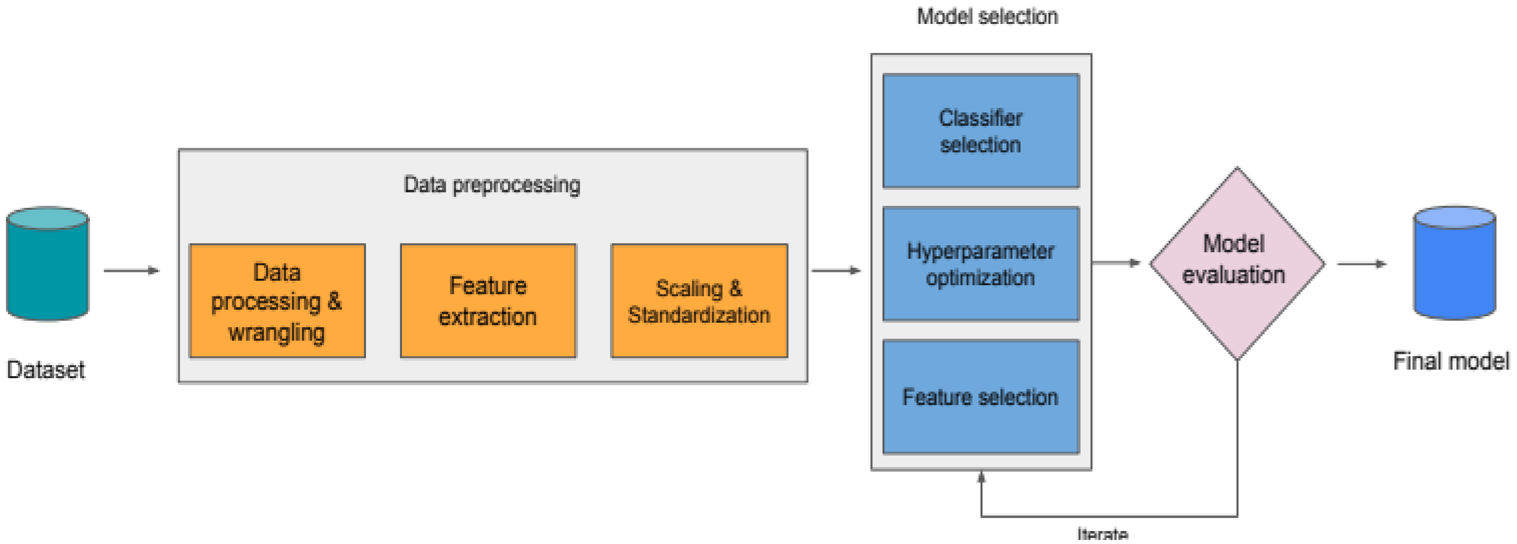
Machine Learning Techniques: Supervised and Unsupervised in E&P

LEARNING OBJECTIVES

- GOAL01: Machine Learning Fundamentals: Classification Clustering
- GOAL02: ML and DL Algorithms: Best Practices
- GOAL03: Modeling Limitations
- GOAL04: Model Selection Criteria

Machine Learning Techniques

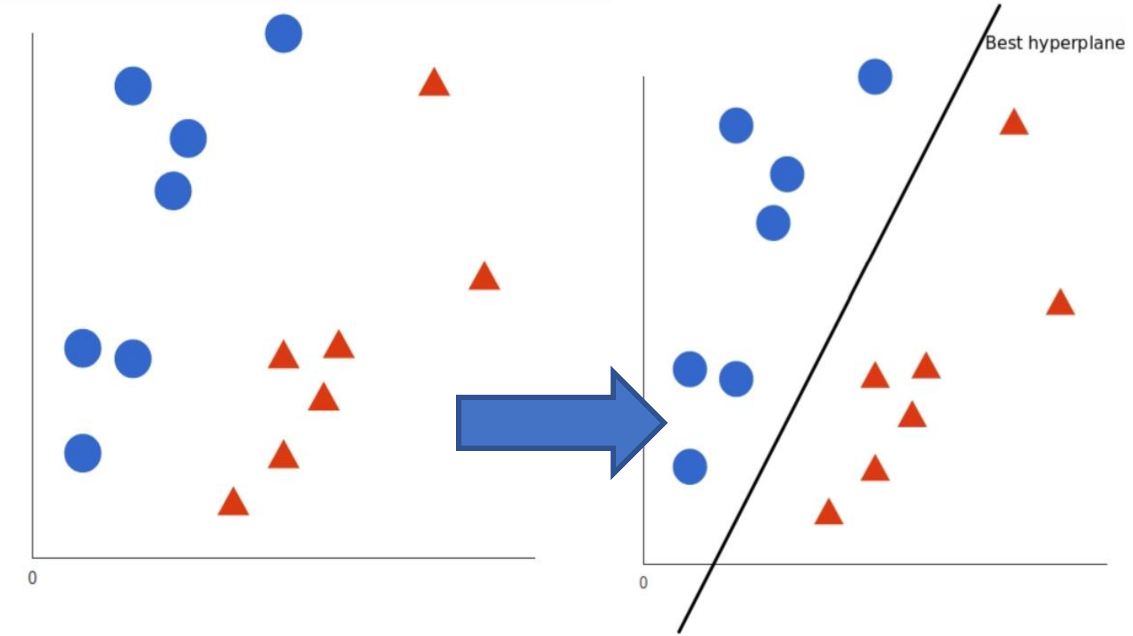
Regression Classification Clustering



Machine Learning Techniques

Classification

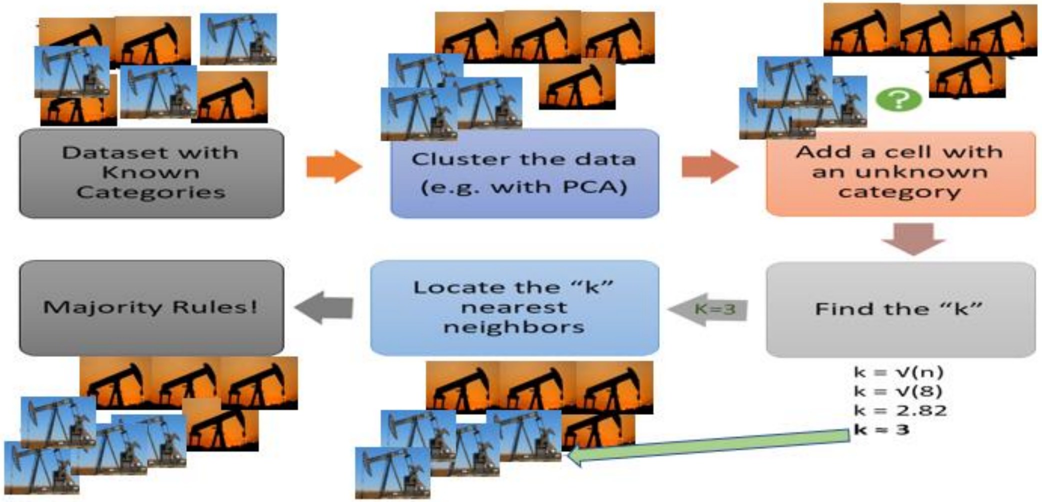
Popular Classification Algorithms:
- Logistic Regression
- Naive Bayes
- **K-Nearest Neighbors**
- Decision Tree
- Support Vector Machines

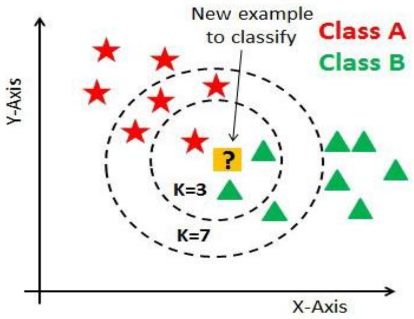


Machine Learning Techniques

Classification Clustering

K-Nearest Neighbor in one picture:

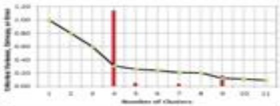


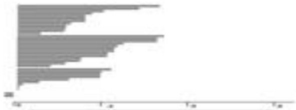


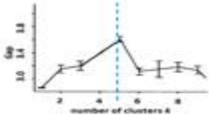
Machine Learning Techniques

Classification Clustering

Determining Optimum Number of Clusters for an Upstream Analysis



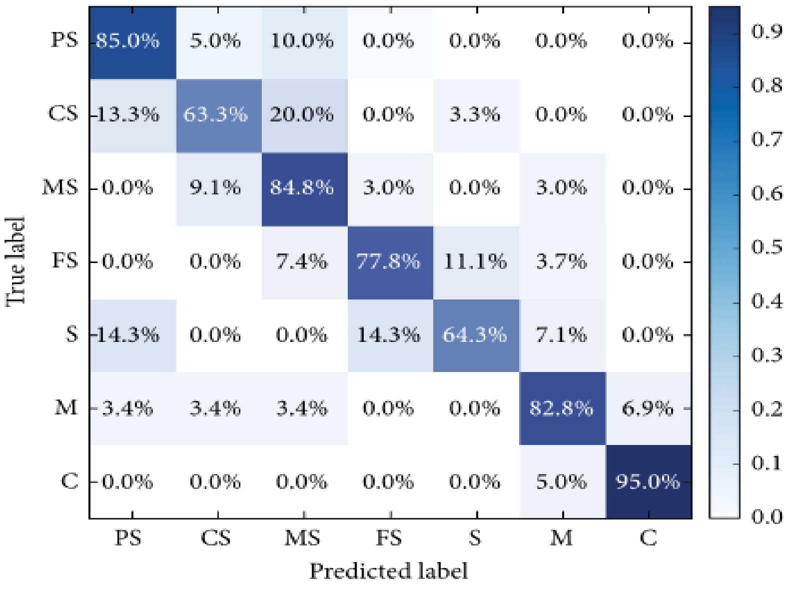




Technique	Elbow Rule	Silhouette Rule	Gap Statistic
How it works	Finds optimal (minimized) within cluster sum of squares (SSE) as a function of number of clusters	Measures how well objects are contained within clusters. Ranges from −1 to +1. High values mean well-matched to own cluster vs. other clusters.	Compares different "k" values with expected values from a dataset with no obvious clustering. The larger the "gap", the better.
Advantages	Ease of use and verification: Clear visual shows an "elbow" (red vertical line in the above image), beyond which more clusters add very little value	Easy to use. Popularity, which means there is plenty of available documentation.	Ease of use: optimal **k** value clear to see (a dashed blue line in the above image) May outperform other methods, according to authors.
Disadvantages	Sometimes ambiguous	More challenging to use and interpret; visual not too intuitive	Functions are poorly documented in R (Steorts, 2017)

Machine Learning Techniques

Classification



The diagonal of the matrix presents the percentage of lithology classes that are correctly classified

Machine Learning Techniques

ML and DL Algorithms used in O&G: Supervised & Unsupervised

Algorithms	Application	Advantages	Disadvantages
ANN Artificial neural networks MLP multi-layer perceptron FF feed forward RBF radial basis function CN convolutional FN functional PN probabilistic	Regression/ classification/ clustering	Learning algorithms are simple With available data it can superior any other model Does not depend on linearity of any function Can be used for problems which are hard or not practical to get a formula for	They are "black box" in nature so it is not easy to be understood or interpreted Lack the ability of generalization as they are exposed to overtraining and might memorize specific data For small datasets, the predictions are not acceptable
		ANN can be used for tasks that linear programs cannot handle Due to the parallel nature of the networks, they can proceed without problems even if an element fails They can learn from experience and avoid reprogramming Applicable in most problems	Neural networks need training to be used The architecture is different from problem to another For big networks, the training and processing time is high
		It is tolerant to faults Can learn from experience Effect of small changes is minor	Needs parallel processing abilities
		Handle nonlinear data Excellent in fitting applications	Exposed to overfitting Can be trapped in local optimum solution Consumes large time in training

Machine Learning Techniques

ML and DL Algorithms used in O&G: Supervised & Unsupervised

Algorithms	Application	Advantages	Disadvantages
FL Fuzzy logic	Classification/clustering	Quick, easy, strong and effect of environment changes is minor Gives a combination of numeric and symbolic picture of systems Can handle problems with strict conditions or even without exact solution Can be described with few data points or approximated datasets	If mathematical model is existing, FL is used only in case of low computational capabilities Not easy to prove the system characteristics as it lacks mathematics
		Simple reasoning, application and can deal with uncertainties and nonlinearity	Lacks robustness
		It is able to detect hyperplane of optimal separation Deals with higher degrees of dimensionality Its kernels can learn precise concepts as they have infinite Vapnik–Chervonenkis dimension Works well usually	Positive and negative examples need to be used to train the model Kernel function choice needs care Consumes memory and computation time Suffers from numerical stability issues while solving the constraint QP

Machine Learning Techniques

ML and DL Algorithms used in O&G: Supervised & Unsupervised

Algorithms	Application	Advantages	Disadvantages
SVM Support vector machine	Regression/ classification/ clustering	Provide high accuracy classifiers.Overfitting occurrence is little, excellent in dealing with noise Preferred for text classification applications that are normally high dimensional problems Intensive memory consumption	It is a binary classification technique, so it needs pairwise classification to perform multi-class classification that means one class against all others, for all classes Runs slowly and require high computational power
		Get useful information from little datasets Has generalization capabilities	Low performance with big data or multi-classification tasks Kernel function parameters affect the performance
		Easy to understand and can be interpreted Data preparation is fast Can deal with numerical and categorized data It has white box interpretable model Statistical tests can be used to validate the model accuracy Robust Efficiently handle huge data in a little time	Even for most simple concepts, the learning of an optimal DT is known as NP-complete Complex DT models cannot generalize the data well Fails to learn some concepts as it is not easy for DT to express them
DT Decision tree	Regression/ classification/ clustering	Nonlinearity among parameters do not affect the performance of DT Interpretable and explainable	Complex Duplication might happen for same sub-tree of other paths
		In case of few predictor variables, it is easy to understand Can be used in building models that contain special data types, such as text	Have huge storage requirements The similarity function selection used to correlate instances is sensitive No clear principles for selecting k, excluding over cross-validation or alike Computational rate is high
		Classes do not need to be linearly divisible Modest and powerful	Tends to disregard the attributes importance Sluggish and expensive

Machine Learning Techniques

ML and DL Algorithms used in O&G: Supervised & Unsupervised

Algorithms	Application	Advantages	Disadvantages
KNN *K* nearest neighbors	Classification/ clustering	Understandable and easy to implement technique Can be trained quickly Robust in case of associated noise It is mainly well suited for multimodal classification	Sensitive to local Data structure Memory restriction Supervised type of learning Sluggish algorithm
		High performance Arise variable measures	Computationally expensive Overfit issues
		Quick in implementation Less complex	Does not depend on variables Disregards original geometry of data
RF Random forest	Classification/ clustering	Robust with noisy data Can learn in increments	Low performance with attribute-related training data
K-means	Classification/ clustering	Data point is allowed to exist in different clusters Normal representation of the behavior of genes	Need to define number of clusters c Membership cutoff value has to be set Initial assignment of centroids affects the clusters
		Changeable model that can adapt different dataset distribution If training data increase, the parameters number does not change	In some cases, the convergence in slow
Fuzzy C-means	Classification/ clustering	Modest and easy to-understand the working algorithm As a topological clustering unsupervised technique, it can deal with dataset nonlinearity Unique in directionality reduction being able to convert high dimensions problem to 1–2 dimensions	Time consuming technique

Machine Learning Techniques

ML and DL Algorithms used in O&G: Supervised & Unsupervised

Algorithms	Application	Advantages	Disadvantages
RNN Recurrent neural network	Regression/ classification/	Can record the information as activations with time Manipulate consecutive information that are random in length	It is affected by the gradient vanishing type Not able to be stacked within extra deep modeling
CNN Convolutional neural network	Regression/ classification/ image processing	Able to detect relevant features only from given dataset Same parameters can be utilized in different problems	Tuning of parameters is difficult Requires large amount of data
		Quick training	Quality might be low
GAN Generative adversarial network	Regression/ classification/	No approximation techniques needed Does not require several entries in the samples	Unstable training Generating discrete data is difficult
DBN Deep belief network	Regression/ classification/	Layer by layer strategy of learning makes it capable of learning the features Deals with non-labelled data and can be safe from the overfitting and underfitting issues	Some pre training algorithms decrease the performance as the input data is clamped Run time is long
		Not affected by the fragmentation of training data thus it reduces over-smooth problem	Lower output quality

Machine Learning Techniques

Limitations of AI Models

Limitation	Reason	Solution
Overfitting	Lack of an appropriate amount of data to be used for training	Using the ratio of input data points to the total number of network weights used by the connections (ρ)
Coincidence	Getting a good match by coincidence for a specific dataset	Using discriminant analysis
Overtraining	When the error keeps decreasing by updating the model structure and the model can be more complex to fit a specific dataset	A training methodology that is named "early stopping" can be used Reinforcement learning with in-stream supervision, for example, the generative adversarial networks
Data availability	Sometimes the gathered data is limited	Single-shot learning in which the AI model is pre-trained on a similar dataset and then is enhanced with experience
Interpretability	The single connections in the models do not affect alone but the whole model connections combined affect results	Local interpretable model and its agnostic explanations The generalized additive models method
Generalization	Model failure in the circumstances different from the set of circumstances, which were used in building the original model	Additional resources are to be utilized for training new datasets
Bias	The nature of black-box models makes it to be prone to biases	Using model-independent perturbations

Machine Learning Techniques

Model Selection Criteria

Vocabulary

When selecting a model, we distinguish 3 different parts of the data that we have as follows:

Training set	Validation set	Testing set
• Model is trained • Usually 80% of the dataset	• Model is assessed • Usually 20% of the dataset • Also called hold-out or development set	• Model gives predictions • Unseen data

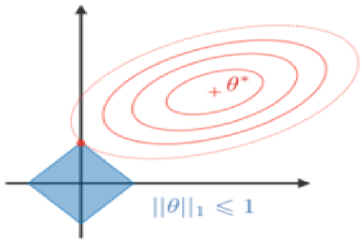
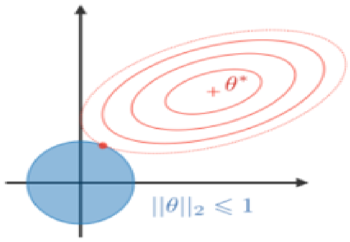
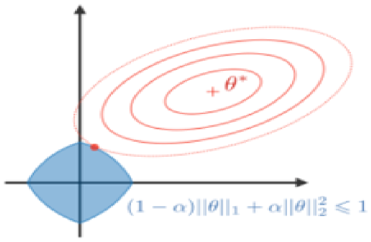
Once the model has been chosen, it is trained on the entire dataset and tested on the unseen test set. These are represented in the figure below:

Machine Learning Techniques

Model Selection Criteria

Regularization

The regularization procedure aims at avoiding the model to overfit the data and thus deals with high variance issues. The following table sums up the different types of commonly used regularization techniques:

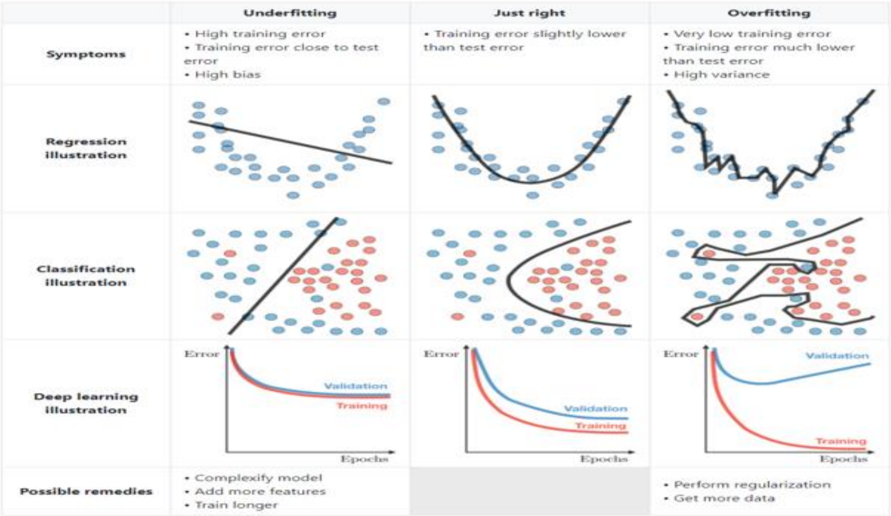
LASSO	Ridge	Elastic Net
• Shrinks coefficients to 0 • Good for variable selection	Makes coefficients smaller	Tradeoff between variable selection and small coefficients
		
$... + \lambda\|\|\theta\|\|_1$ $\lambda \in \mathbb{R}$	$... + \lambda\|\|\theta\|\|_2^2$ $\lambda \in \mathbb{R}$	$... + \lambda\Big[(1-\alpha)\|\|\theta\|\|_1 + \alpha\|\|\theta\|\|_2^2\Big]$ $\lambda \in \mathbb{R}, \alpha \in [0,1]$

Machine Learning Techniques

Model Selection Criteria

Bias/Variance Tradeoff

The simpler the model, the higher the bias, and the more complex the model, the higher the variance.



	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Deep learning illustration			
Possible remedies	• Complexify model • Add more features • Train longer		• Perform regularization • Get more data

Module 05

Deep Learning Techniques: Upstream E&P Deep Learning

MODULE 05

Deep learning techniques require substantial amounts of labeled data and significant computational resources for training. However, they have demonstrated remarkable capabilities in handling complex data and achieving state-of-the-art performance in various tasks. It's essential to carefully design deep learning architectures, preprocess the data, and fine-tune the models to extract the most meaningful insights from exploration and production data.

Deep learning techniques have gained significant attention in recent years for their ability to handle complex and high-dimensional data in various domains, including exploration and production in the oil and gas industry. Deep learning models, particularly neural networks with multiple layers, can automatically learn hierarchical representations from the data, enabling them to capture intricate patterns and relationships.

Here's an overview of deep learning techniques commonly applied to exploration and production data:

- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)
- Autoencoders
- Generative Adversarial Networks (GANs)

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 05

Deep Learning Techniques: Upstream E&P Deep Learning

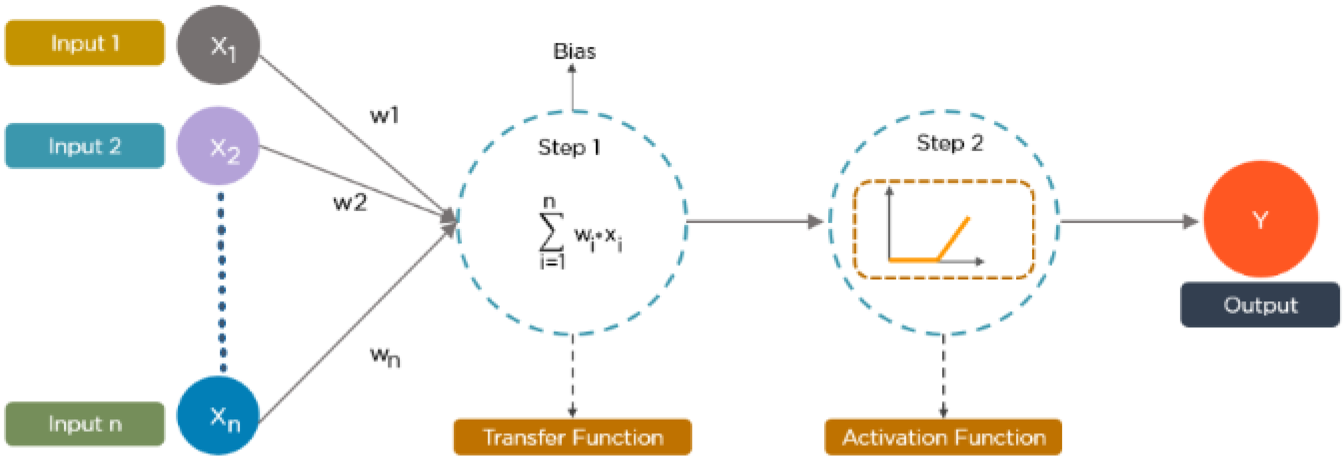
LEARNING OBJECTIVES

- ➢ GOAL01: Deep Learning Fundamentals
- ➢ GOAL02: Deep Learning Seismic Data
- ➢ GOAL03: Deep Learning Architectures used in Upstream

Deep Learning Techniques

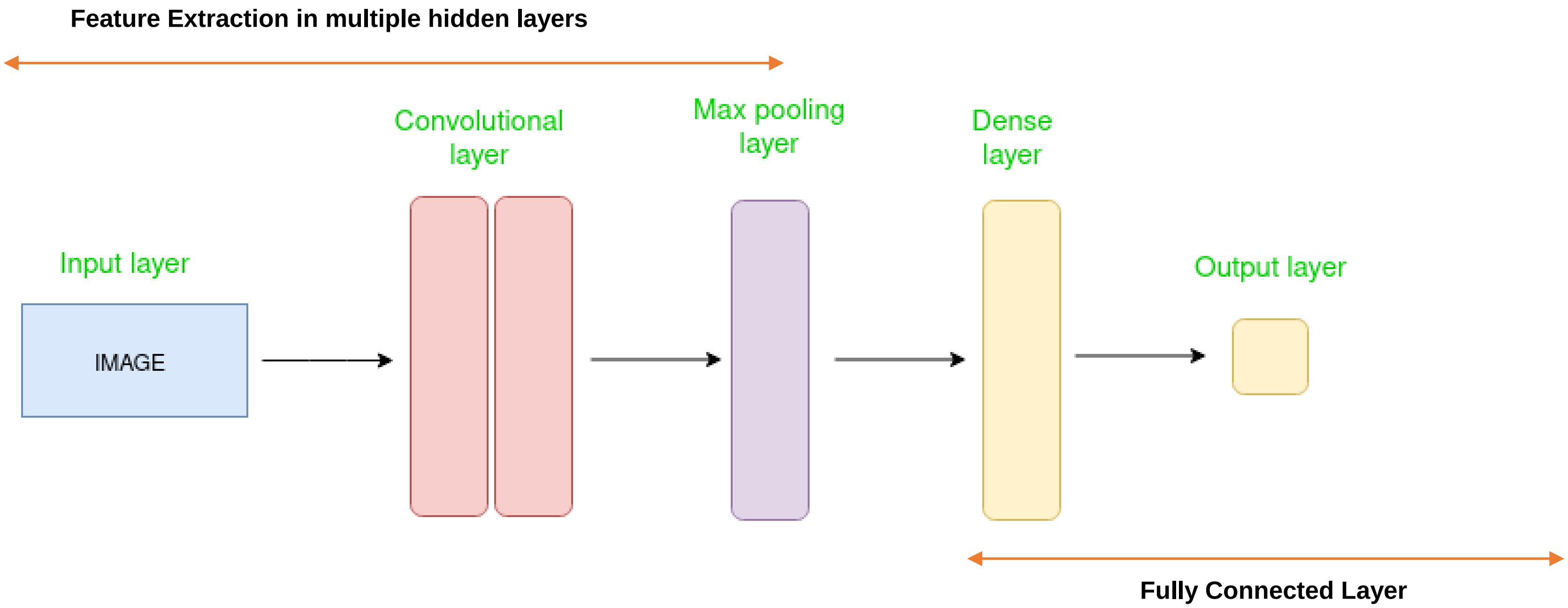
Demystifying Deep Learning

We shall focus on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) architectures.



Deep Learning Techniques

Demystifying Deep Learning: CNNs

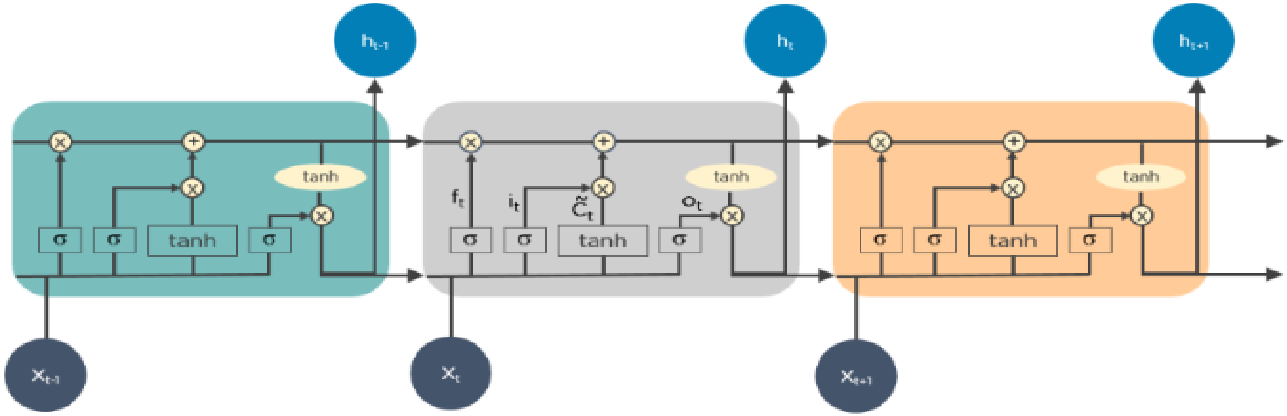


Deep Learning Techniques

Demystifying Deep Learning: RNNs

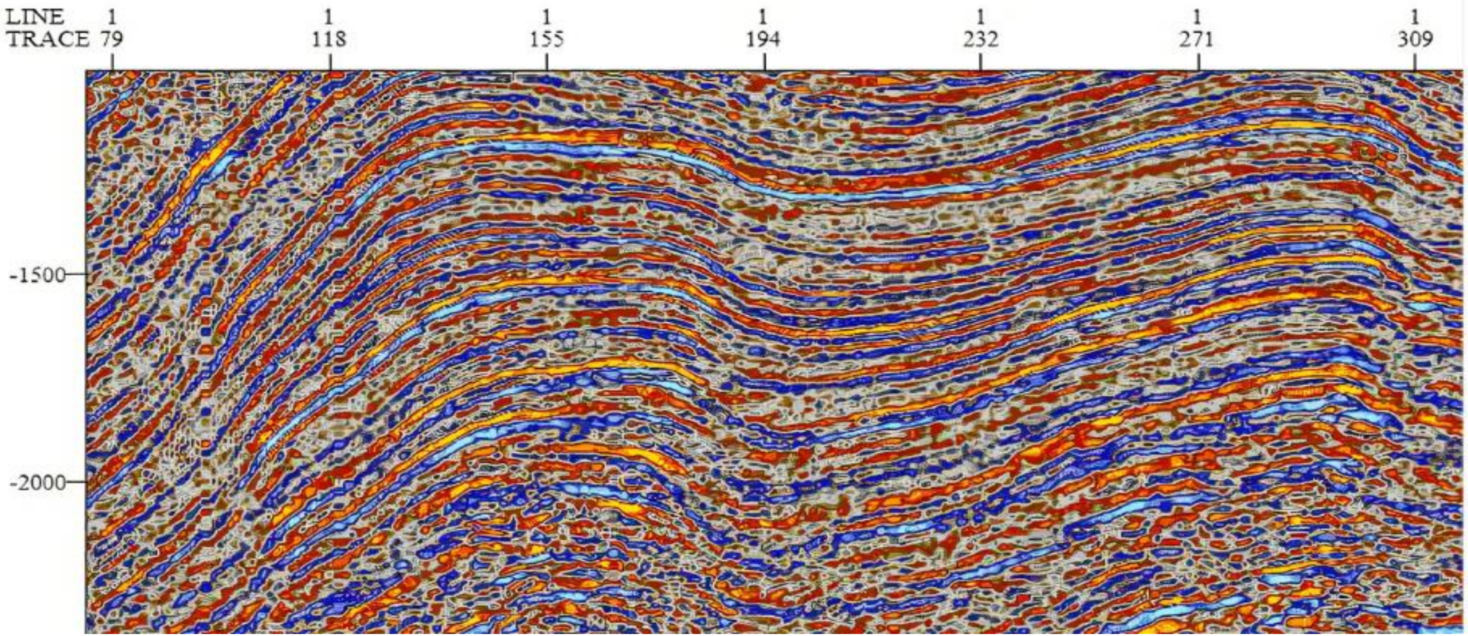
An RNN model attempts to optimally generate features from past events (remember past events) and use these features along with conventional model inputs to predict a series of interval targets or a sequence of categorical targets.

Long Short Term Memory Networks (LSTMs)



Deep Learning Techniques

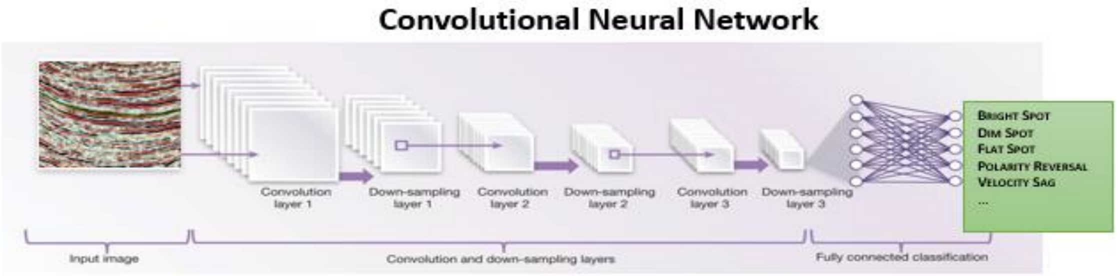
A CNN-based framework to classify anticlines structures on seismic data

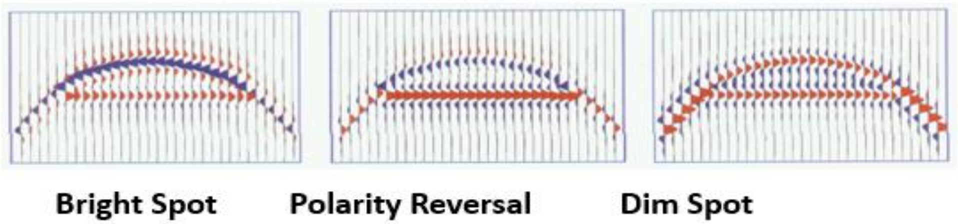


Deep Learning Techniques

Seismic Images

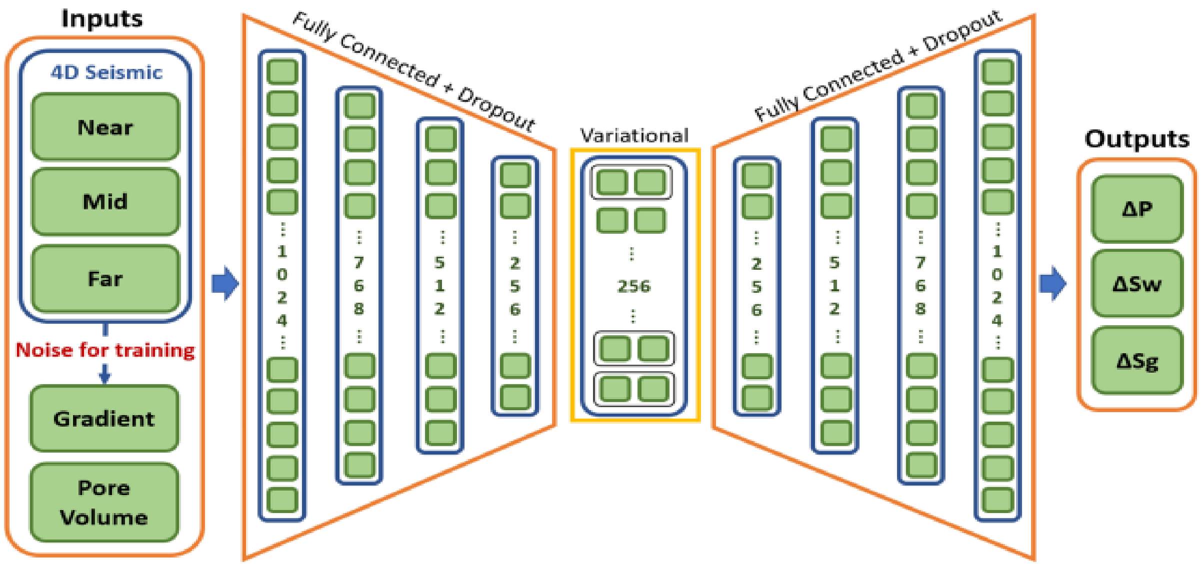
- From seismic image pixels, the first hidden layer identifies the edges
- From the edges, the second hidden layer identifies the corners and contours
- From the corners and contours, the third hidden layer identifies the parts of objects
- Finally, from the parts of objects, the fourth hidden layer identifies whole objects





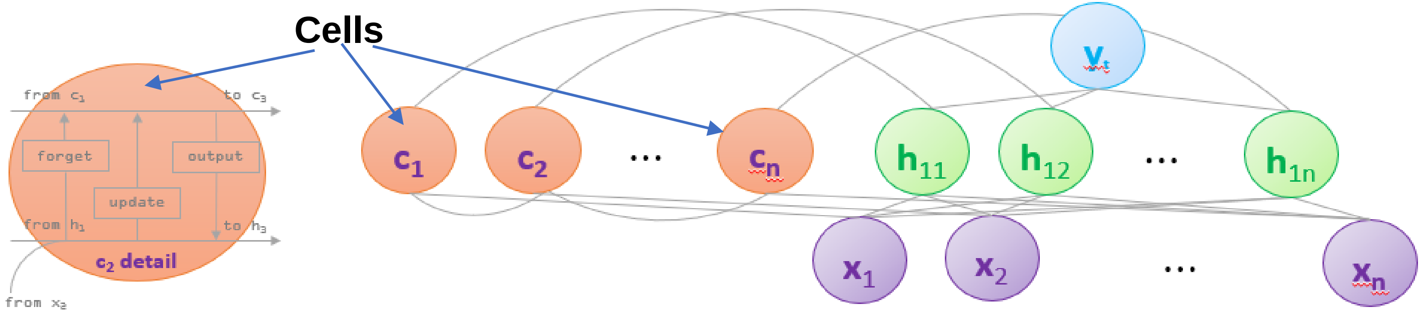
Deep Learning Techniques

4D Seismic Inversion - DNN



Deep Learning Techniques

LSTM - Recurrent Neural Network (RNN)



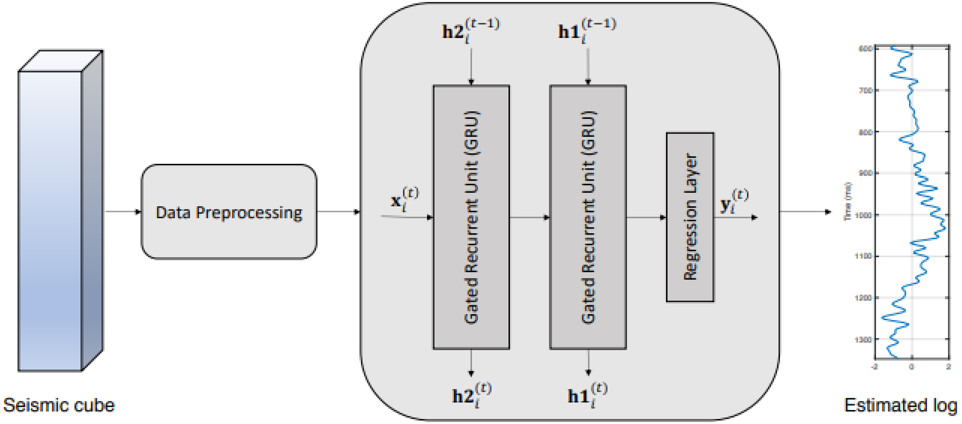
An LSTM RNN uses a more sophisticated network structure in which past information can be remembered or forgotten

Many of the recent advances in deep learning have been applied to LSTM RNN models enabling significant breakthroughs in sequence learning

Deep Learning Techniques

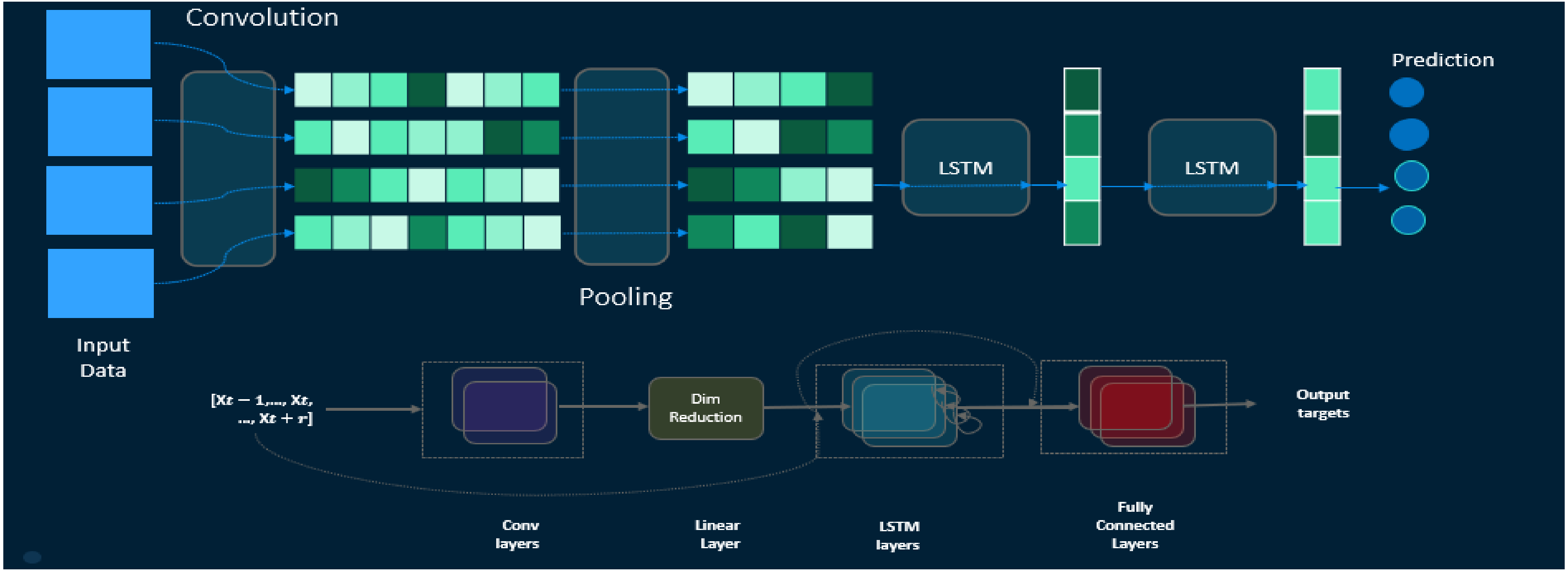
Recurrent Neural Network (RNN)

Petrophysical property estimation from seismic data using recurrent neural networks



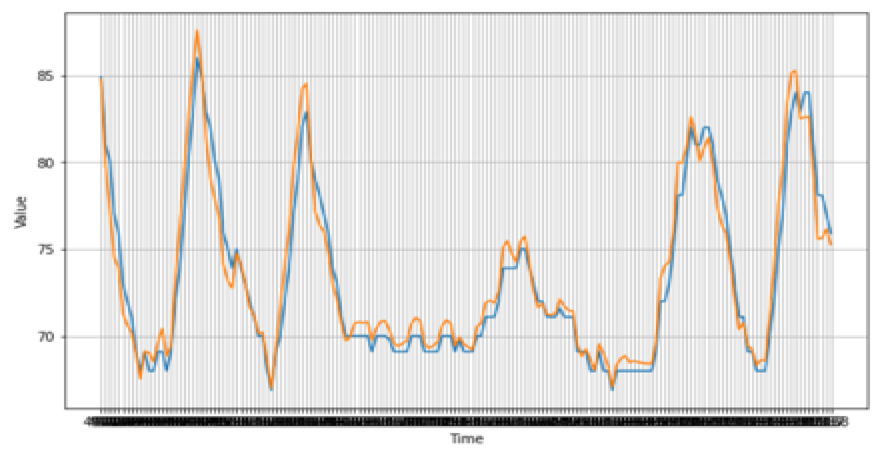
The proposed workflow with 2 layers of GRU and a regression layer

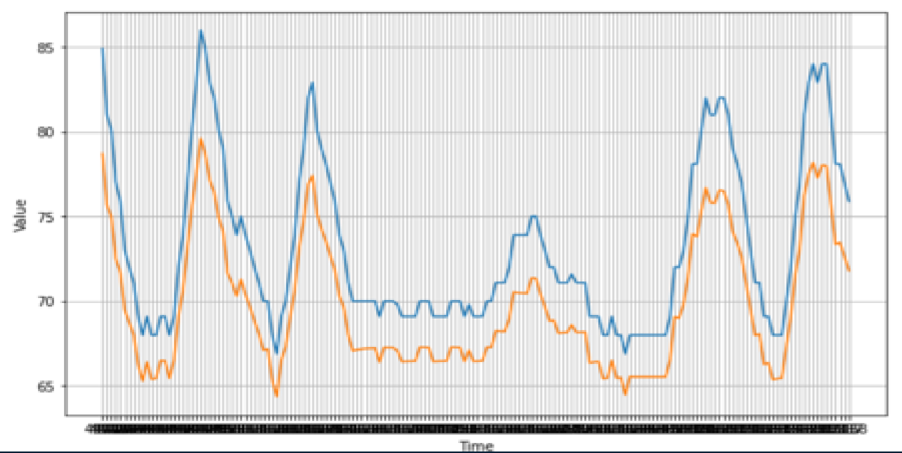
Deep Learning Techniques



CNNs +RNNs

Deep Learning Techniques





1 week horizon

CNN-LSTM
MAE = 0.94

LSTM
MAE = 3.48

CNNs +RNNs: Weekly Production Forecasting

Module 06
Case Studies: Completion Strategy and Automated Tops

MODULE 06

This Module introduces two case studies based on a data-driven analytical methodology to address a business value proposition for a completion strategy optimization in an unconventional reservoir in the USA. We shall follow the SEMMA process introduced in Module 02 under Process and Methodology.

I shall share a Society of Petroleum Engineers technical paper detailing this case study. And there is a demonstration of the case study in third-party analytics software.

The second case study under investigation in this Module is Tops Bypassed Pay. We shall discuss an automated workflow with domain input to identify the tops of historical datasets generated from well logs. The analytical workflow follows a SEMMA process to cleanse data, cluster well data, and select reference wells that provide labeled information to train machine learning techniques to automate major tops of a field under study.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 06

Case Studies: Completion Strategy and Automated Tops

LEARNING OBJECTIVES

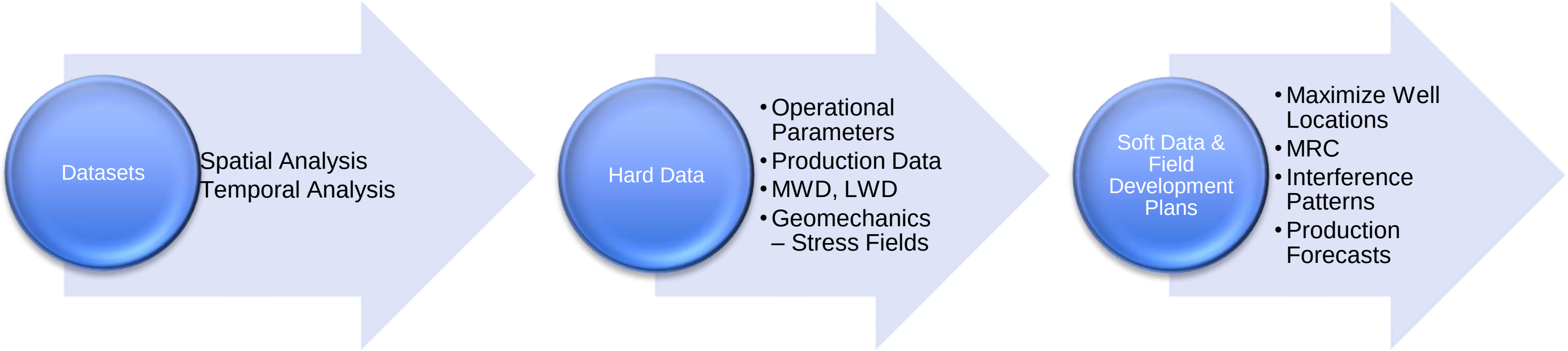
- GOAL01: Completion Strategy Optimization
- GOAL02: Automated Tops Identification – Bypassed Pay

BAUERBERG KLEIN
TRAINING & CONSULTING

Case Studies

Completion Strategy Optimization

How do you analyze risk and uncertainty, strengthen confidence in completion strategies, and quantify the impact of exploitation plans on attaining predefined targets?



Case Studies

Completion Strategy Optimization

Case Studies

Completion Strategy Optimization

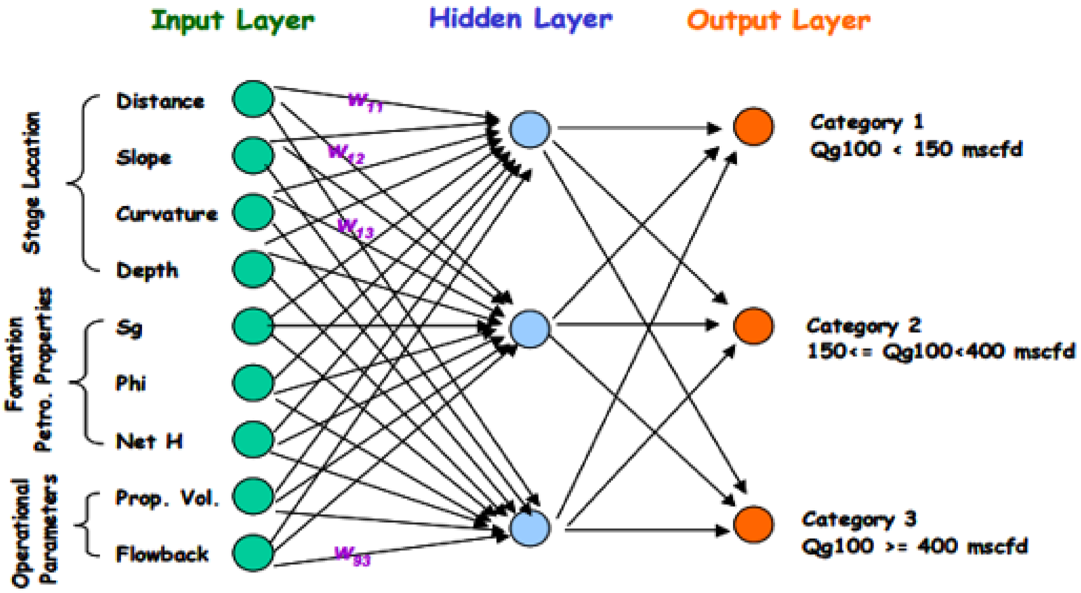
Dataset for ML Workflow:

General Well Data	• Location, Section, wellbore diagrams • 211 wells, 2399 stages
Stimulation Data	• Stimulated treatment data • 149 wells
Production Data	• 12 months of cumulative production data • Gas and water rates of production
Production Logging tool Data	• 119 wells, 166 PLTs, 32 wells with multiple PLTs
Formation Physical Property Data	• 412 wells, distribution analysis of petrophysical sandstone data
Flowback Data	• 129 wells

Geological Parameters	Formation Petrophysical Features	Operational Parameters
Distance from the global maximum location at the peak of the anticline	Sg Gas saturation	Total volume of proppant allocated for each wellbore stage
The slope of the structure gradient (1st derivative)	Porosity	Flowback initiation timing of stimulation
Curvature (2nd derivative)	Net feet of petrophysical pay	
True vertical depth from the top of the structure		

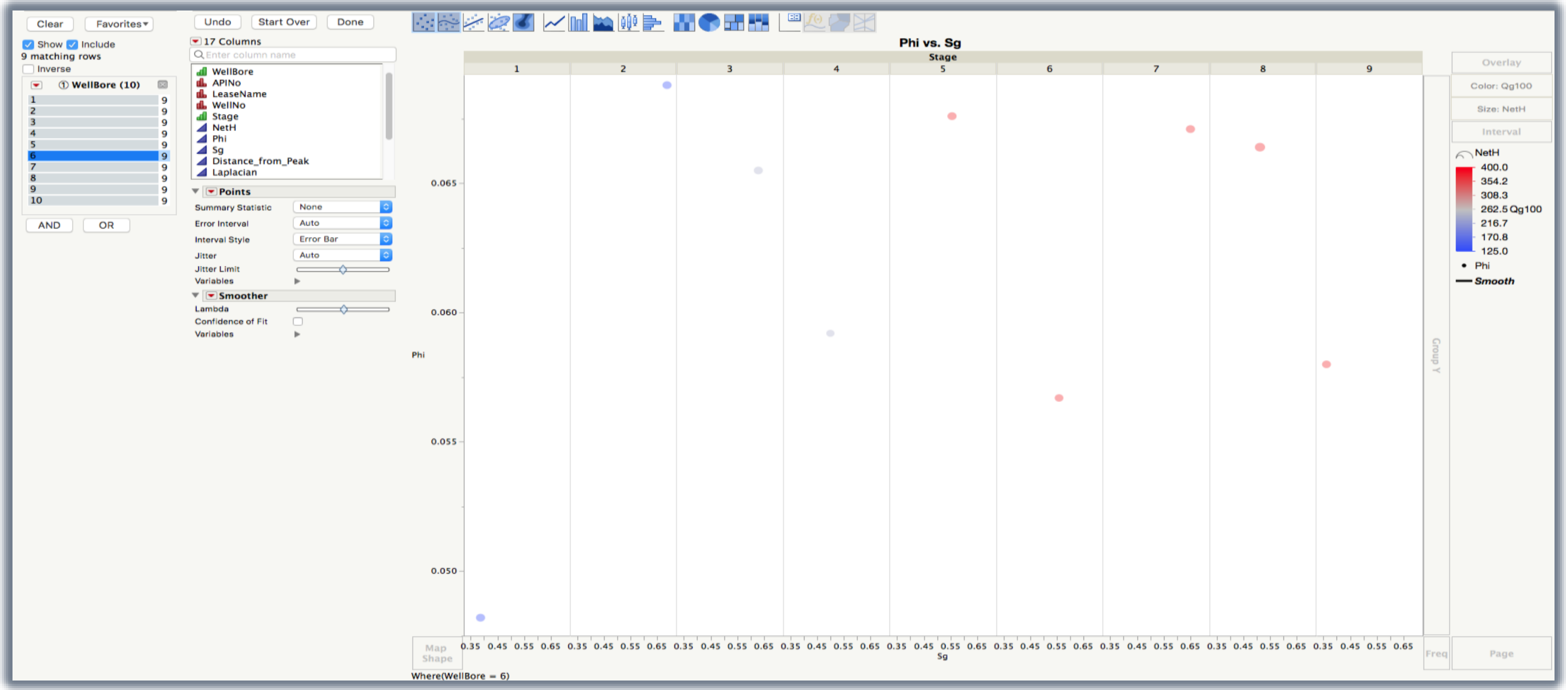
Case Studies

Completion Strategy Optimization



Case Studies

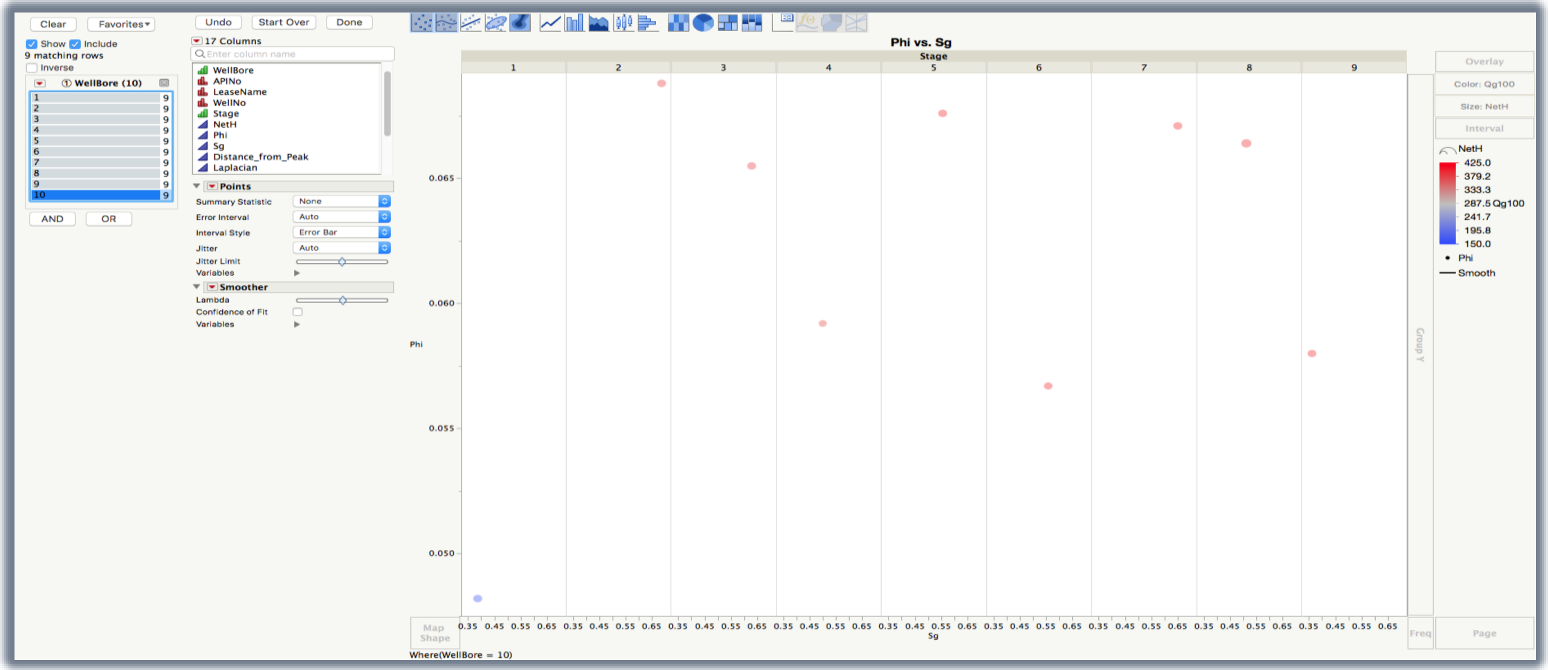
Completion Strategy Optimization



Petrophysical Parameters

Case Studies

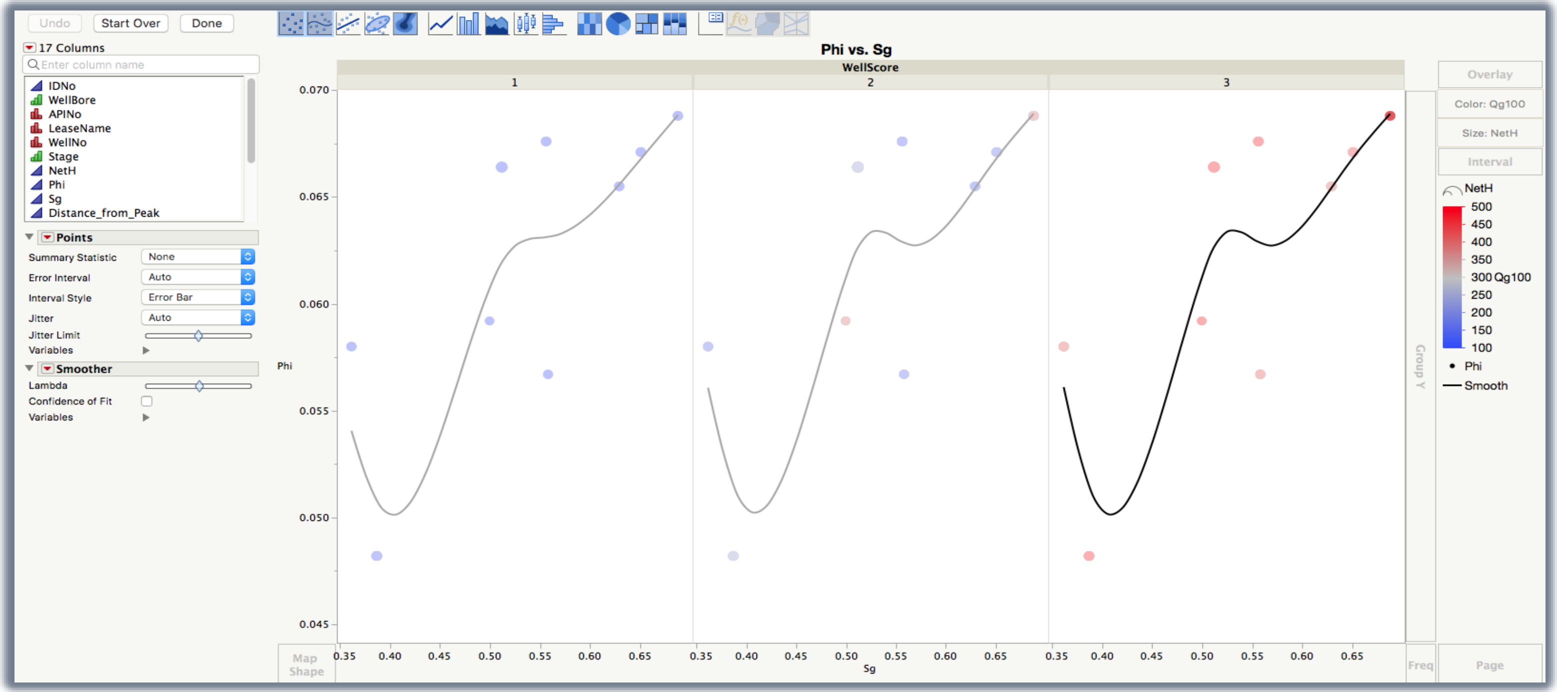
Completion Strategy Optimization



Petrophysical Parameters

Case Studies

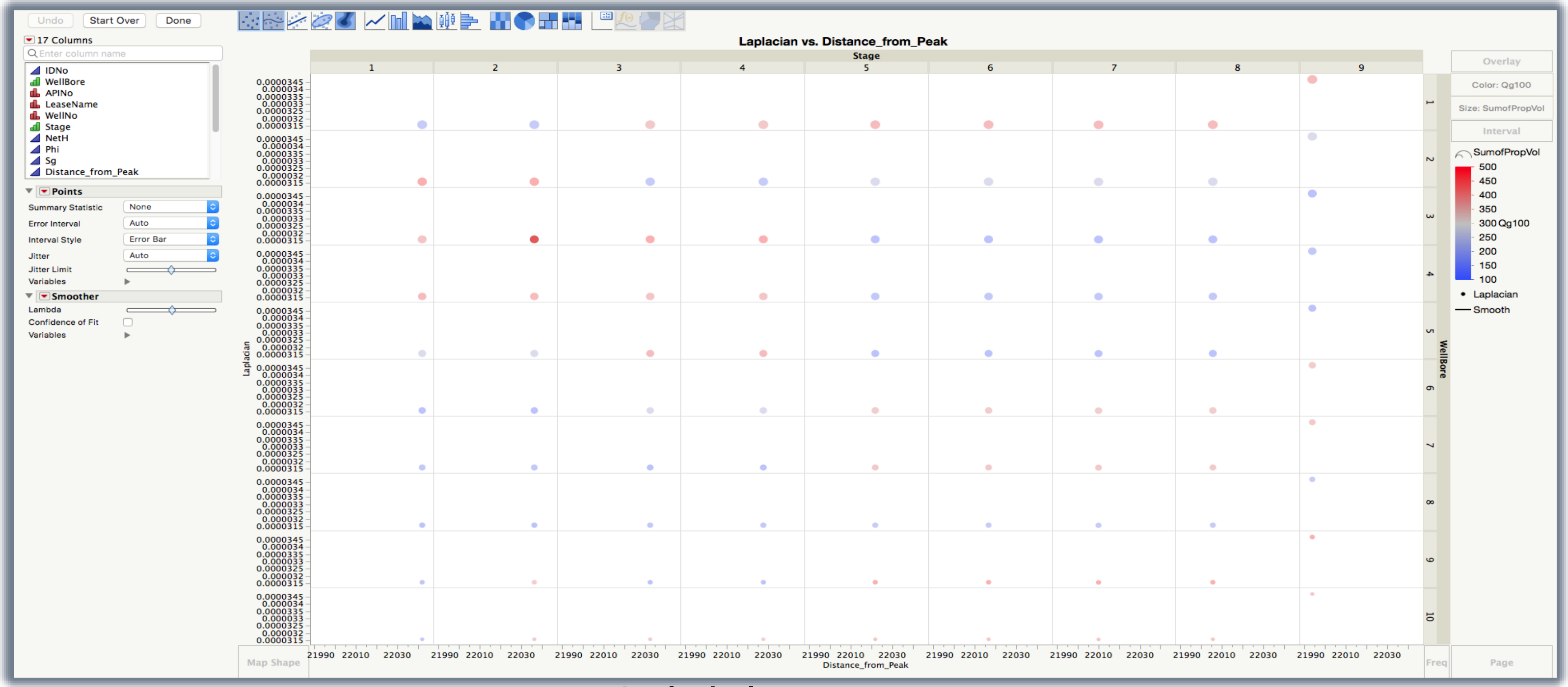
Completion Strategy Optimization



Petrophysical Parameters

Case Studies

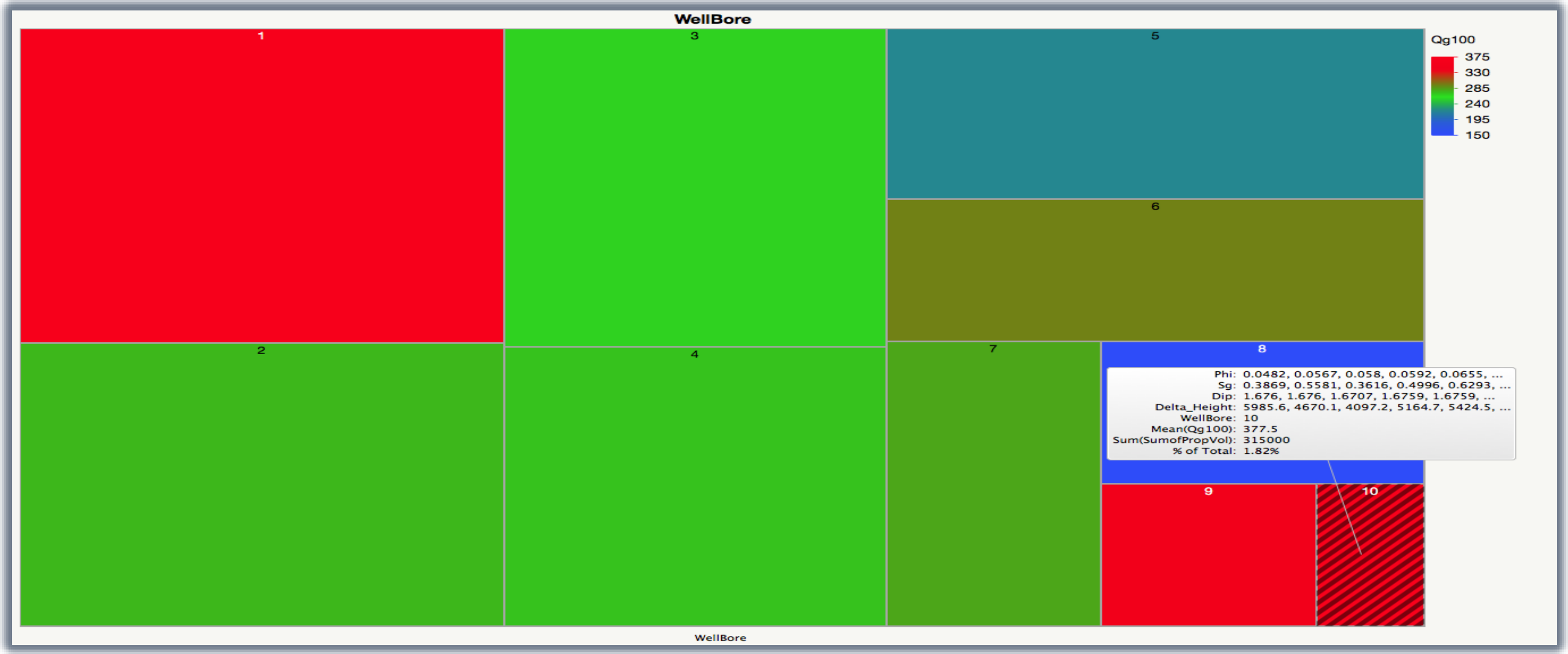
Completion Strategy Optimization



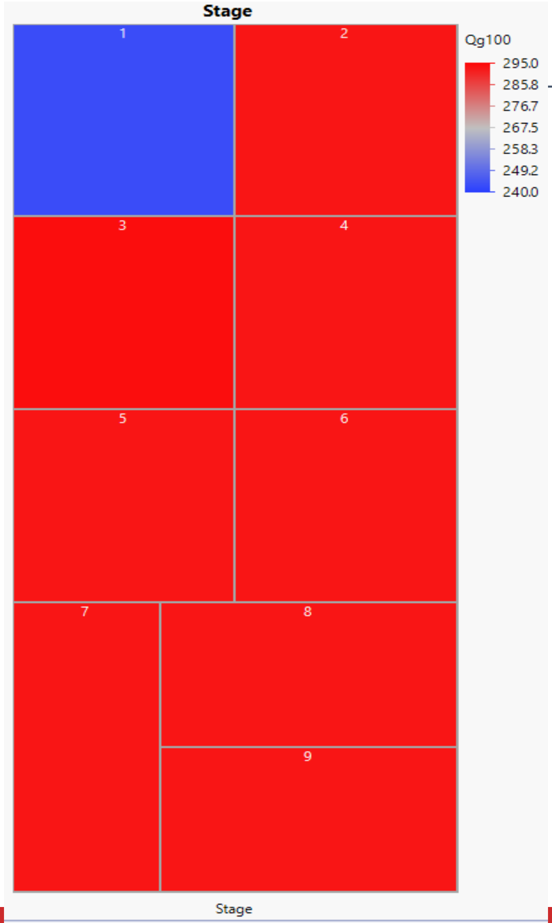
Geological Parameters

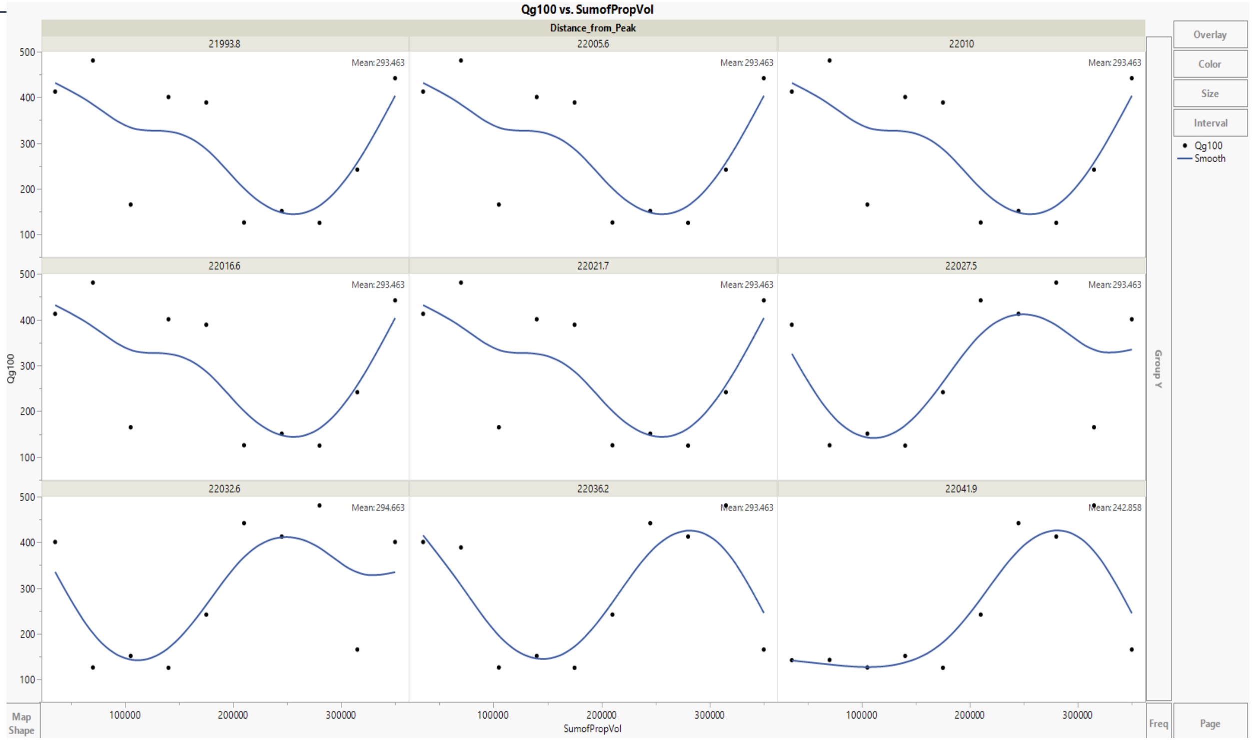
Case Studies

Completion Strategy Optimization



Geological Parameters





Geological Parameters

Case Studies

Completion Strategy Optimization

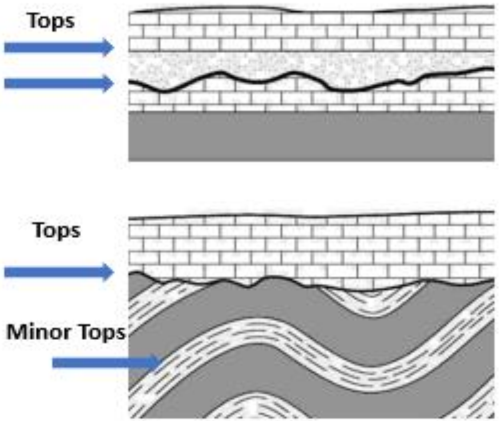
- Important to map Business Problem to a Data Mining Problem
- **Critical Workflows: Data Management** and **Dimensionality Reduction**
- SEMMA Process enables **repeatable** and **scalable** soft-computing methodologies
- Exploratory Data Analysis: Get a feel for your data!
- Model and Score
- Operationalize: Avoid academic exercise!
- Ensure new data re-trains supervised models

- 49 stages selected for proppant increase resulted in incremental 1,962 Mscf/d at the cost of $2.23 million
- Economic ROI favorable at gas prices > $3.0 Mscf/d
- Breakeven at $2.60 Mscf/d

Case Studies

Tops Bypassed Pay

Rock Strata: Tops and Bottoms of Geologic Layers

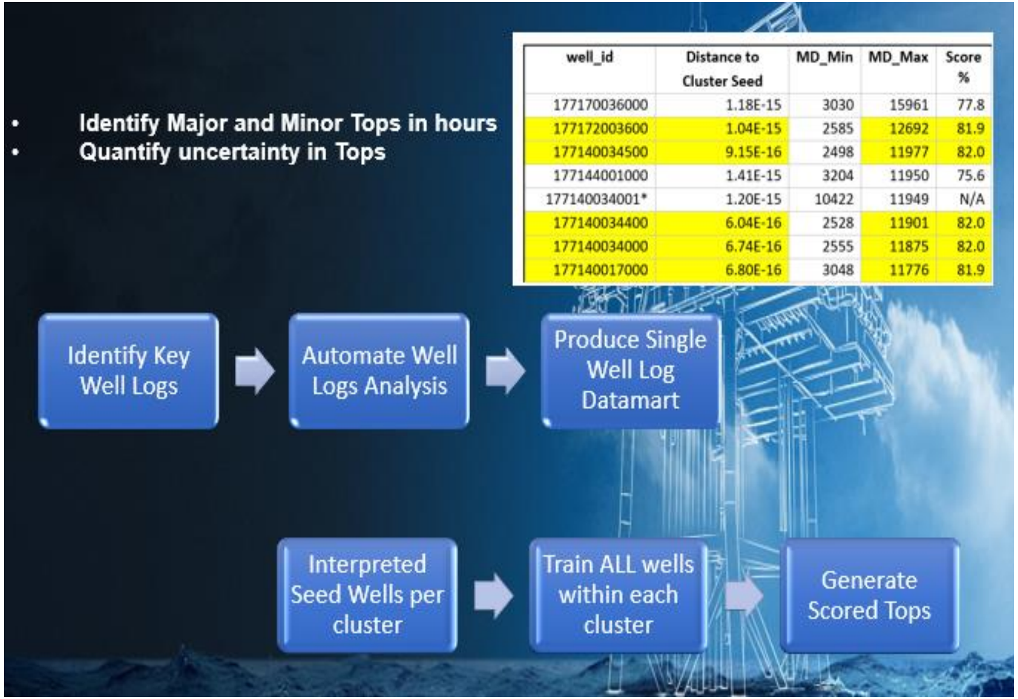


There are three basic types of contacts:

1. **Depositional contacts**, where a sediment layer is deposited over preexisting rock.
2. **Fault contacts**, where two units are juxtaposed by a fracture on which sliding has occurred.
3. **Intrusive contacts**, where one rock body cuts across another rock body.

Case Studies

Tops Bypassed Pay



well_id	Distance to Cluster Seed	MD_Min	MD_Max	Score %
177170036000	1.18E-15	3030	15961	77.8
177172003600	1.04E-15	2585	12692	81.9
177140034500	9.15E-16	2498	11977	82.0
177144001000	1.41E-15	3204	11950	75.6
177140034001*	1.20E-15	10422	11949	N/A
177140034400	6.04E-16	2528	11901	82.0
177140034000	6.74E-16	2555	11875	82.0
177140017000	6.80E-16	3048	11776	81.9

POR: Porosity
CNLC: Compensated Neutron Porosity
Sw: Water Saturation
BHC: Sonic
EGR: Gamma Ray
TSPR, TSSS, TSSW: Thomas-Stieber Logs Vsh: Shale Volume

- Measurements taken at a regular step (Redwater: 3 to 20 cm, depending on the well; Bay Marchand: 0.5 ft)
- Each log has several measured and derived variables
- Bay Marchand has only MD, POR, SW, and VSH

Variable Name	Description	Type
MD	Measured Depth	
BHC	Borehole Compensated Sonic	Primary
CNLC	Compensated Neutron Porosity	Primary
EGR	Corrected gamma ray	Primary
TSPR, TSSS, TSSW	Thomas-Stieber logs	Derived
VSGC	Gamma ray corrected shale volume	Derived
POR	Porosity	Derived
SW	Water saturation	Derived
VSH	Shale volume	Derived

Case Studies

Tops Bypassed Pay

9 clusters with 23 to 140 wells assigned to a different cluster

Number of wells covered: 418 out of 1182 (35.4% of the data)

All but one cluster show results above 60% (67.8% to 91.3%)

Overall prediction score: 70.0%

Training wells per cluster: 1 to 3

Total training wells: 14 (3.4% of the data)

Cluster #	Cluster size	Training wells	Score, %
1	140	3	52.3
2	134		
3	54	2	80.4
4	48	2	74.1
5	35	1	80.6
6	29		
7	29	1	89.3
8	29	1	81.1
9	28	2	67.8
10	27		
11	27	1	70.4
12	27		
13	26		
14	23	1	91.3

Case Studies

Tops Bypassed Pay

- Measured Depth (MD) to a top of a rock layer
- SME identified major rock layer tops
- Analytics automated rock layer tops
- Reduced SME decision cycles from 6 months to 4 days
- Accuracy > 70%
- Tops Automated Picker identified potential drilling locations to exploit bypassed pay

Well ID	Formation Name	MD (m)
1W40552030050000	CLRD	416.14
1W40552030050000	SSPK	572.46
1W40552030050000	BFS	646.59
1W40552030050000	VKNG	691.88
1W40552030050000	JLFU	719.73
1W40552030050000	MNVL	733.86

Module 07

Case Studies: Seismic Attributes

MODULE 07

This Module introduces two case studies based on a data-driven analytical methodology to address a business value proposition using seismic attributes. We shall follow the SEMMA process introduced in Module 02 under Process and Methodology.

The first technique studies Self-Organizing Maps (SOMs), an unsupervised neural network algorithm. SOMs are a valuable tool for exploratory data analysis and visualization, which map from a high-dimensional input space to a low-dimensional lattice, preserving the topology of the data set as faithfully as possible. We shall identify critical features to optimize gas production in an unconventional reservoir.

The second case study under investigation in this Module is Acoustic Impedance. We shall discuss an automated workflow with domain input to identify important historical datasets that can predict the Acoustic Impedance based on five seismic attributes. The analytical workflow follows a SEMMA process to cleanse data, perform Exploratory Data Analysis, and generate Tukey diagrams to understand feature relationships and statistical predictive power.

We shall close with a demonstration using a Jupyter Notebook to generate Acoustic Impedance logs.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 07

Case Studies: Seismic Attributes

LEARNING OBJECTIVES

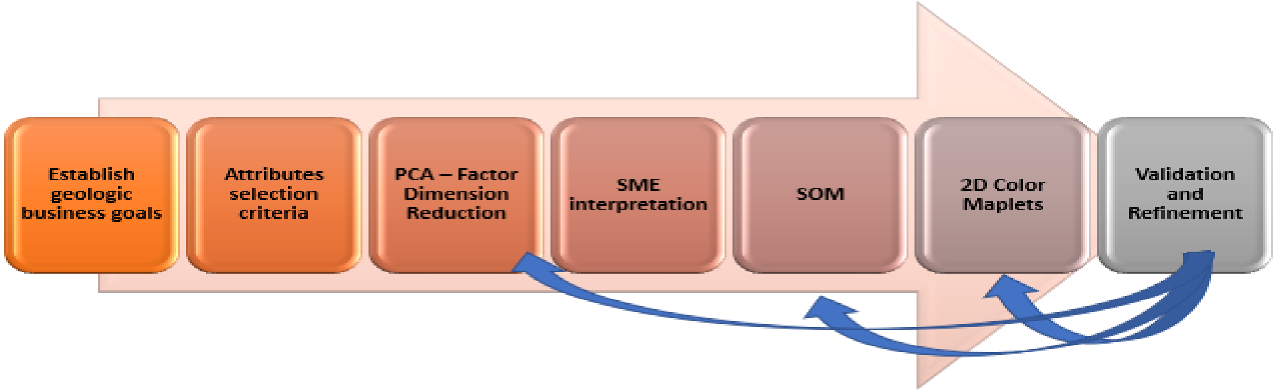
- ➢ GOAL01: Case Studies – Seismic Attributes
- ➢ GOAL02: Self-Organizing Maps (SOMs)
- ➢ GOAL03: Case Studies – Acoustic Impedance

Case Studies

Classifying Multiple Seismic Attributes

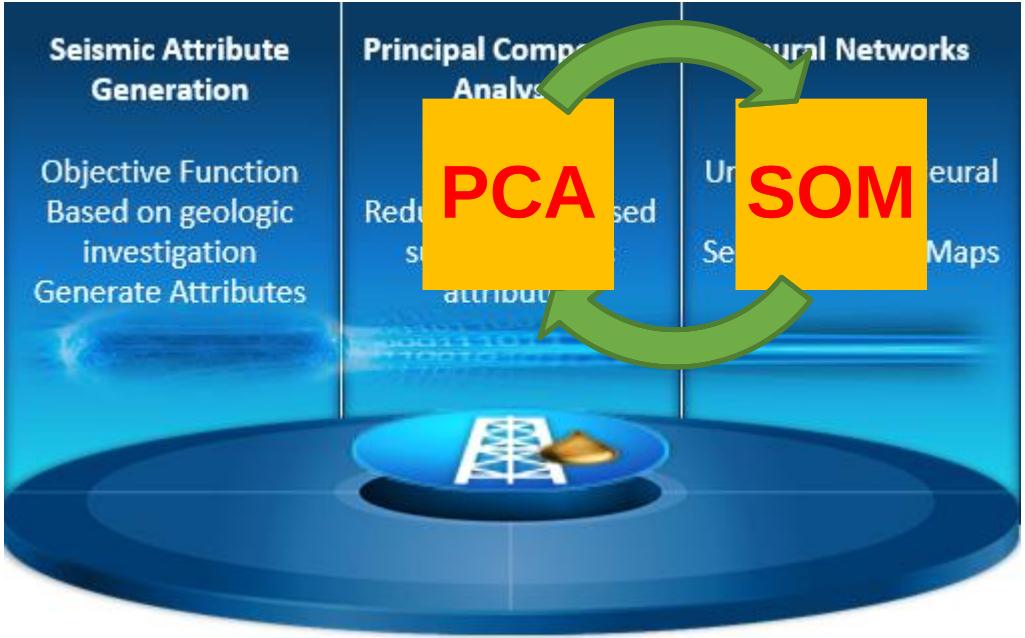
Unsupervised Neural Networks: Self-Organizing Maps

In this research, unsupervised seismic interpretation from multi-attribute data was analyzed by using an ML technique: SOMs.



Case Studies

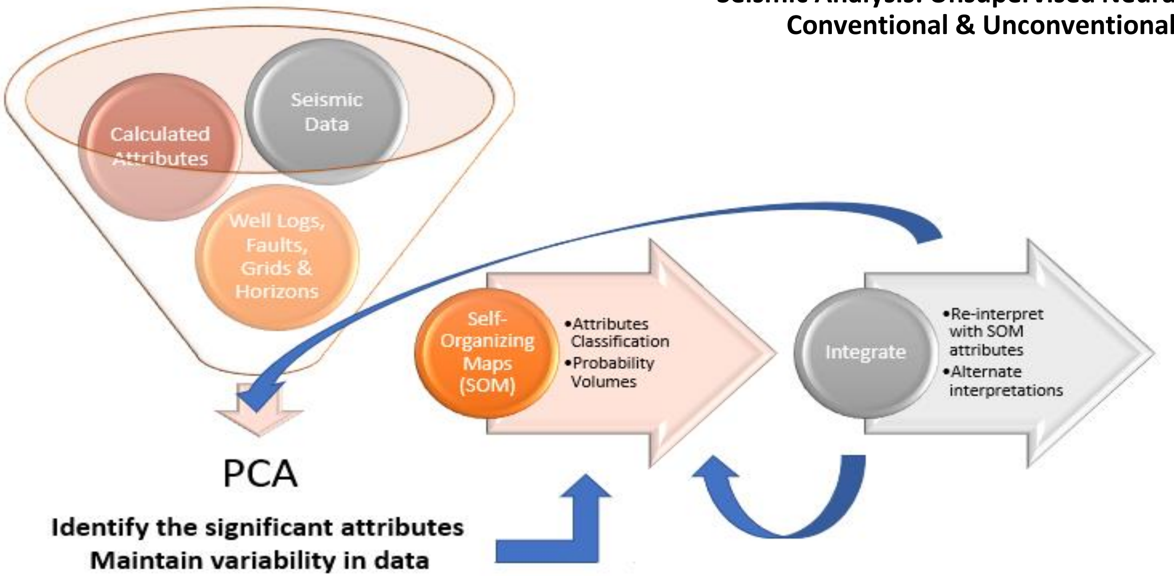
Classifying Multiple Seismic Attributes



Case Studies

Classifying Multiple Seismic Attributes

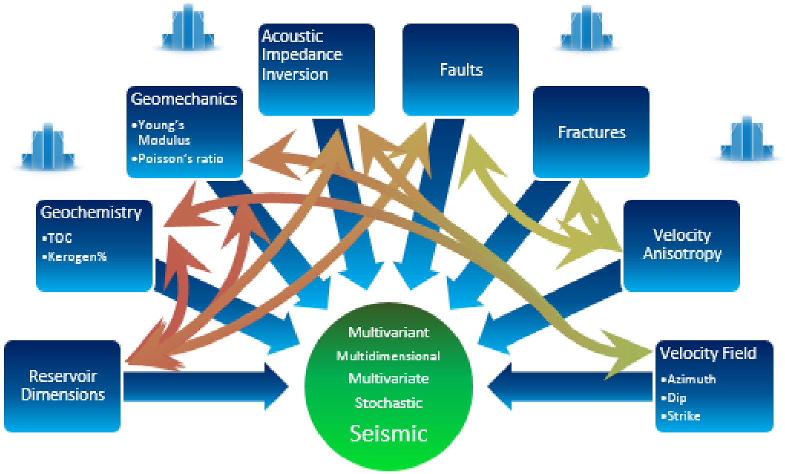
Seismic Analysis: Unsupervised Neural Network & PCA
Conventional & Unconventional Reservoirs



Case Studies

Classifying Multiple Seismic Attributes

- Reservoir geology
    - Thickness and Lateral extent
    - Mineralogy
    - Porosity and Permeability
- Geochemistry
    - Total Organic Content (TOC)
    - Maturity and Kerogen Richness
- Geomechanics
    - Acoustic impedance inversion
    - Young's Modulus
    - Poisson's Ratio (Vp/Vs)
- Faults, Fractures, and Stress regimes
    - Coherency and Curvature
    - Fault Volumes
    - Velocity Anisotropy
    - Stress maps

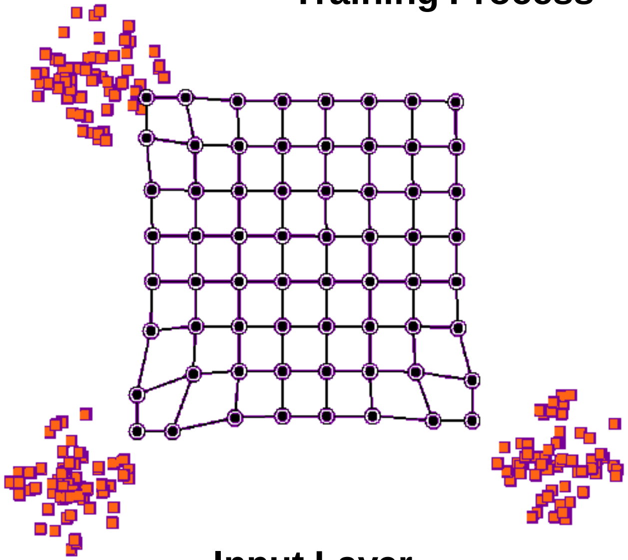


- Pre-Stack Time Migration Traces
    - Attenuation
    - Bandwidth
    - Envelope slope
    - Instantaneous
    - MuRho
    - S-Impedance
    - Trace envelope
    - Young's brittleness
    - Poisson's Ratio
    - Poisson's brittleness
    - Shear Impedance
    - P- impedance
    - Brittleness coefficient
    - Spectral decomposition volumes
    - Instantaneous attributes

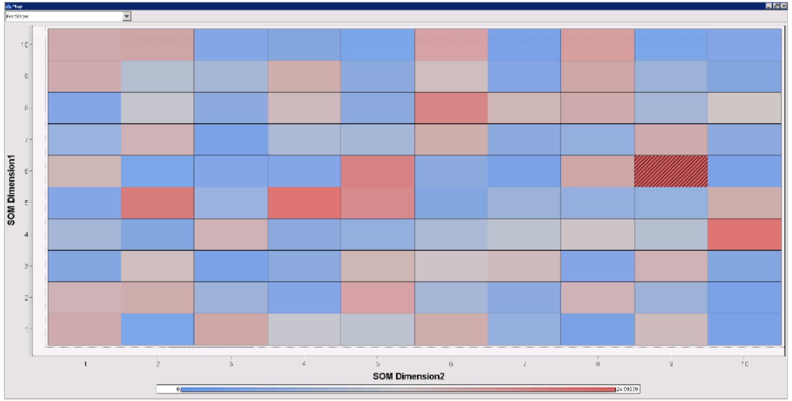
Case Studies

Classifying Multiple Seismic Attributes

Training Process



Input Layer



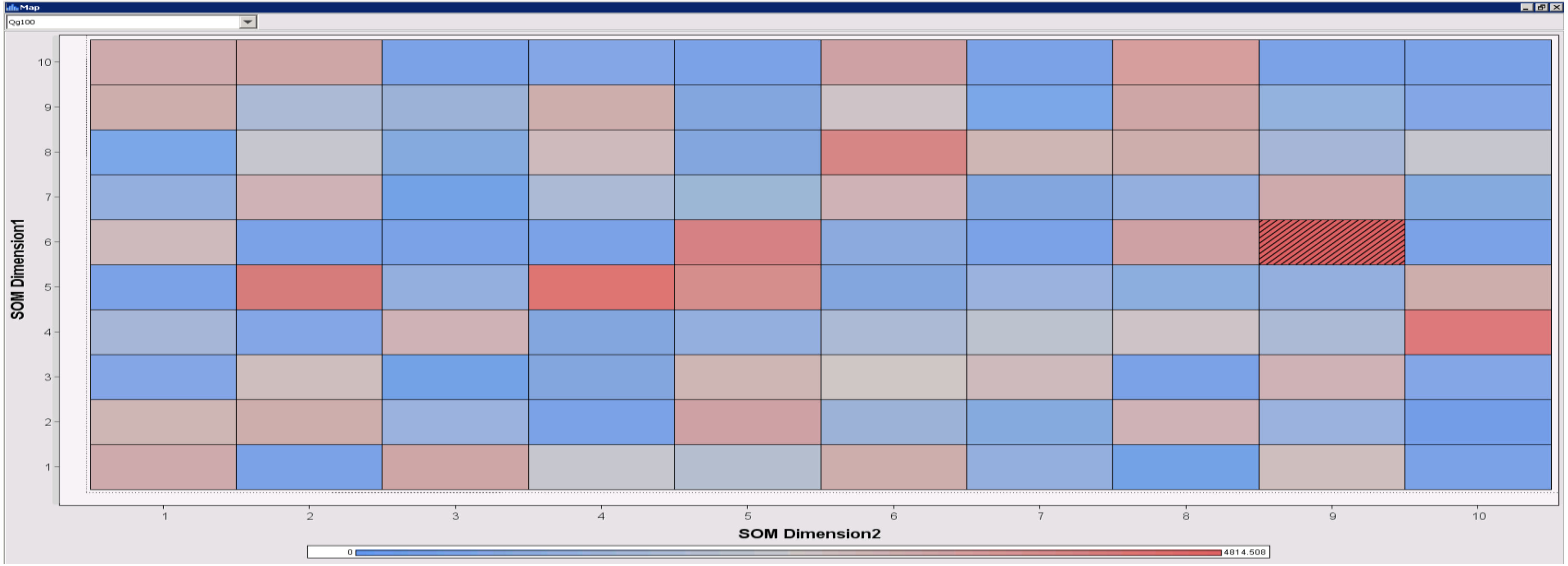
Maplet

Output Layer

Case Studies

Classifying Multiple Seismic Attributes

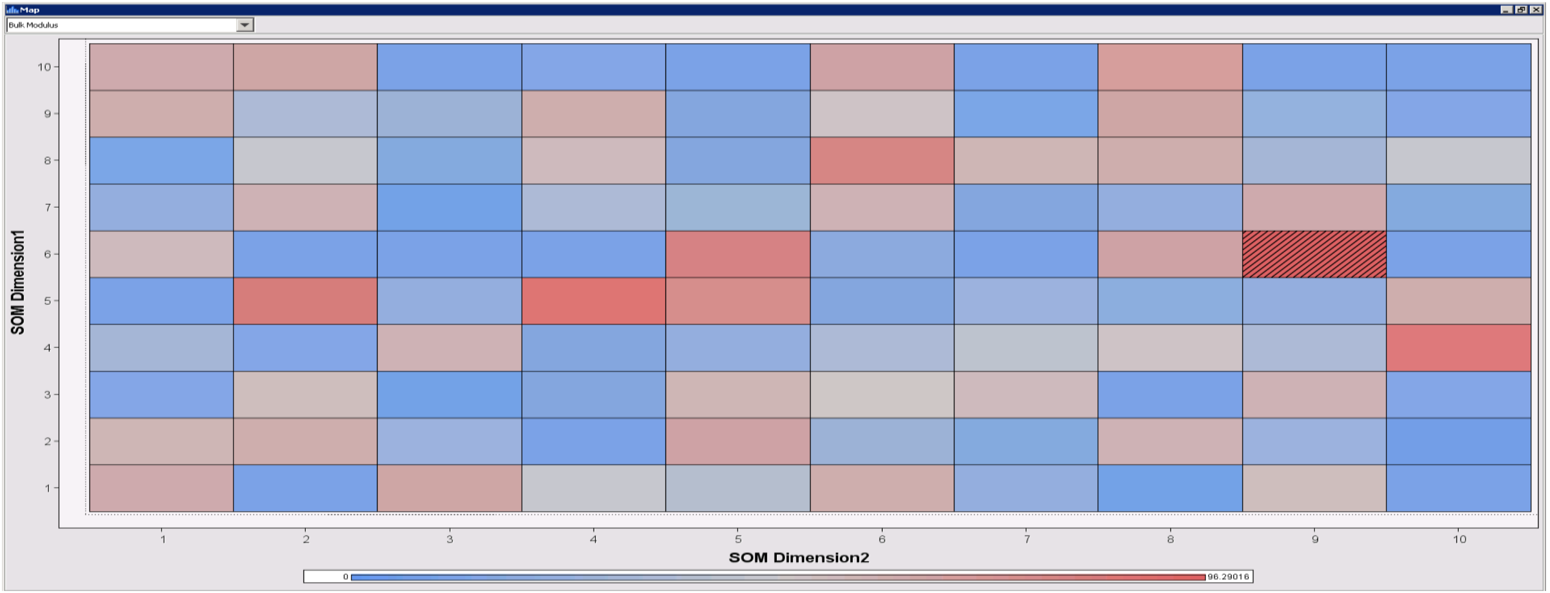
Self-Organizing Maps: Unsupervised NN: Qg100 Maplet



Case Studies

Classifying Multiple Seismic Attributes

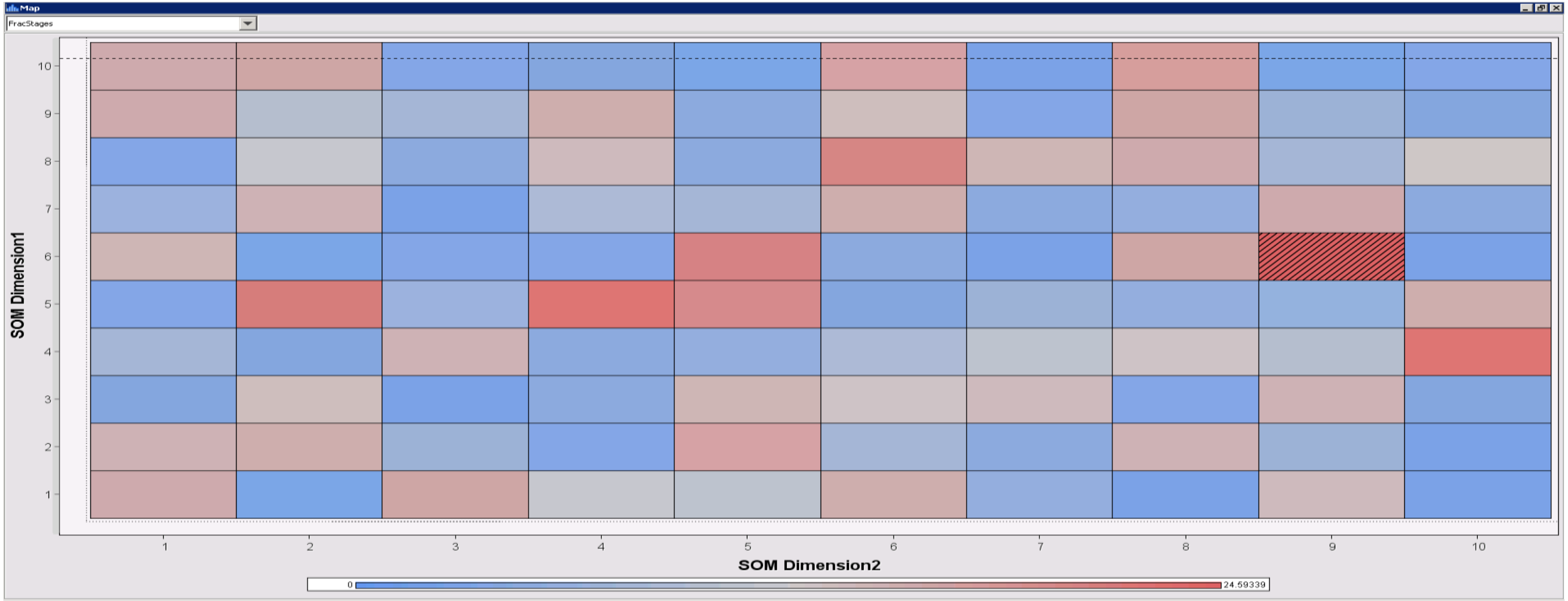
Self-Organizing Maps: Unsupervised NN: Bulk Modulus Maplet



Case Studies

Classifying Multiple Seismic Attributes

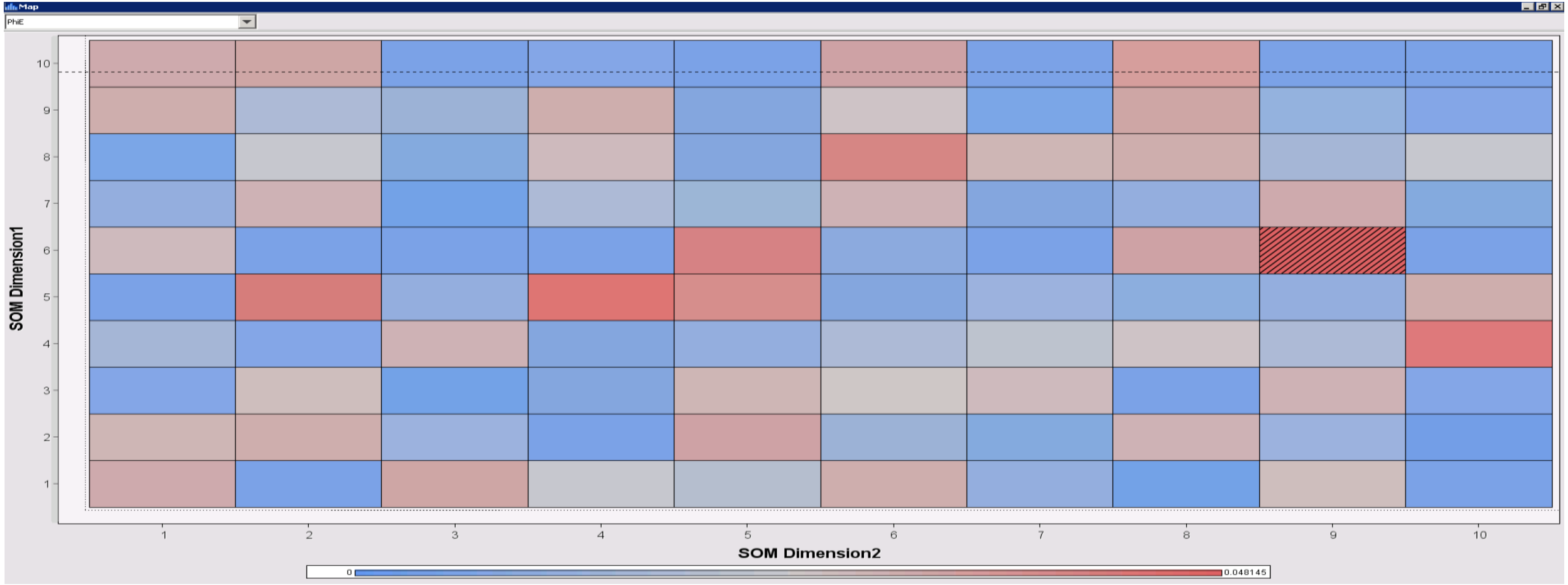
Self-Organizing Maps: Unsupervised NN: Instantaneous Phase Maplet



Case Studies

Classifying Multiple Seismic Attributes

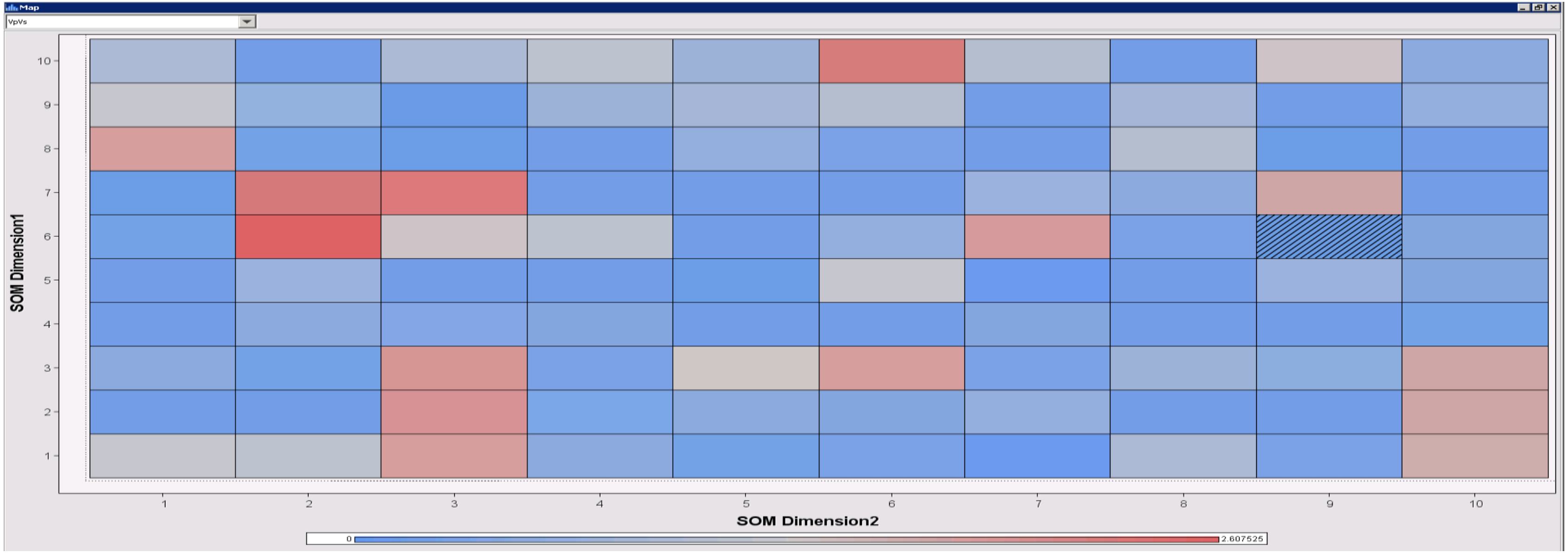
Self-Organizing Maps: Unsupervised NN: Instantaneous Frequency Maplet



Case Studies

Classifying Multiple Seismic Attributes

Self-Organizing Maps: Unsupervised NN: VpVs Maplet

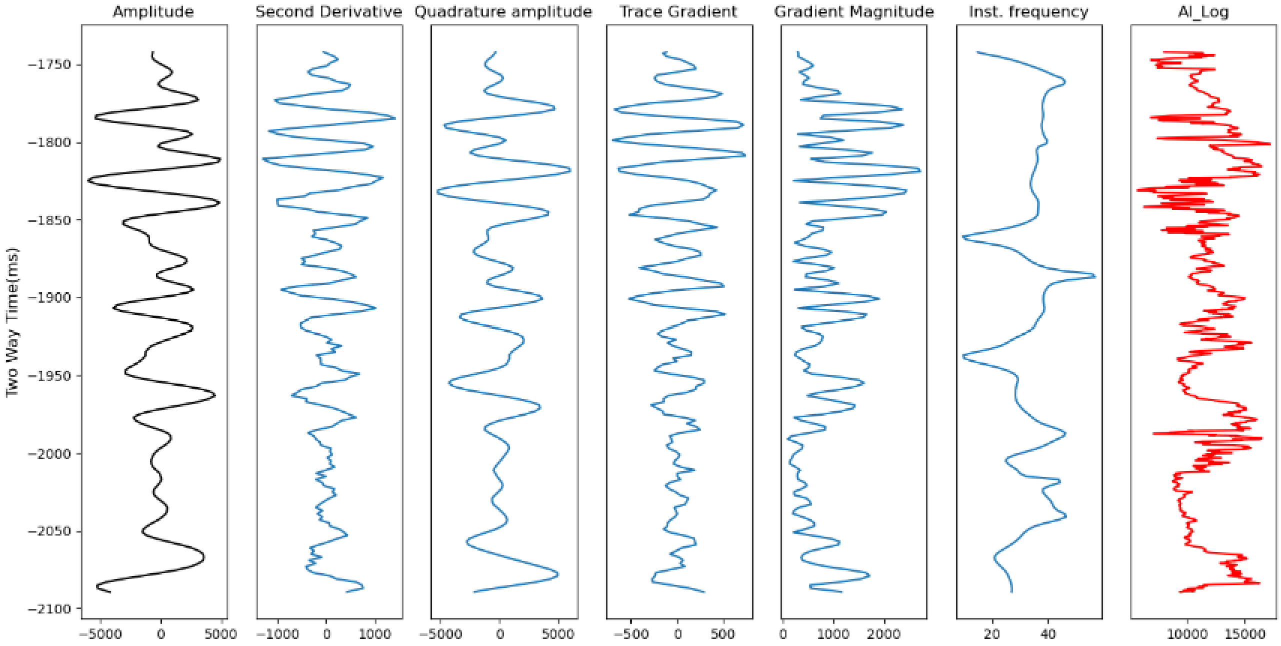


Acoustic Impedance Estimation from Seismic Data Using ML in Well-Log Resolution

Variable/Feature	Description
Depth	Depth in well (m)
Amplitude	Seismic Trace Amplitude
AI_Log	Acoustic Impedance calculated from Sonic and Density logs
Derivative2	Second time derivative of the input seismic volume
QuadrA	Quadrature Amplitude attribute; imaginary part of the analytic signal calculated by phase shifting original trace by 90 degrees
TraceGrad	Gradient along the trace is generated.
GradMag	Magnitude of the instantaneous gradient.
IFreq	Instantaneous frequency, time derivative of phase angle
AI_Inv	Acoustic Impedance Inversion determined from 50 well logs and 3D seismic cube at well locations

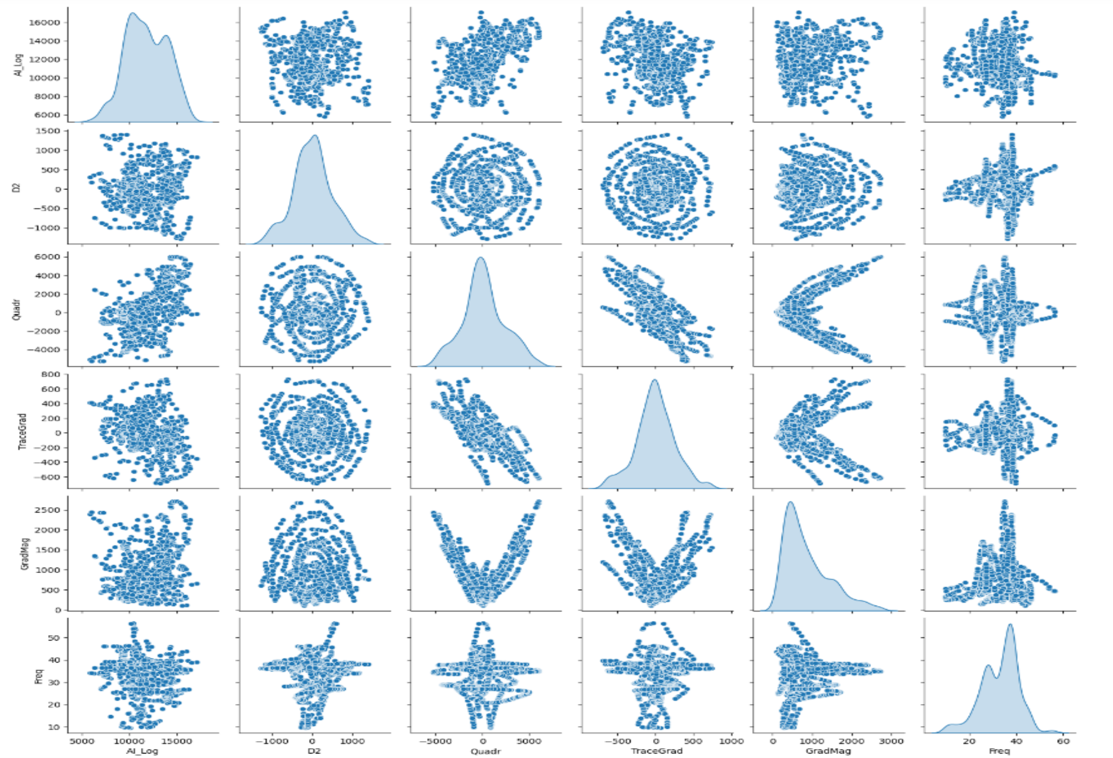
Case Studies

Seismic Attributes



Case Studies

Seismic Attributes: Pair plot



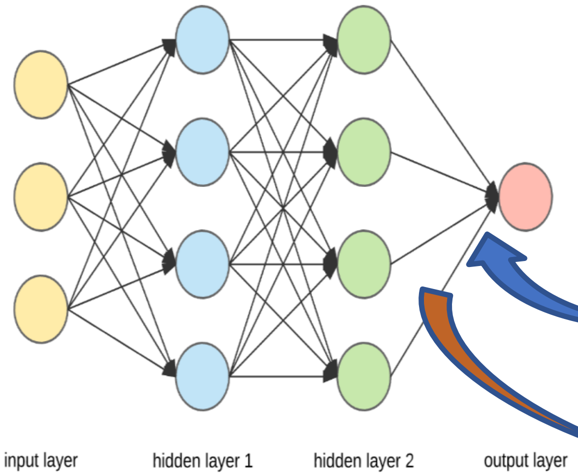
Case Studies

Seismic Attributes: Descriptive + Normalization/Scaling + Neural Network

	count	mean	std	min	25%	50%	75%	max
D2	4154.0	6.233901	483.644848	-1301.850220	-269.884384	0.937068	287.638557	1426.181519
Quadr	4154.0	68.946043	2222.350184	-5258.812012	-1135.513886	-83.835155	1076.680450	6001.270996
TraceGrad	4154.0	-1.895277	261.936301	-692.938904	-146.534729	-10.918543	157.479198	737.373962
GradMag	4154.0	852.799978	580.266098	104.631599	417.605415	653.331024	1130.101135	2719.670166
Freq	4154.0	33.217643	8.299750	9.398151	28.239159	35.046831	38.467464	56.897087

Descriptive Analysis

Sequential Neural Network



Normalization

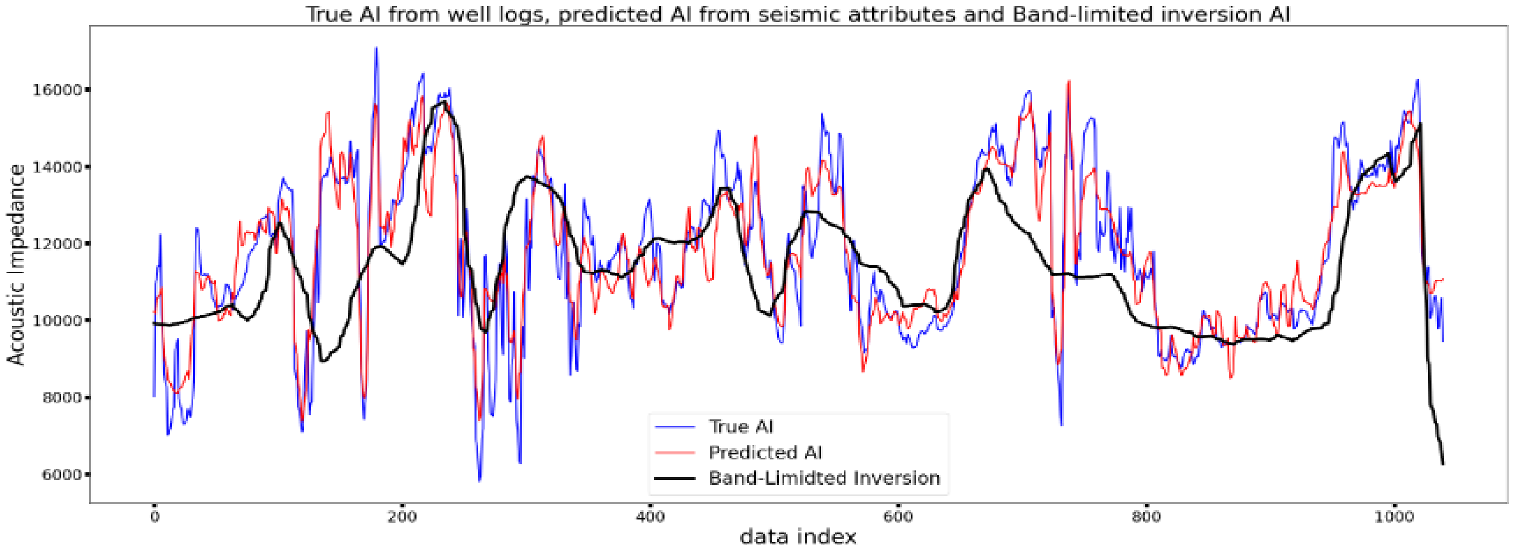
	D2	Quadr	TraceGrad	GradMag	Freq
7127	-0.110911	-0.186294	-0.423878	-0.951503	-2.239415
7134	-0.033997	-0.197319	-0.512131	-0.936389	-2.139127
7148	0.116923	-0.237370	-0.536117	-0.929738	-1.854823
7151	0.148651	-0.248342	-0.521002	-0.931481	-1.782884
7152	0.159211	-0.251999	-0.515962	-0.932060	-1.758911

	loss	mae	mse	val_loss	val_mae	val_mse	epoch
795	79123.000000	207.771271	79123.000000	91841.007812	223.593536	91841.007812	795
796	80983.390625	205.555542	80983.390625	180168.718750	354.522369	180168.718750	796
797	78248.476562	202.256042	78248.476562	142529.468750	289.698700	142529.468750	797
798	80460.343750	205.290558	80460.343750	79298.562500	174.930710	79298.562500	798
799	79822.710938	207.968369	79822.710938	87984.718750	203.654663	87984.718750	799

Model Training Progress

Case Studies

Seismic Attributes: Descriptive + Normalization/Scaling + Neural Network



Module 08

Case Studies: Drilling Program & Completion Study & Virtual Assistant for Fluids and Lithology

MODULE 08

This Module introduces two case studies based on a data-driven analytical methodology to address a business value proposition to optimize drilling and completions and identify fluids and petrophysical properties in an onshore field. We shall follow the SEMMA process introduced in Module 02 under Process and Methodology.

The first case study details a repeatable and scalable data-driven analytical process to optimize drilling and completion strategies in a brownfield with upstream historical datasets.

The second case study under investigation in this Module is Lithology-Fluids and Rocks pattern recognition. We shall discuss an automated workflow with domain input to identify important historical datasets that can predict an African asset's rocks and fluid contents. The analytical workflow follows a SEMMA process to cleanse data, perform Exploratory Data Analysis, and generate Tukey diagrams to understand feature relationships and feature engineering for derived variables and statistical predictive power.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 08

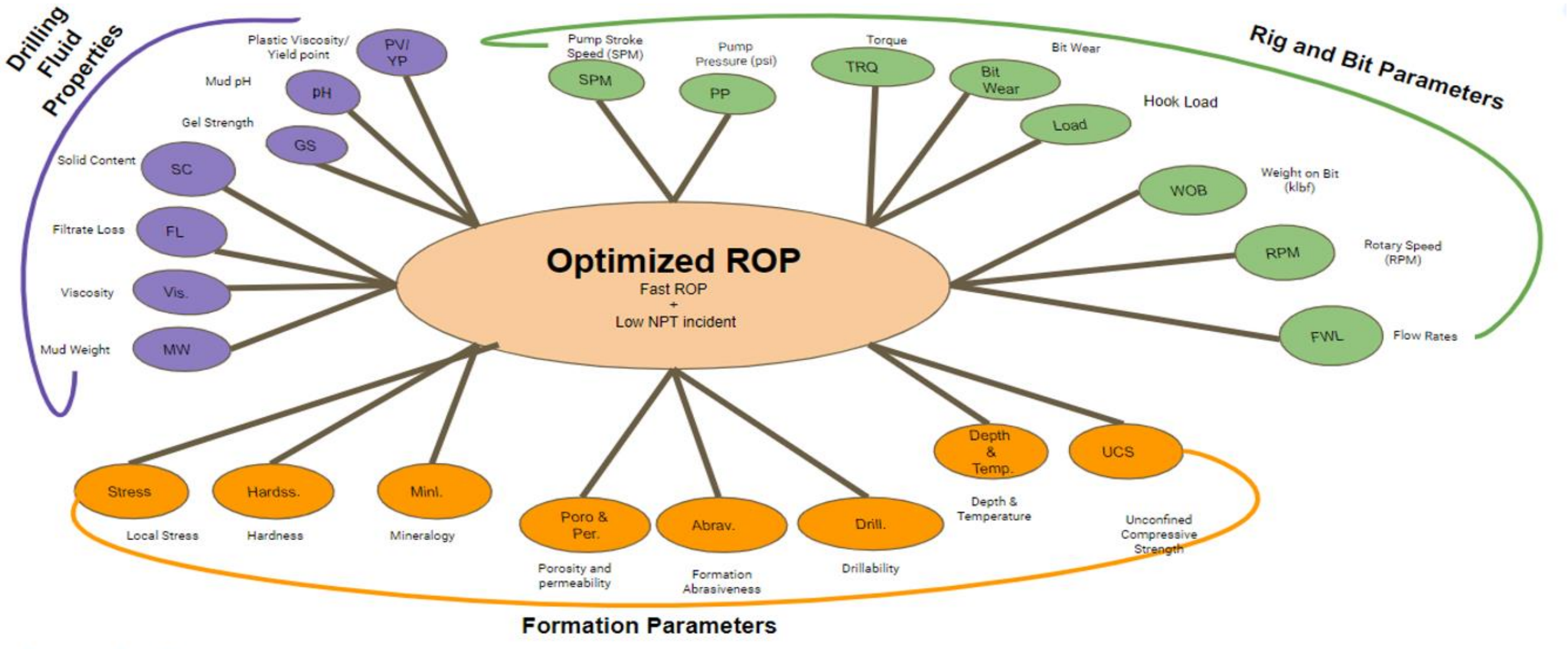
Case Studies: Drilling Program & Completion Study and Virtual Assistant for Fluids and Lithology

LEARNING OBJECTIVES

- ➢ GOAL01: Case Studies – Drilling and Completion in Unconventional Reservoirs
- ➢ GOAL02: Case Studies – Fluids and Lithology Virtual Assistant

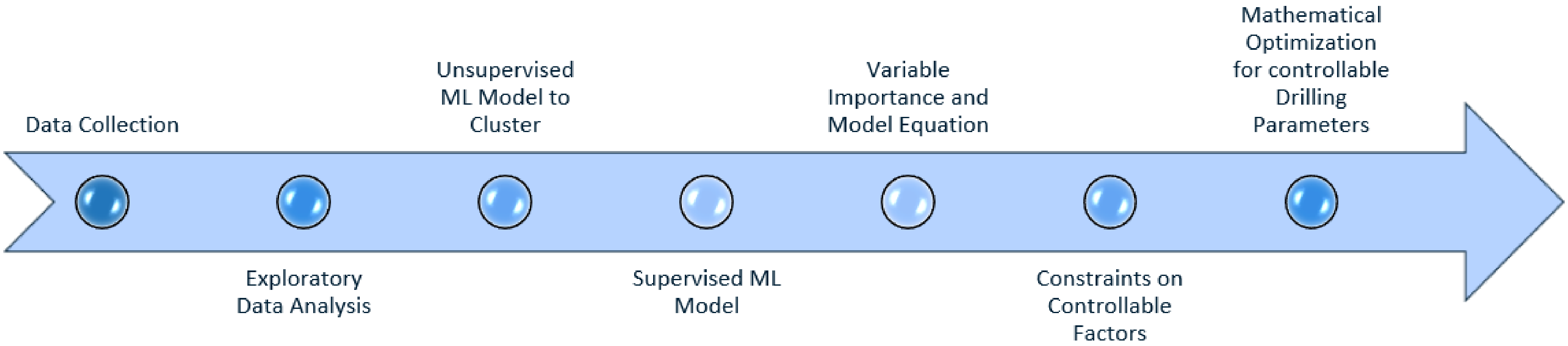
Case Studies

Drilling Optimization Process Workflow



Case Studies

Drilling Optimization Process Workflow

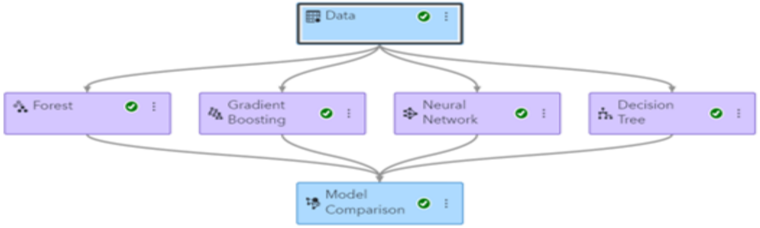


Case Studies

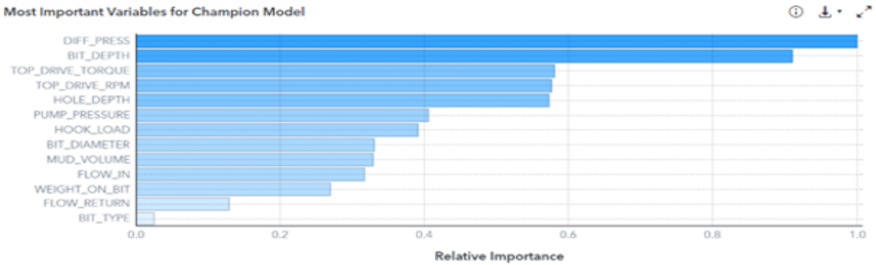
Drilling Optimization Process Workflow

The Process

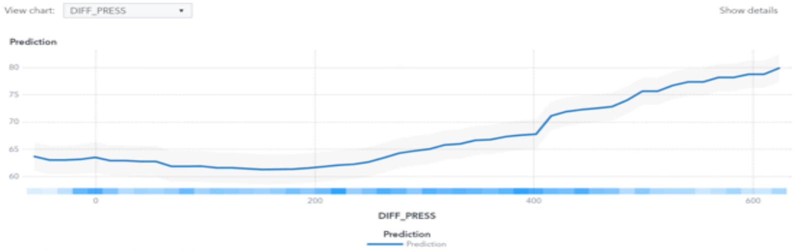
Run pipeline

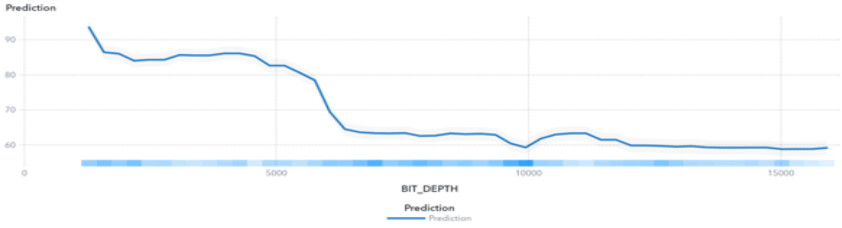


Variable Importance



Explainable AI





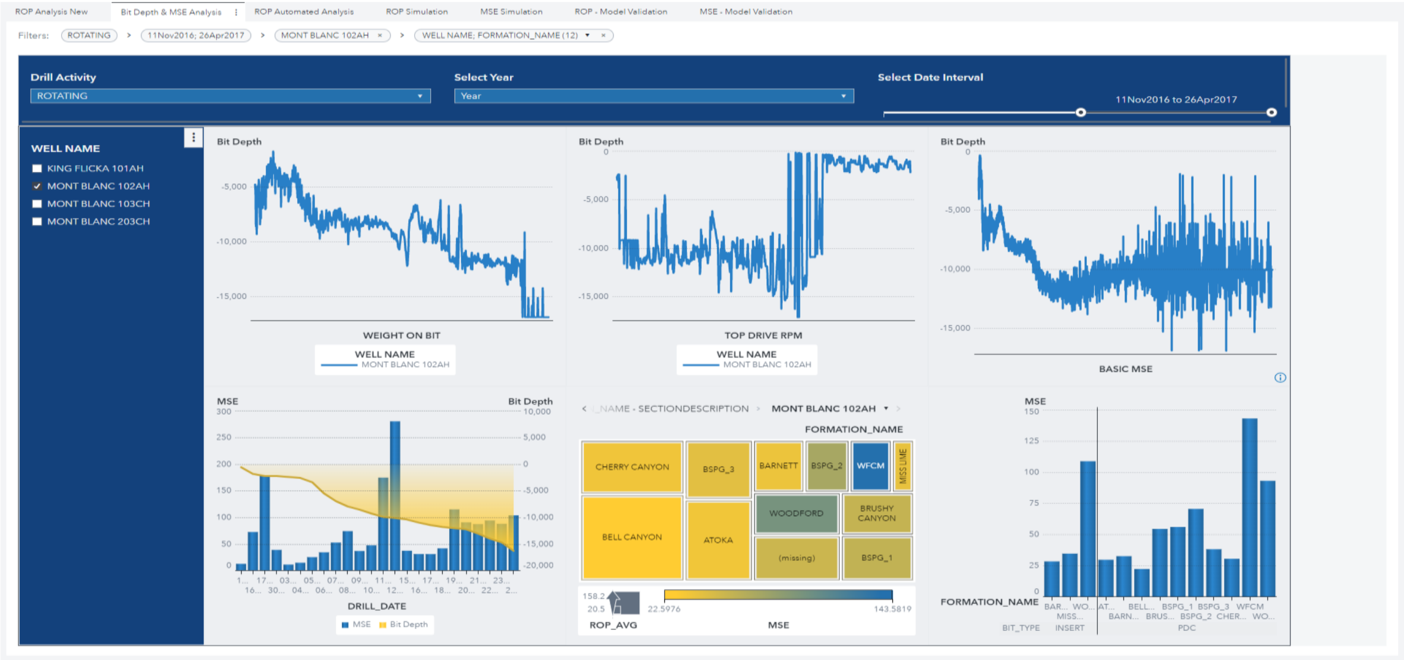
BAUERBERG KLEIN
TRAINING & CONSULTING

Case Studies

Drilling Optimization Process Workflow

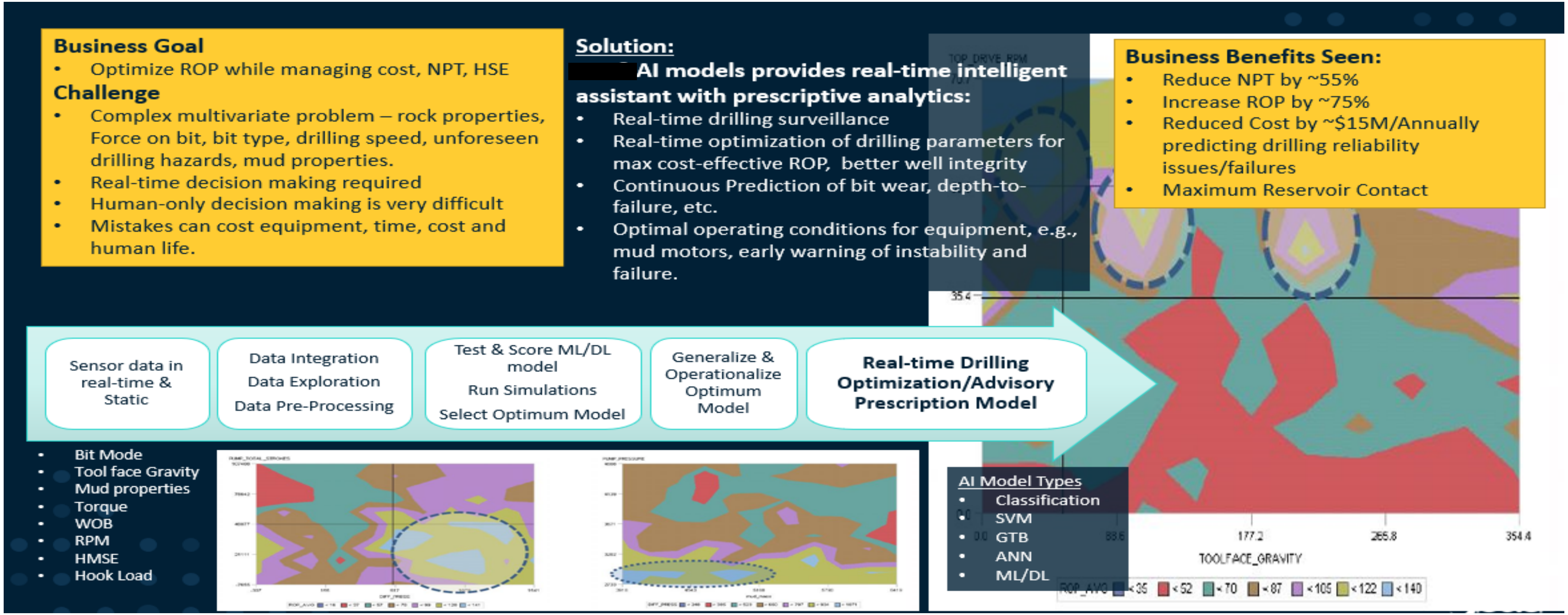
Case Studies

Drilling Optimization Process Workflow



Case Studies

Drilling Optimization Process Workflow



Case Studies

Lithology-Fluids Pattern Recognition

- Well Logs best suited for Lithofacies Classification?
- Classify Lithofacies based on Supervised Learning
  - Support Vector Machine
  - Gradient Tree Boosting
  - Artificial Neural Network
  - Random Forest
- Predicting Stratigraphic Units from Well Logs

Case Studies

Lithology-Fluids Pattern Recognition

Manual interpretation of lithofacies from wireline log data is traditionally performed by an expert, can be subject to biases, and is substantially laborious and time-consuming for large datasets.

Automating the facies classification process using machine learning is a potentially intuitive and efficient way to facilitate facies interpretation based on large-volume data. An expert traditionally performs manual interpretation of lithofacies from wireline log data.

- Multi-Well Log Analysis
- Lithology Characteristization
- Automated Lithology Classification
- Lithology Classification Model

The input parameters used to train AI models include LWD, MWD, Drilling Data, Gas Components data.

Data will be from multiple reservoirs, fields, formations, wells

Data would be for more than 75 wells.

Actual Lithology and fluid tags would be identified and used as the target variable.

Additional Derived variables would be created that would be able to explain the lithology facies better.

Different classes of lithology facies and fluid-type relationships will be modeled.

The automated Machine Learning process to predict field pattern type recognition

The AI assistant will suggest the best approach to follow to the domain experts.

Provide workflow automation that reduces work time and raises efficiency with real-time interpretation.

Case Studies

Lithology-Fluids Pattern Recognition

Typical Input Data: Facies-Fluids: *Feature Engineering*

Table 1 – LWD Curves

Logs While Drilling (LWD)	Gas Components	Drilling Params.
Gamma Ray	Total Gas (TG)	Weight on Bit (WOB)
Resistivity Shallow	Methane (C1)	Rate of Penetration (ROP)
Resistivity Deep	Ethane (C2)	
Neutron	Propane (C3)	
Density	IsoButane (iC4)	
	NormalButane (nC4)	
	IsoPentane (iC5)	
	NormalPentane (nC5)	

Table 2 – Derived Features

Derived	Forumla
LHR	C1+C2/C3+iC4+nC4+iC5+nC5
CH	iC4+nC4+iC5+nC5/C3
WH	C2+C3+iC4+nC4+iC5+nC5/C1+C2+C3+iC4+nC4+iC5+nC5
C1/C2	C1/C2
C1/C3	C1/C3
C2/C3	C2/C3
%C1	(C1/C1+C2+C3+iC4+nC4+iC5+nC5)*100
%C2	(C2/C1+C2+C3+iC4+nC4+iC5+nC5)*100
%C3	(C3/C1+C2+C3+iC4+nC4+iC5+nC5)*100
%C4	(iC4+nC4/C1+C2+C3+iC4+nC4+iC5+nC5)*100
%C5	(iC5+nC5/C1+C2+C3+iC4+nC4+iC5+nC5)*100

IPTC
International Petroleum Technology Conference

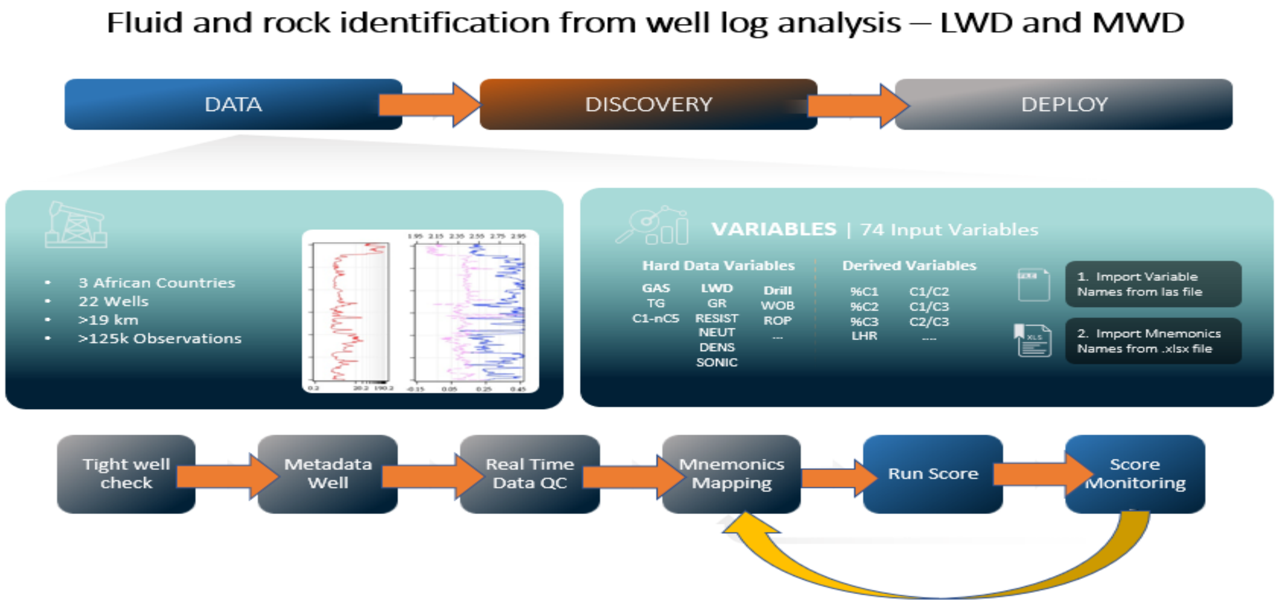
IPTC-19701-MS

Artificial Intelligence and Machine Learning Techniques Provide Operations Geologists With an Automated and Reliable Lithology-Fluid Pattern Recognition Assistant: A Case History in a Clastic Reservoir in West Africa

Davide Baldini and Luca Piazza, ENI SpA; Luca Barbanotti, SAS Institute

Case Studies

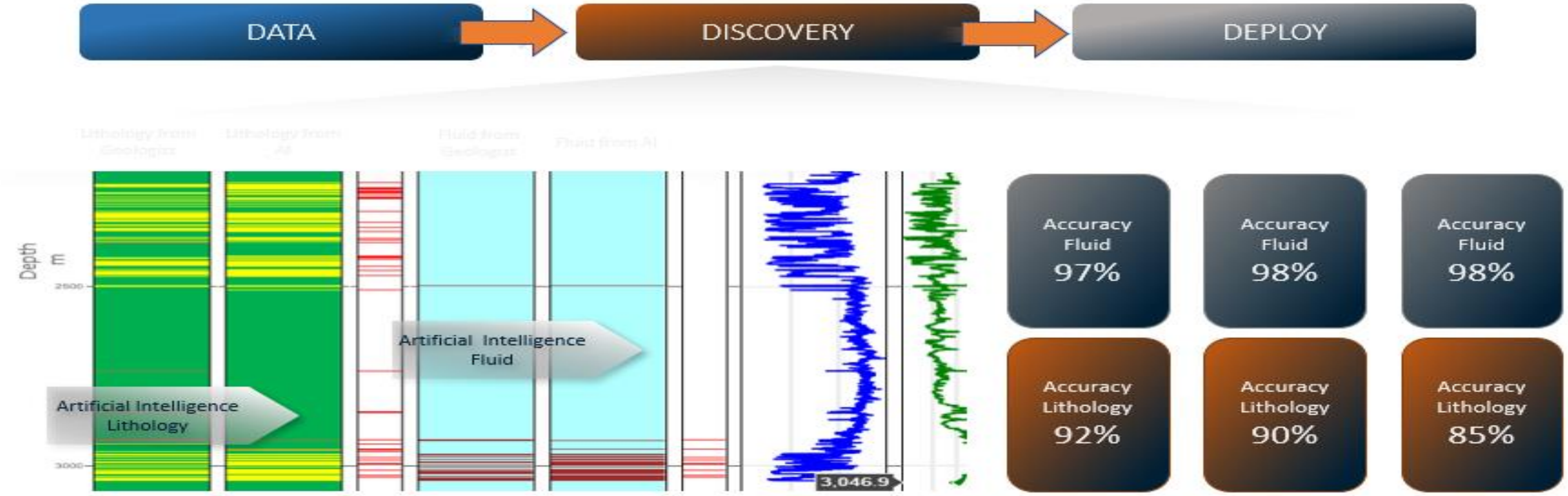
Lithology-Fluids Pattern Recognition



Case Studies

Lithology-Fluids Pattern Recognition

Fluid and rock identification from well log analysis – LWD and MWD



Module 09

Case Studies: Time-Series Analysis and Production Forecasting

MODULE 09

This Module introduces the six principles of forecasting in a time-series dataset.

We shall implement these principles in the case study to optimize production data collected in a brownfield.

The SEMMA process takes on a journey to analyze temporal data using several time-series statistical and machine-learning methods. The well, reservoir, and field production forecasting uses spatial and temporal data to optimize the re-engineering of a brownfield.

We shall show the use of both supervised and unsupervised techniques. And we shall introduce a deep neural network architecture called a Recurrent Neural Network for time-series analysis. RNNs are discussed in Module 05.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 09

Case Studies: Forecasting Principles & Production Forecasting Techniques

LEARNING OBJECTIVES

- ➢ GOAL01: Forecasting - Six Principles
- ➢ GOAL02: Forecasting Techniques & Forecasting Data-Driven Workflows
- ➢ GOAL03: Case Study: NOC Re-engineering Brownfield

Case Studies

Six Principles of Forecasting

1. Forecasting is a stochastic problem
2. All forecasts are wrong
3. Some forecasts are useful
4. All forecasts can be improved
5. Forecast accuracy is never guaranteed
6. Having a second opinion is preferred

Case Studies

Production Forecasting

A time series is a sequence of observations *Y1*, … *Yt-1*, *Yt*, where the observation at time *t* is denoted by *Yt*.

Rule Induction for TS Forecasting

- Exponential Smoothing (ES)
- Auto-Regressive Integrated Moving Average (ARIMA)
- Random Walk (RW)
- Neural Networks (NN)

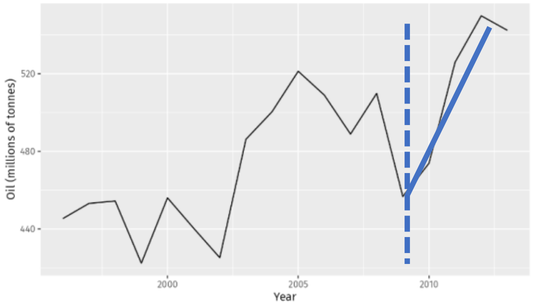
Case Studies

Exponential Smoothing

Simple and low cost
- Less Data Memory Storage
- Fast Computational Speed

Not as accurate:
- ARIMA
- FFNN

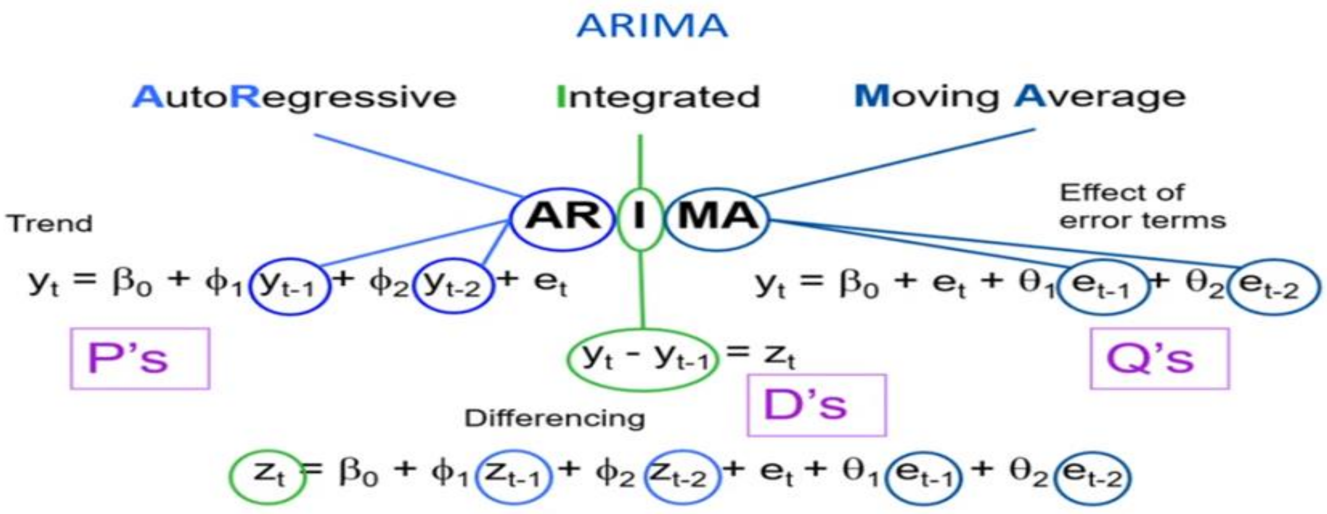
- Simple Exponential Smoothing (SES)
- Double Exponential Smoothing (DES)
- Triple Exponential Smoothing (Holt-Winters Method)



Oil Production in Aramco from 1996 - 2013

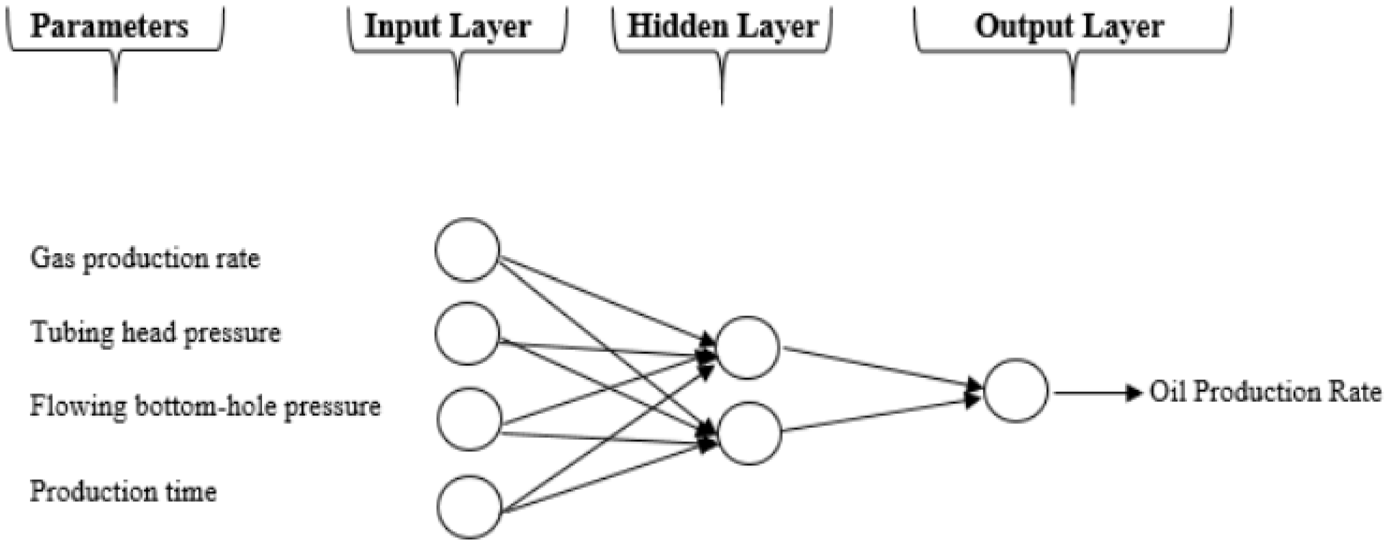
Case Studies

AutoRegressive Integrated Moving Average



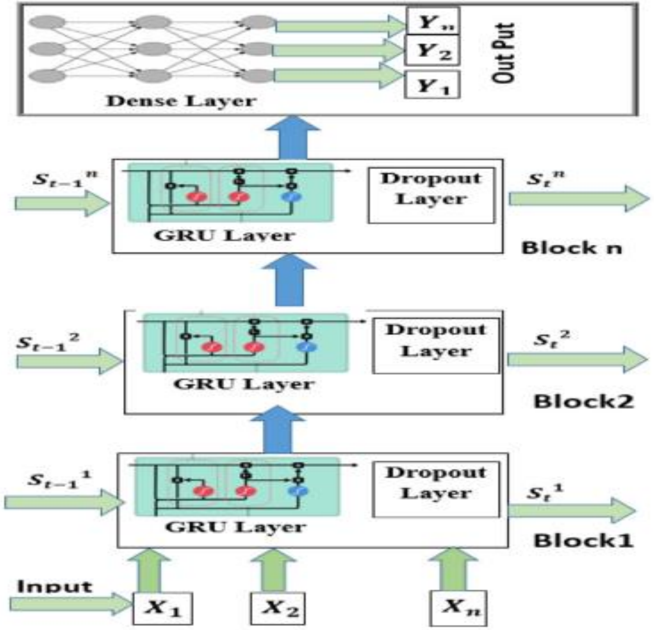
Case Studies

Artificial Neural Networks - ANNs



Case Studies

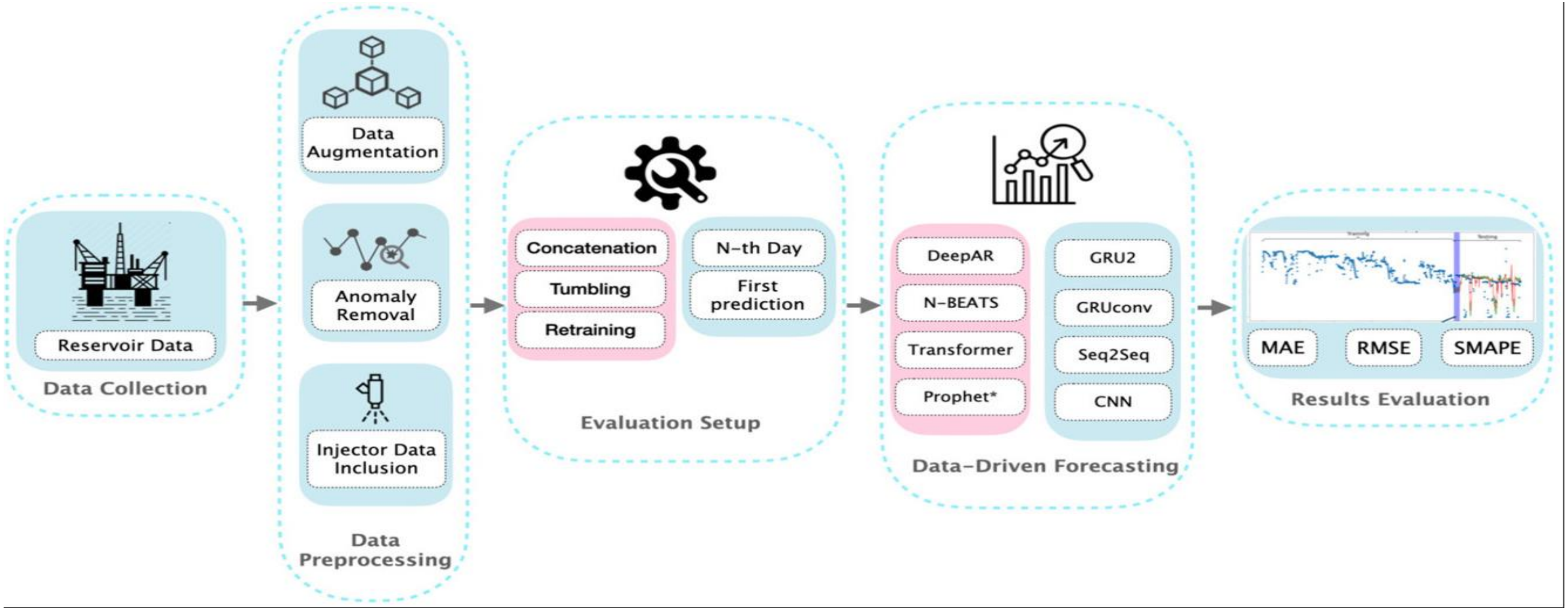
Deep Gated Recurrent Unit Network (DGRU - Deep Learning Neural Networks)



The proposed model can handle the temporal dependencies of complex time-series data at a deep level. It consists of stacks of several layers, where each layer solves part of the task and passes the results to the next layer. Since each layer combines the learned representations of the previous layer and feeds them to a higher layer, better representations of the data can be achieved in the model.

Case Studies

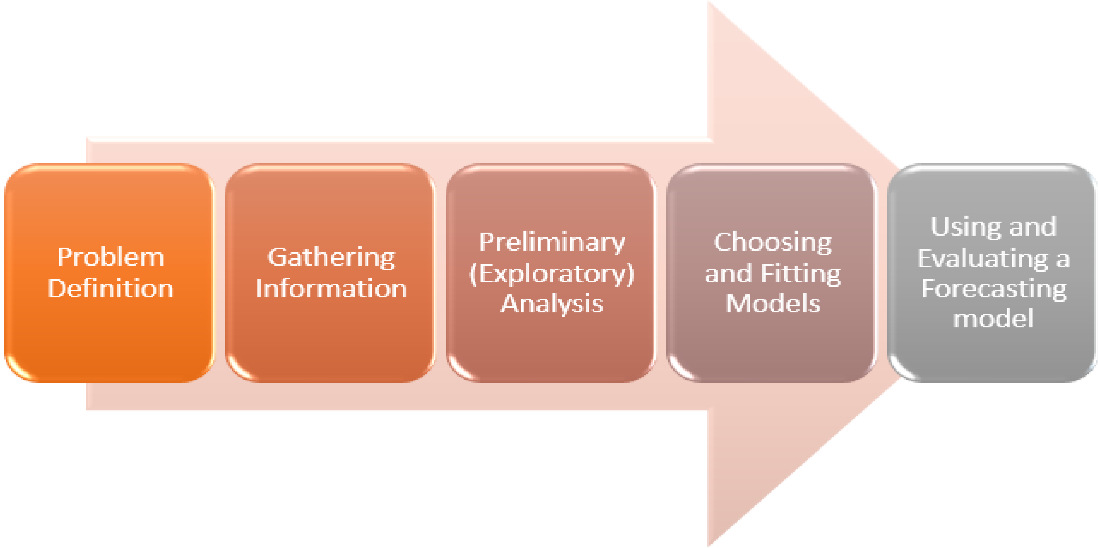
Repeatable and Scalable Methodology for Forecasting



Case Studies

Production Forecasting

Five Basic Steps in a Forecasting Task



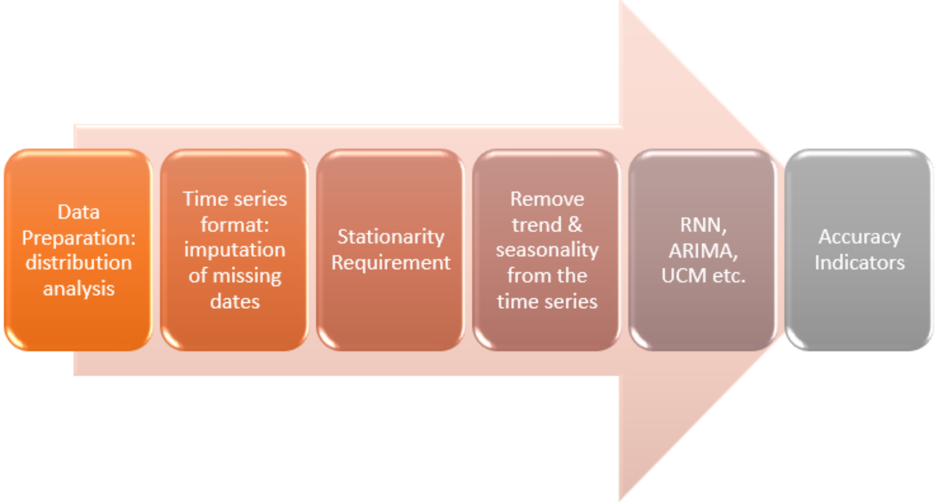
Case Studies

Production Forecasting

Time-Series Data Forecasting

Well Production Workflow

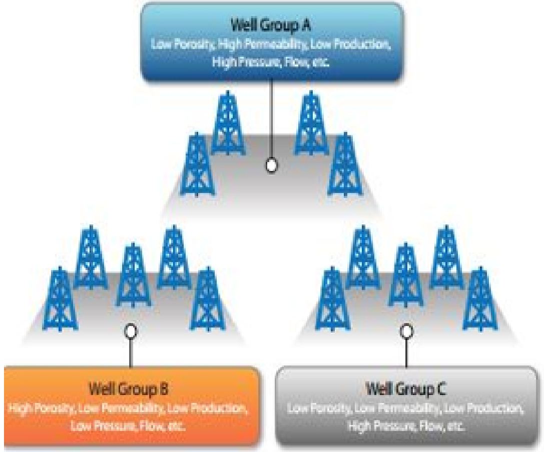
Data-driven analytical workflow to forecast oil/gas production in a well.



Case Studies

Production Forecasting

CLUSTER ANALYSIS: WELL PROFILES

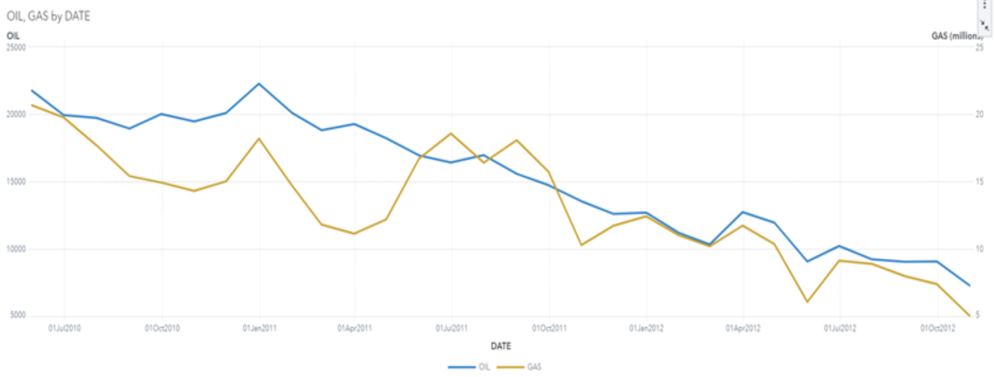


Cumulative oil or gas production
Water cut (Percentage determined by water production/liquid production)
B exponent (Decline type curve)
Initial rate of decline
Initial rate of production
Geomechanics and Petrophysical Properties
Geological Parameters
Operational Parameters
Completions Design

Case Studies

Production Forecasting – Let’s “stationarize’ our temporal data

Oil & Gas Production Decline



DATE	OIL	GAS
31May2010	21760.419361	20678171.017
30Jun2010	19950.990314	19770066.046
31Jul2010	19739.3784	17706084.114
31Aug2010	18953.365184	15434317.195
30Sep2010	20029.094784	14955523.1
31Oct2010	19481.346546	14335004.844

Stationary Time Series	Non-Stationary Time Series
Statistical properties of a stationary time series are independent of the point in time where it is observed.	Statistical properties of a non-stationary time series is a function of time where it is observed.
Mean, variance and other statistics of a stationary time series remains constant. Hence, the conclusions from the analysis of stationary series is reliable.	Mean, variance and other statistics of a non-stationary time series changes with time. Hence, the conclusions from the analysis of a non-stationary series might be misleading.
A stationary time series always reverts to the long-term mean.	A non-stationary time series does not revert to the long term mean.
A stationary time series will not have trends, seasonality, etc.	Presence of trends, seasonality makes a series non-stationary.

Module 10
Time-Series Analysis and Production Forecasting

MODULE 10

This Module introduces a case study to optimize the technical potential of a National Oil Company (NOC.) Technical Potential (TP) forms the basis for future expectations by defining what is achievable and thus highlights the gap between potential performance and what is realized in hydrocarbon production. This knowledge transforms into initiatives that drive the processes for minimizing the gap. Assessment and forecasting TP workflows provide the appropriate tools for NOCs to drive the operator contractors towards better performance targets.

We shall demonstrate a case study to forecast the fluid rates in a brownfield, analyzing historical production data of wells across multiple reservoirs. The proposed methodology is a Deep Learning Long-Short Term Memory architecture.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 10

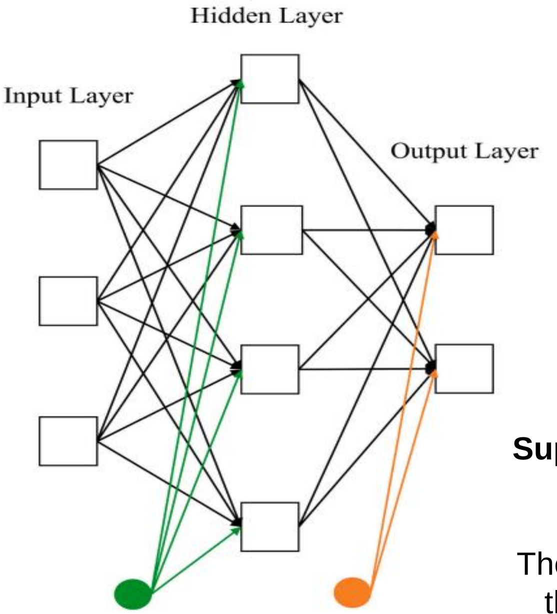
Case Studies: Time-Series Analysis and Production Forecasting

LEARNING OBJECTIVES

- GOAL01: Time Series Patterns Used in Forecasting
- GOAL02: Case Study: Forecasting Well Data from the Volve Field
- GOAL03: Case Study: Production Forecasting using DL Models
- GOAL04: Case Study: Drilling Time series identifying Lost Circulation

Case Studies

Well Production Forecasting – Volve Field



Feed-Forward Neural Network

A Feed-Forward Neural Network (FNN) is an ML algorithm formulated based on biological neural network functionalities. FNN comprises many calculating units known as artificial neurons or nodes. It has been demonstrated to be more successful in approximating the complex non-linear relationships between input and output vectors of a database than the conventional regression methods.

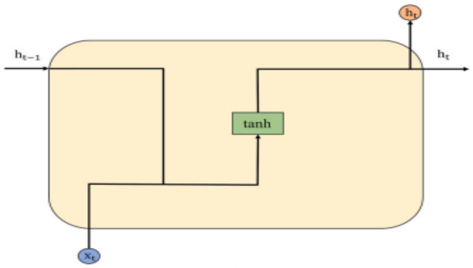
Support Vector Regression is a subset of a Support Vector Machine for regression analysis.

The fundamental idea regarding the mechanism of **Particle Swarm Optimization** is that each particle corresponds to a potential solution to an optimization problem.

Case Studies

Well Production Forecasting – Volve Field

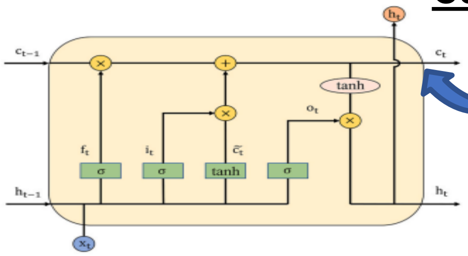
Recurrent Neural Networks



(a)

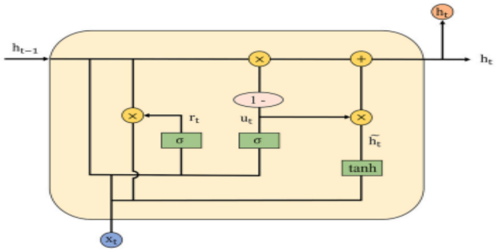
Simple RNN

Cell State "ct"



(b)

Long Short-Term Memory (LSTM)



Gated Recurrent Unit (GRU)

RNN is a subset of ANN established to handle the input data with sequential characteristics. Fundamentally, RNN can preserve any previous information to the current task, and such ability widens its application in different aspects, including time series analysis.

Case Studies

Well Production Forecasting – Volve Field

Training and Blind Validation Results

Datasets	Models	R^2	RMSE
Training	SVR-TE	0.9951	13.88
	SVR-PSO	0.9944	14.68
	FNN-BP	0.9948	14.00
	FNN-PSO	0.9945	14.92
	Simple RNN	0.9945	14.46
	LSTM	0.9962	12.03
	GRU	0.9962	12.17
Validation	SVR-TE	0.9880	21.37
	SVR-PSO	0.9889	20.79
	FNN-BP	0.9911	19.13
	FNN-PSO	0.9923	15.75
	Simple RNN	0.9921	18.27
	LSTM	0.9910	19.51
	GRU	0.9940	15.75
Testing	SVR-TE	0.9764	30.83
	SVR-PSO	0.9936	16.61
	FNN-BP	0.9936	16.44
	FNN-PSO	0.9898	19.91
	Simple RNN	0.9941	15.37
	LSTM	0.9922	17.64
	GRU	0.9915	18.24

Best

Datasets	Models	R^2	RMSE
Blind Validation	SVR-TE	0.9476	7.34
	SVR-PSO	0.9644	6.04
	FNN-BP	0.9538	6.89
	FNN-PSO	0.9574	6.61
	Simple RNN	0.9665	5.87
	LSTM	0.9712	5.45
	GRU	0.9700	5.56

Datasets	Models	R^2	RMSE
All	SVR-TE	0.9935	16.52
	SVR-PSO	0.9952	14.21
	FNN-BP	0.9956	13.65
	FNN-PSO	0.9952	14.15
	Simple RNN	0.9957	13.51
	LSTM	0.9961	12.69
	GRU	0.9964	12.28

Best

Case Studies

Production Forecasting

Data Input Space

- Location data for each well
- Historical production of oil, water and gas for each well
- Historical pressures in some wells allow to build an average behavior for each reservoir
- Petrophysical data

Alias	Well	Reservoir	Date	Days	Oil_Rate__BOPD	Water_Rate__BWPD	Gas_Rate__MSCF_D	K__mD	h__ft	Depth_avg__ft	P__psi	Pwf__psi
BCS0006:M06	BCS0006	M06	19633	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19664	15	299.9	0.9	62.9	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19694	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19725	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19756	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19784	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19815	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19845	0	0	0	0	3.206	189.3	9932	4950	4950
BCS0006:M06	BCS0006	M06	19876	4	346.3	4.5	103.5	3.206	189.3	9932	4950	4806
BCS0006:M06	BCS0006	M06	19906	4	377.3	5	112.8	3.206	189.3	9932	4950	4793
BCS0006:M06	BCS0006	M06	19937	16	391.6	0.8	48.3	3.206	189.3	9932	4950	4787
BCS0006:M06	BCS0006	M06	19968	20	545.9	1.1	74.4	3.206	189.3	9932	4950	4723

- Prepare time series data for training an RNN forecasting model
- Implement an RNN model to predict the next 3 steps ahead (time *t+1* to *t+3*) in the time series
- Use a simple encoder-decoder approach in which the final hidden state of the encoder is replicated across each time step of the decoder
- Enable early stopping to reduce the likelihood of model overfitting
- Evaluate the model on a test dataset

	Date	K, mD	h, ft	P, psi	Pwf, psi
9071	41944	1.24	77.0	2977	1907
9072	41974	1.24	77.0	2976	1687
9073	42005	1.24	77.0	2974	683
9074	42036	1.24	77.0	2973	921
9075	42064	1.24	77.0	2972	662

Define input and outputs for the model

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 1024)              6144

 dense_1 (Dense)             (None, 512)               524800

 dense_2 (Dense)             (None, 256)               131328

 dense_3 (Dense)             (None, 128)               32896

 dense_4 (Dense)             (None, 3)                 387

=================================================================
Total params: 695,555
Trainable params: 695,555
Non-trainable params: 0
_________________________________________________________________
```

Print model summary

Case Studies

Production Forecasting: Hyperparameters

Well Production Workflow: Predictive Accuracy Indicators

Data-driven analytical workflow to forecast oil/gas production in a well.

"The function we want to minimize or maximize is called the objective function or criterion. When we are minimizing it, we may also call it the cost function, loss function, or error function.".

I use **MSE**: In the case of regression problems where a quantity is predicted, it is common to use the mean squared error (MSE) loss function instead. This returns a very good accuracy for your model.

- Maximum A Posteriori (MAP), a Bayesian method
- Maximum Likelihood Estimation (MLE), frequentist method

It may be more important to report the "accuracy" and "root mean squared error" (RMSE) for models used for classification and regression, respectively.

Loss function evaluates/diagnoses model learning
Metric measures model accuracy

- Loss: Used to evaluate and diagnose model optimization only.
- Metric: Used to evaluate and choose models in the case study context.

```
deep_model.compile(optimizer='adam',
                   loss='mse',
                   metrics=['accuracy'])
```

Compile and **Fit** the model to the training data

```
deep_model.fit(
    X_train,
    y_train,
    epochs=1000,
    shuffle=True,
    verbose=2
)
```

Case Studies

Production Forecasting: Hyperparameters

Optimizer

The optimizer performs the necessary computations to adapt to the network's weight and bias variables during training. Those computations invoke the calculation of gradients that indicate the direction in which the weights and biases must be changed during training to minimize the network's cost function.

The goal of machine learning and deep learning is to **reduce the difference between the predicted output and the actual output**. This is also known as a **cost function or loss function**. Cost functions are convex functions.

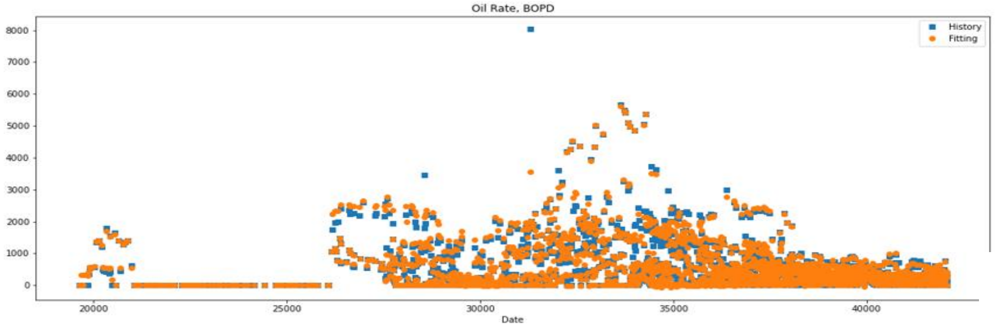
We aim to **minimize** the cost function by finding the optimized weight value. We also need to ensure that the algorithm **generalizes** well. This will help better predict the data that was not seen before.

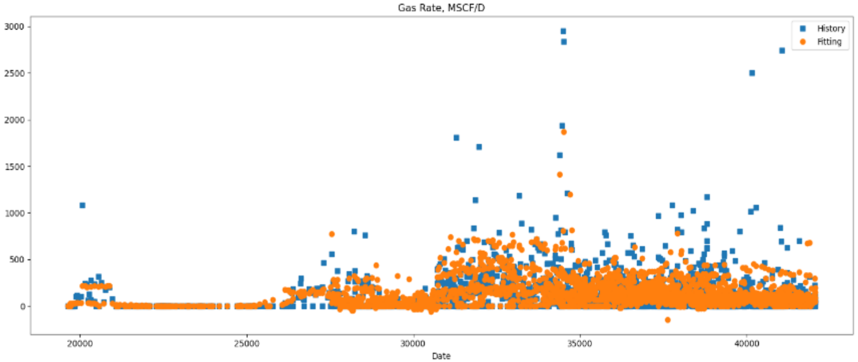
Nadam-Nesterov-accelerated Adaptive Moment Estimation

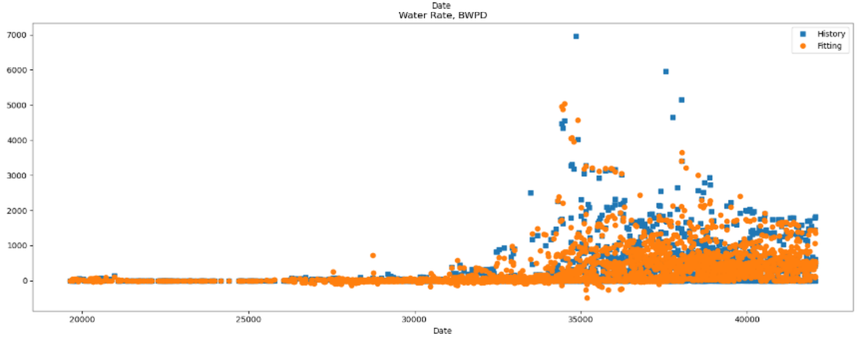
- Nadam combines NAG and Adam
- Nadam is employed for noisy gradients or gradients with high curvatures
- The learning process is accelerated by summing up the exponential decay of the moving averages for the previous and current gradient

Case Studies

Production Forecasting: Oil, Gas and Water







- **Sequential Model – Encoder - Decoder**
- **Optimizer: "nadam"**
- **Loss Function: MSE**
- **Epochs - 800**

Deep Neural Network
Loss: 45208.87030354782
Accuracy: 0.8470692038536072

Case Studies

Production Forecasting: Results

	Well	Reservoir	Date	Qo, BOPD	Qw, BWPD	Qg, MSCF/D
9071	BCS0070	M18	41944	467.636810	1478.841187	466.472656
9072	BCS0070	M18	41974	816.668823	1209.927368	727.219116
9073	BCS0070	M18	42005	649.015991	1453.822021	263.700012
9074	BCS0070	M18	42036	668.285583	1070.237549	300.476562
9075	BCS0070	M18	42064	665.108887	1468.606567	262.656342

	Well	Reservoir	Date	K, mD	h, ft	Depth_avg, ft	P, psi	Pwf, psi
0	BCS0006	M06	42095	3.206	189.3	9932	3070.6	2998
1	BCS0006	M06	42125	3.206	189.3	9932	3067.9	2998
2	BCS0006	M06	42156	3.206	189.3	9932	3065.0	2998
3	BCS0006	M06	42186	3.206	189.3	9932	3062.3	2998
4	BCS0006	M06	42217	3.206	189.3	9932	3059.4	2998

	Well	Reservoir	Date	K, mD	h, ft	Depth_avg, ft	P, psi	Pwf, psi	list	Completion
1440	BCS0070	M18	43770	1.24	77.0	9357	2901.093551	695	[BCS0070, M18]	BCS0070:M18
1441	BCS0070	M18	43800	1.24	77.0	9357	2899.969240	695	[BCS0070, M18]	BCS0070:M18
1442	BCS0070	M18	43831	1.24	77.0	9357	2898.811652	695	[BCS0070, M18]	BCS0070:M18
1443	BCS0070	M18	43862	1.24	77.0	9357	2897.658333	695	[BCS0070, M18]	BCS0070:M18
1444	BCS0070	M18	43891	1.24	77.0	9357	2896.583286	695	[BCS0070, M18]	BCS0070:M18

Case Studies

Deep Learning Time Series Analysis

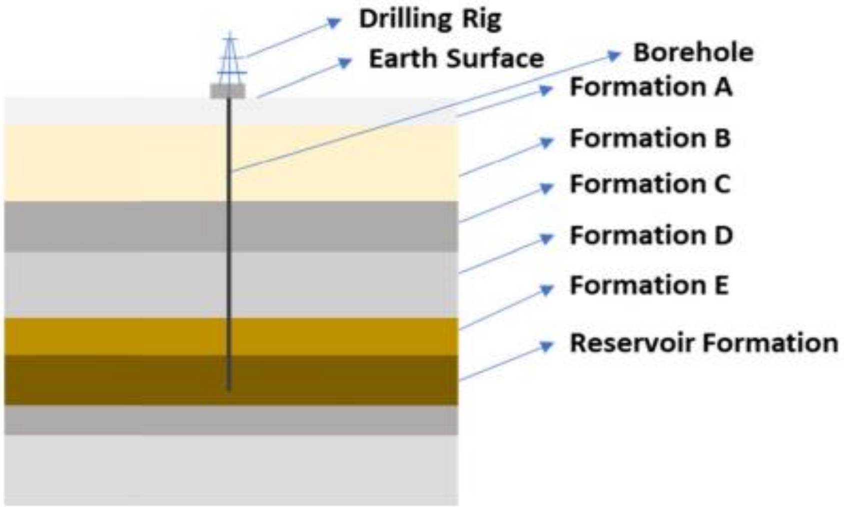


Figure 01

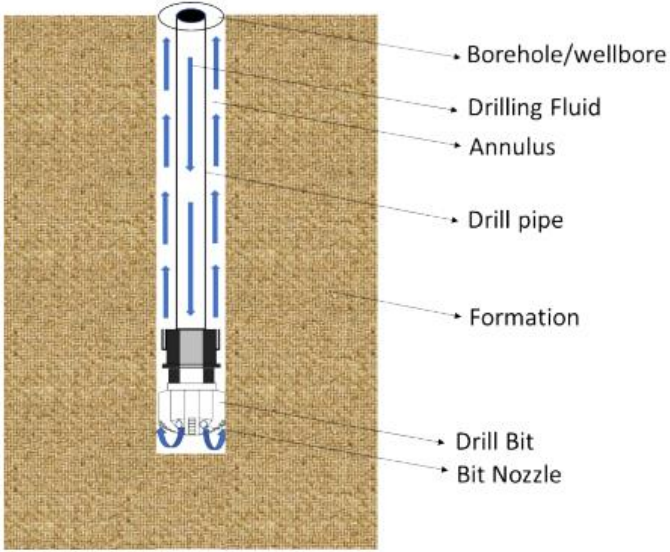
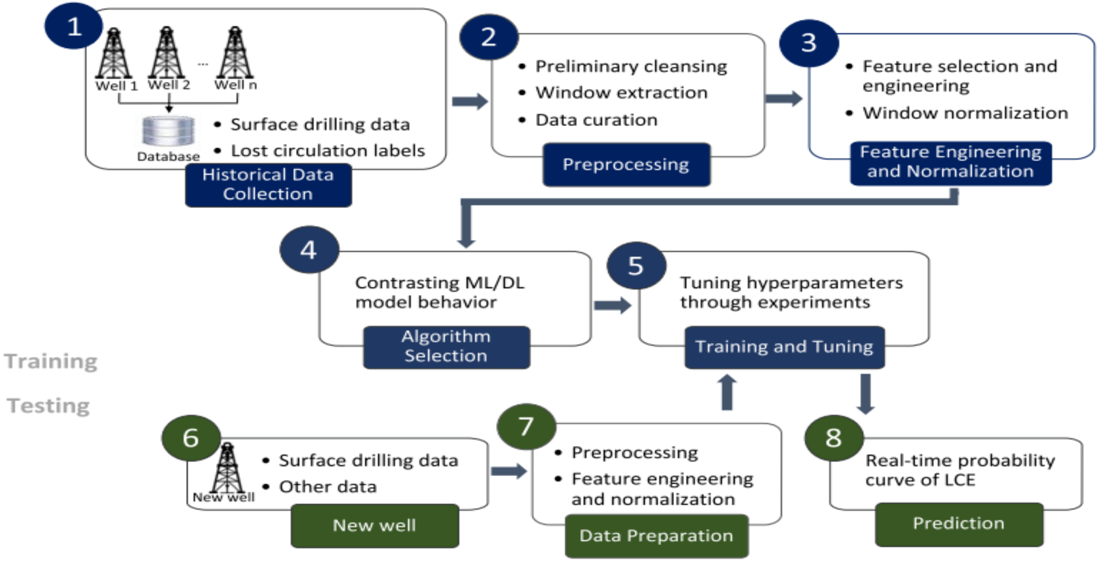


Figure 02

Case Studies

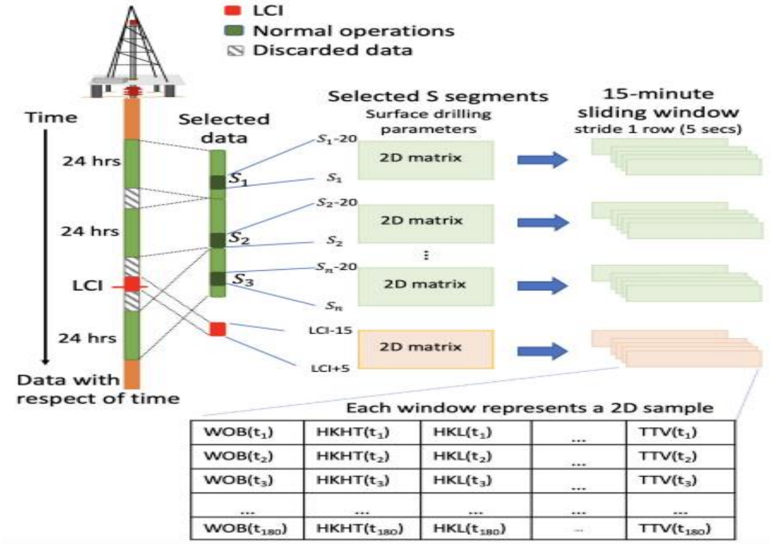
Deep Learning Time Series Analysis



Case Studies

Deep Learning Time Series Analysis

Figure 01



Parameter	Units
Weight on bit (WOB)	Kilo pound (klbf)
Hook height (HKHT)	Feet (ft)
Hook load (HKL)	Kilo pound (klbf)
Torque (TQ)	Pound-foot (kft.lbf)
Stand pipe pressure (SPP)	Pounds per square inch (psi)
Flow-in rate (FLWIN)	Gallons per minute (gpm)
Flow-out rate (FLWOUT)	0-100%
Rate of penetration (ROP)	Feet per hour (ft/hr)
Revolutions per minute (RPM)	Rpm
Total mud system volume (PVT)	Barrels
Trip tank volume (TTV)	Barrels

Table 01

Case Studies

Deep Learning Time Series Analysis

1. Data Preprocessing
2. Feature Engineering
3. Window normalization
4. Data split
5. Algorithm tuning

LC1 Reported by Drilling Crew

Figure 01

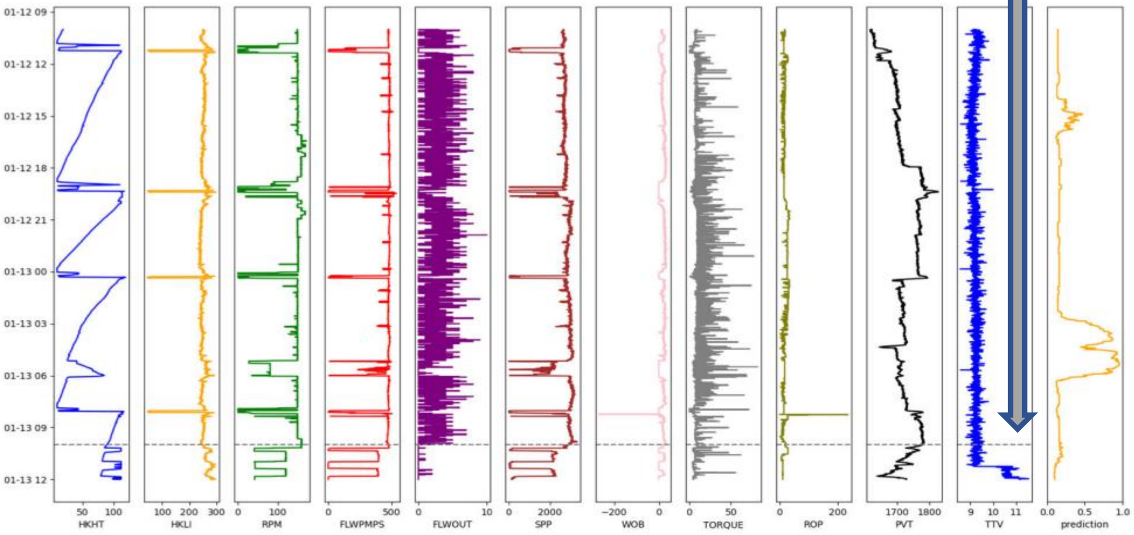


Table 01

Criteria	Description
Relevance	Drilling operations generate various forms of discrete and continuous data streams, hence, it is crucial to exclude sporadic and irrelevant features, which could hamper DL model training in latter stages.
Range check	Elimination or re-sampling data within soft upper and lower boundaries to eliminate extreme and physically meaningless drilling feature values (e.g., negative ROP).
Gap filling	Drilling data sometimes have one or more missing features across specific intervals of time, which must be either imputed or ignored during training. Whereas imputation is outside the scope of this work, wells with missing data across the intervals of interest are ignored during training. During evaluation, data processed through the model must be either complete or imputed a priori before utilizing the proposed model.
Duplication check	Drilling databases sometimes contain repeated or redundant features which hold perfectly or nearly identical values, hence these redundant features are checked and dropped at this stage. Note that redundancy is often encountered when the same physical quantity is measured using two different tools. For instance, in managed pressure drilling, FLWOUT is a Coriolis flow meter in addition to the standard rig flapper sensor located at the mud return line (a pipe through which mud flows from the well head to the mud tanks).
Unit standardization	Features stored in different measuring units (e.g., pounds versus kilo pounds) across different wells are standardized to be consistent and comparable in magnitude.

Module 11
Digital Twins: Upstream E&P

MODULE 11

This Module introduces case studies where Digital Twins are implemented. We shall explore the Digital Twin methodology and the various versions used in the upstream Exploration and Production (E&P) value chain.

Generative AI (GAI) is gaining traction across business verticals. We shall discuss Digital Twins from the perspective of generating or augmenting synthetic datasets for machine-learning data-driven analytical workflows in E&P.

Reinforcement Learning (RL) is a subfield of machine learning that focuses on how an agent can learn to make sequential decisions in an environment to maximize cumulative reward. It is inspired by how humans and animals learn through trial and error and interact with their surroundings. We shall compare RL with supervised and unsupervised methods.

We introduce a Reservoir Simulation case study using RL.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 11

Digital Twins: Upstream E&P

LEARNING OBJECTIVES

- ➢ GOAL01: Digital Twins Introduction
- ➢ GOAL02: Digital Twins in the O&G industry
- ➢ GOAL03: Case studies implementing Digital Twins in upstream
- ➢ GOAL04: Reinforcement Learning: A ML Technique for Digital Twins

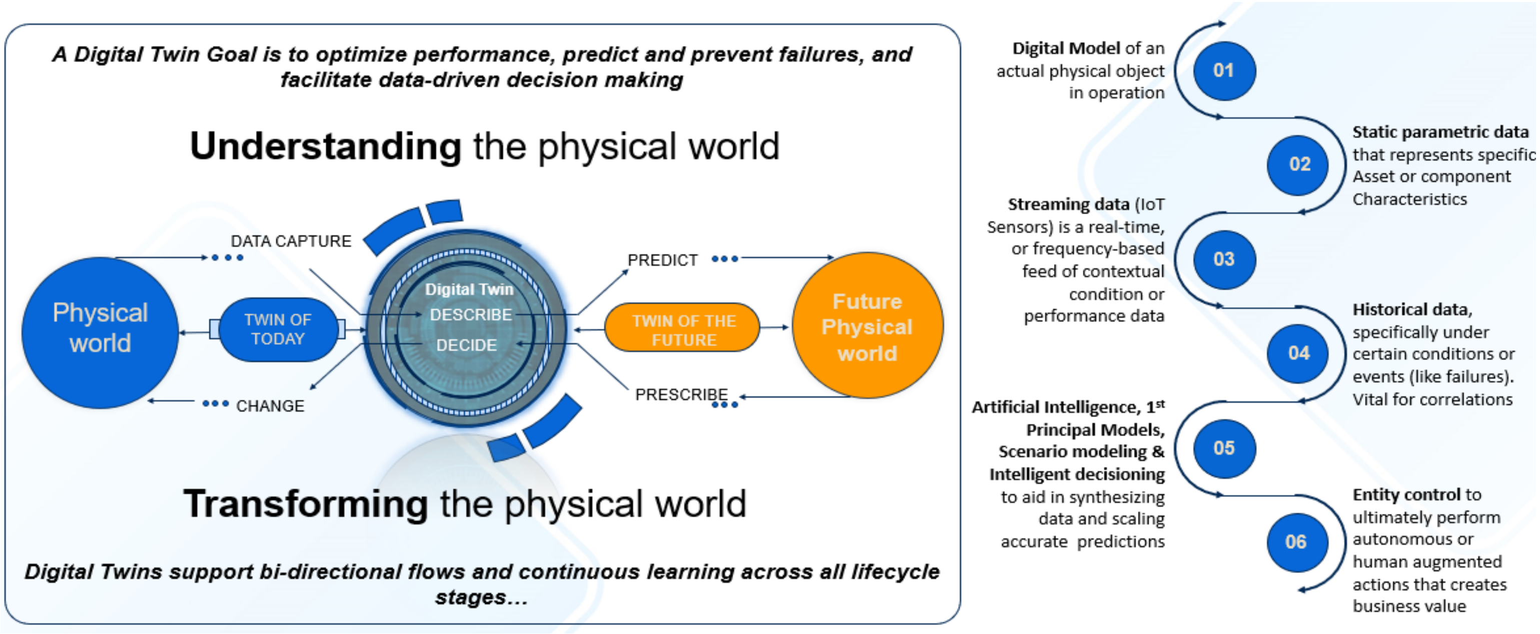
Digital Twins

Digital Twins: Targeted Outcomes

1. Improved performance
2. Enhanced predictability (reduced downtime)
3. Increased innovation through virtual testing
4. Improved collaboration and decision-making
5. Reduced costs (Maintenance, Labor, Raw Materials)

Digital Twins

Digital Twins



Digital Twins

Digital Twins

Exploration and Drilling

- **Reservoir Simulation**: Digital twins can model underground reservoirs to predict how they'll behave. This aids in optimizing extraction techniques and maximizing the reservoir's yield.
- Drilling Optimization: By simulating the drilling process, companies can identify potential issues (like equipment failures or geological hazards) and adjust the drilling strategy accordingly.

Asset Performance Management

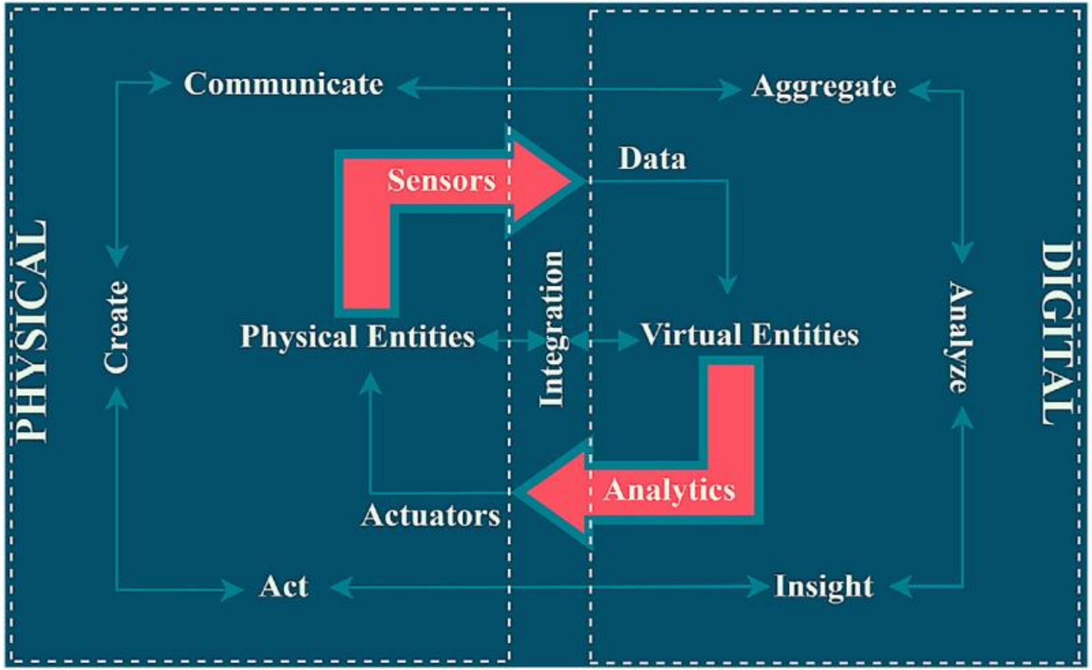
- Predictive Maintenance: Digital twins can predict when equipment might fail using sensors and real-time data. This helps companies fix problems before they happen, reducing downtime.
- Operational Optimization: Digital twins can model the entire operation of an asset (like an oil rig or refinery). Companies can find the most efficient way to run their operations by simulating different conditions.
- Production Optimization: Flow Simulation: Digital twins model the flow of oil and gas through pipelines and other infrastructure. This can help identify bottlenecks or inefficiencies in the system.

Real-time Monitoring and Control

- Remote Operations: Particularly useful in offshore or remote sites, digital twins allow operators in centralized control rooms to monitor and control equipment from a distance.
- System Performance: By continuously comparing the digital twin's performance with the real-world asset, discrepancies can be spotted immediately, leading to quick interventions.

Digital Twins

Digital Twin in the Hydrocarbon Industry

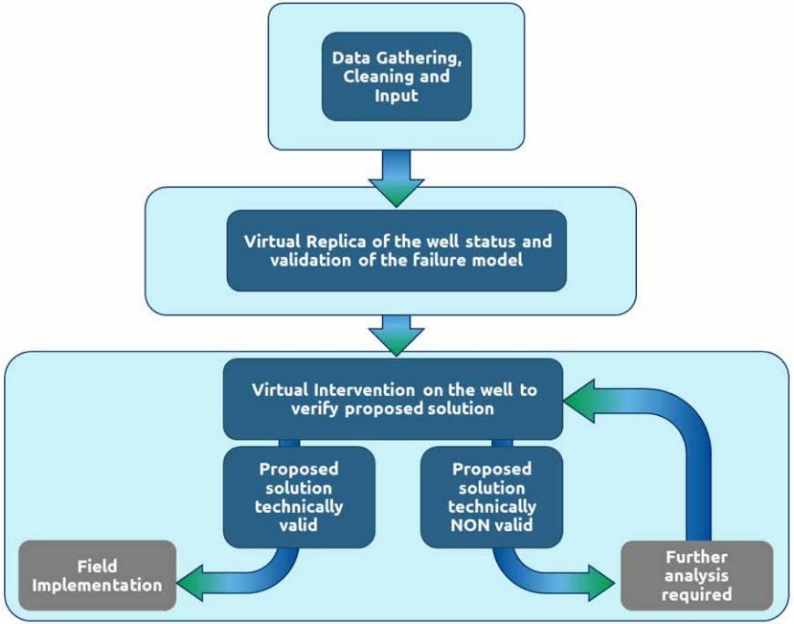


Digital Twin (DT) modeling is the foundation for the next generation of real-time production monitoring and optimization systems. It is a solution that boosts productivity by combining information, simulation, and visualization throughout the entire value chain of an operational firm, from subsurface equipment to central production plants.

Digital Twins

Digital Twin in the Risk Assessment Process

Generative AI – Digital Twin

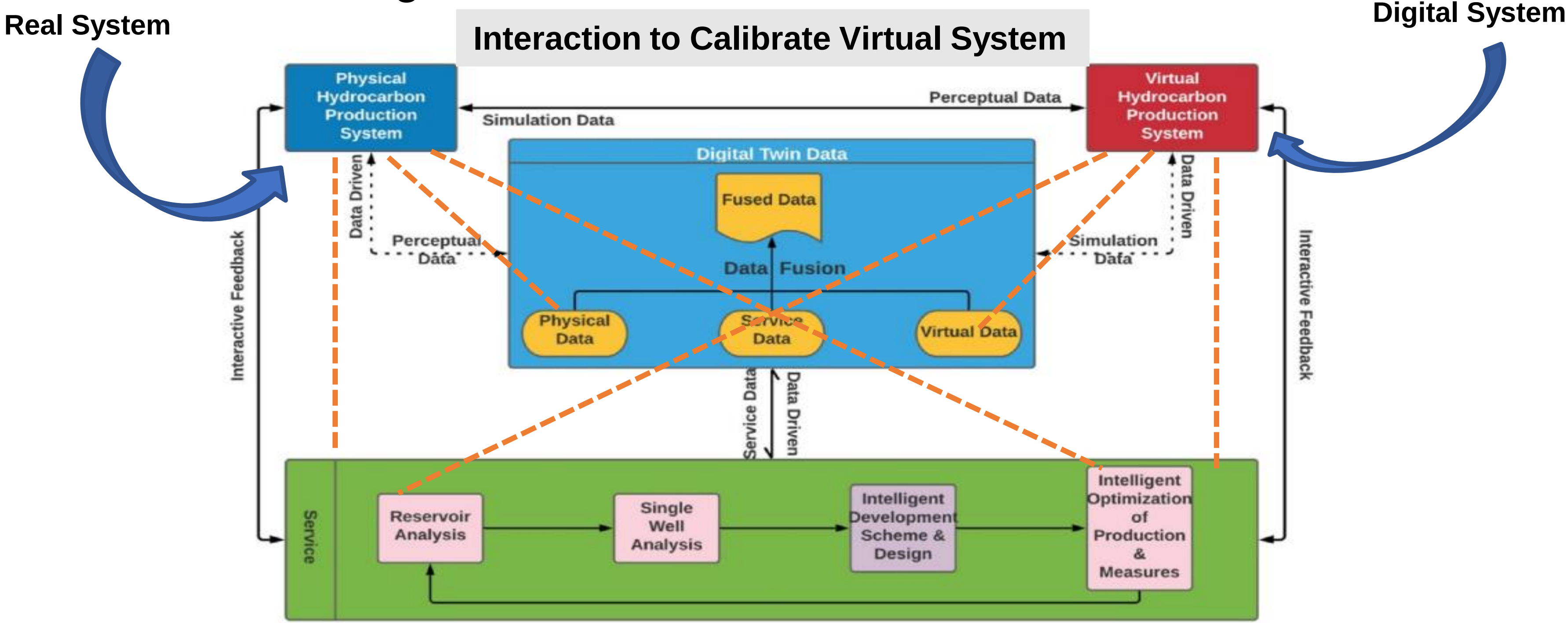


Risk Assessment Process

Digital twin standard workflow

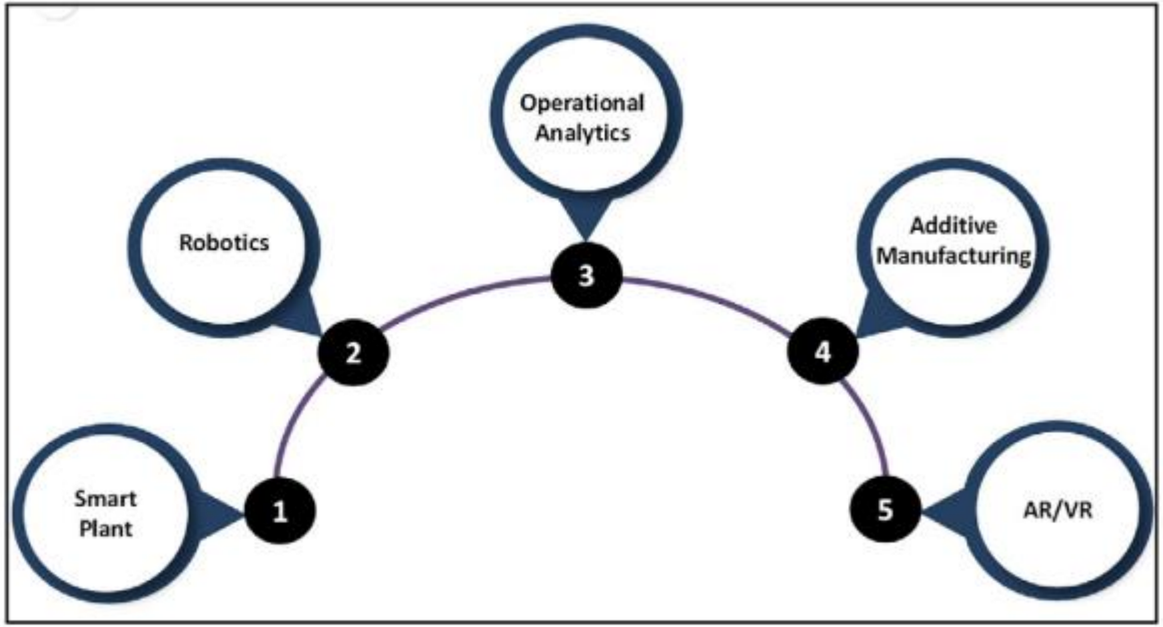
Digital Twins

Digital Twin Framework for O&G Production



Digital Twins

Strategies to Achieve a Digital Twin Model



Digital Twins

Digital Twin of a Well

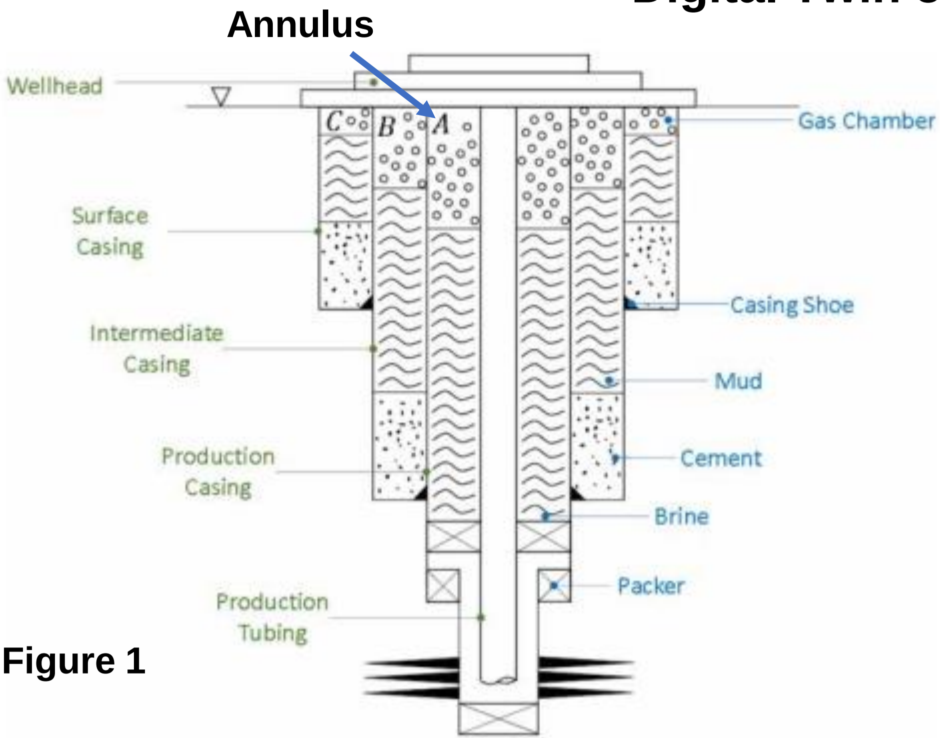


Figure 1

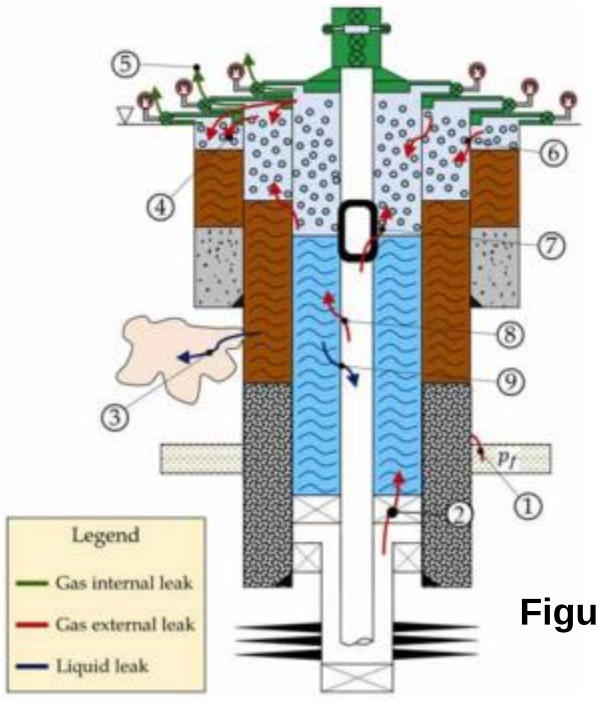


Figure 2

Digital Twins

Digital Twin – Smart Water Optimization Workflow

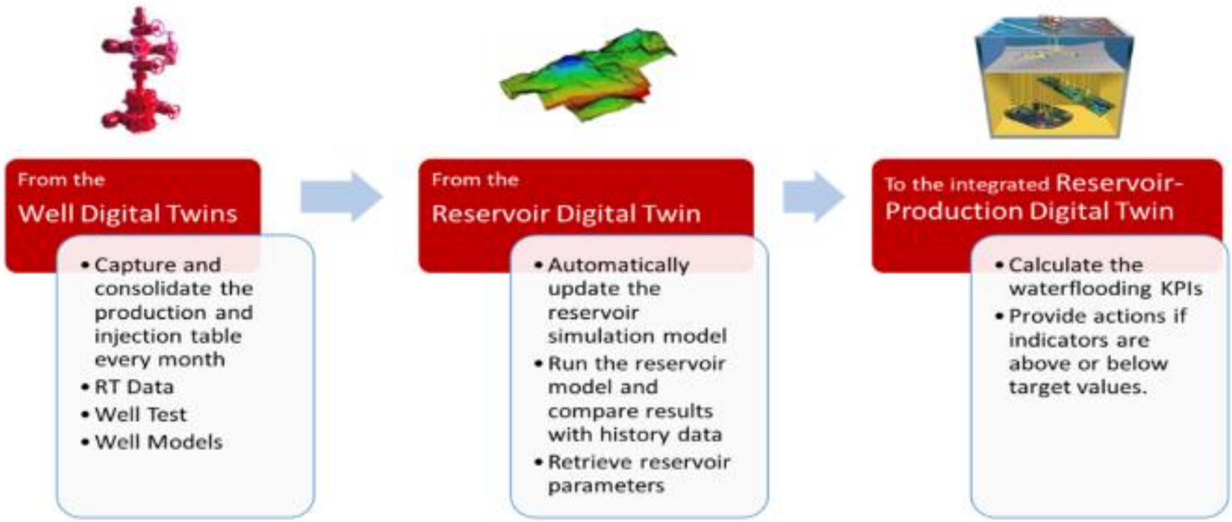
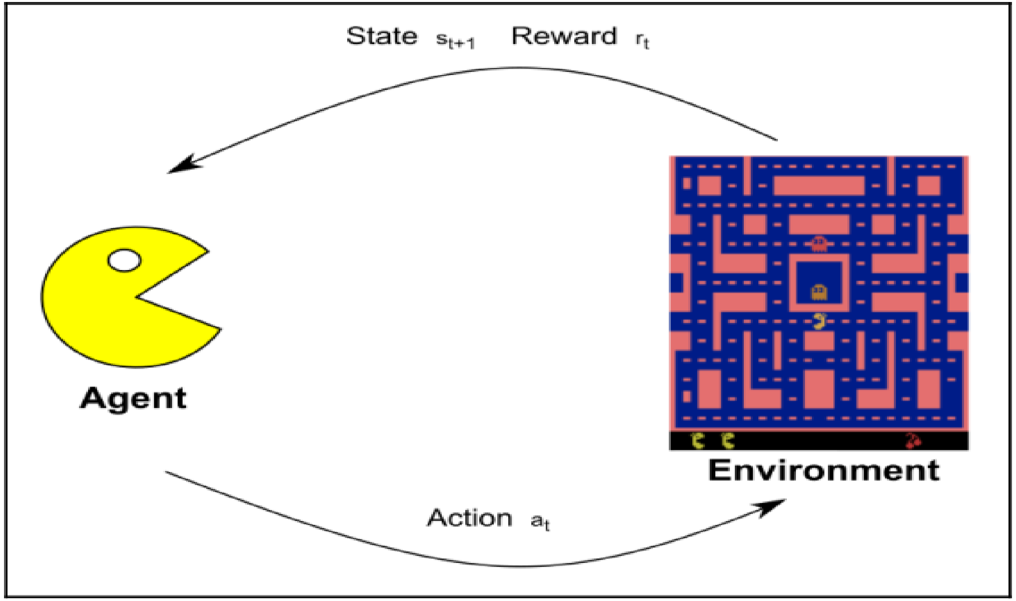


Figure 01

Digital Twins

What is Reinforcement Learning?



Reinforcement learning is learning what to do — how to map situations to actions — to maximize a numerical reward signal. The learner is not told which actions to take but must discover which ones yield the most reward by trying them. In the most interesting and challenging cases, actions may affect the immediate reward, the next situation, and all subsequent rewards. These two characteristics — trial-and-error search and delayed reward — are the two most important distinguishing features of reinforcement learning.

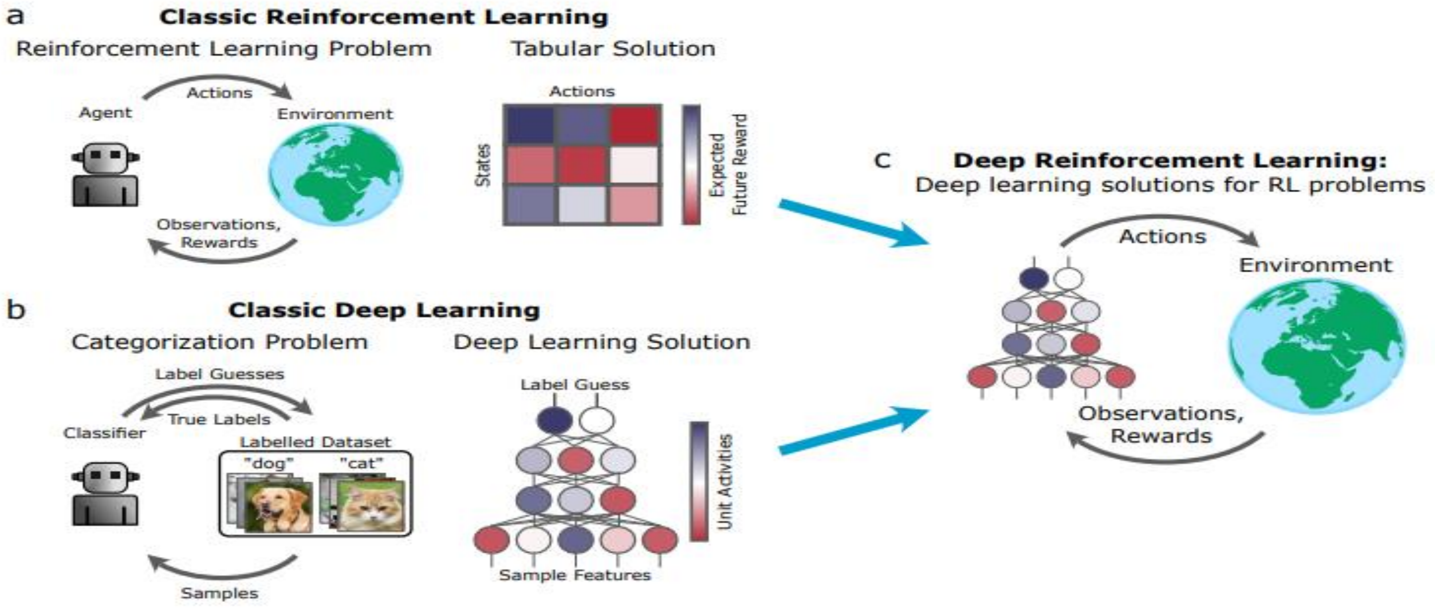
Digital Twins

Reinforcement vs. Supervised/Unsupervised

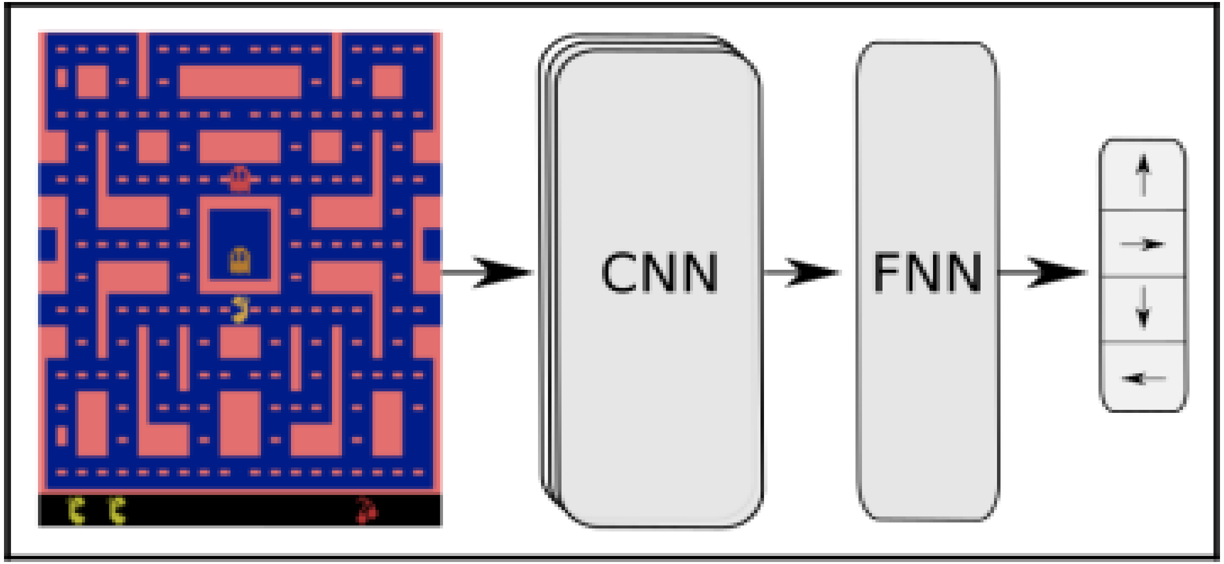
Reinforcement Learning	Supervised/Unsupervised Learning
Objective: choose "best" actions	Objective: Predict, classify or simplify
Environment is uncertain	Environment is known (x is known)
Training involves exploring the environment	Training involves finding patterns in data or is entirely absent
Training process involves determining the "best" policy	Training process involves fitting the "best" model
Explicit dependency of rewards on previous actions	Individual points are independent of each other

Digital Twins

Deep Reinforcement Learning (DRL)

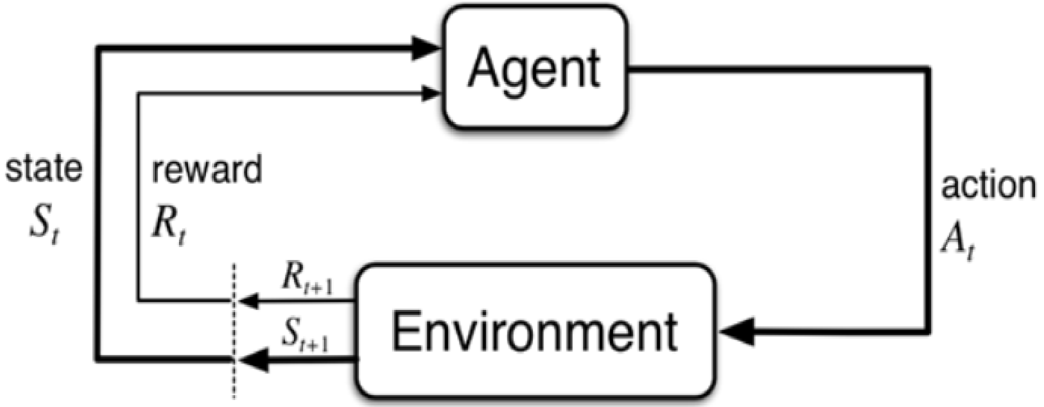


Digital Twins



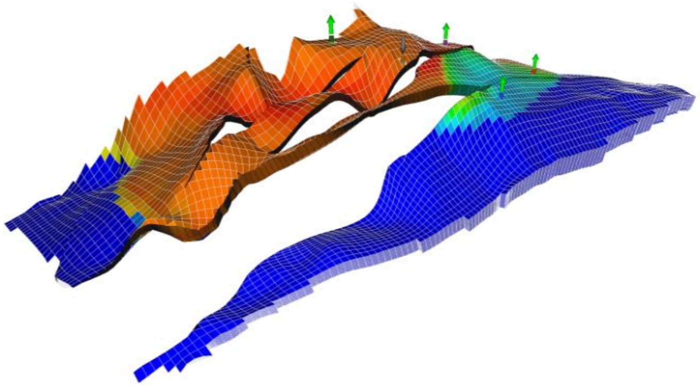
Deep Reinforcement Learning for Petroleum Reservoir Optimization

Digital Twins



Deep Reinforcement Learning for Petroleum Reservoir Optimization

Reservoir Simulation



Module 12
PINNs: Physics-Informed Neural Networks & Explainable AI and Generative AI

MODULE 12

This Module introduces PINNs, Physics-Based Neural Networks., recently proposed for solving partial differential equations. Unlike typical ML algorithms that require a large dataset for training, PINNs can train the network with unlabeled data. The applicability of this method has been explored for the flow and transportation of multiphase flow regimes in porous media.

We shall introduce a case study to manage reservoir pressure by implementing a PINN.

The module also details Explainable AI (XAI), a set of processes and methods that allow human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI describes an AI model, its expected impact, and potential biases.

We shall also explore Generative AI, discussing the Pros and Cons of these techniques in the oil and gas industry.

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN

TRAINING & CONSULTING

Harness Upstream Geophysical and Petrophysical Data with AI Workflows

INDEX

INDEX

Module 12

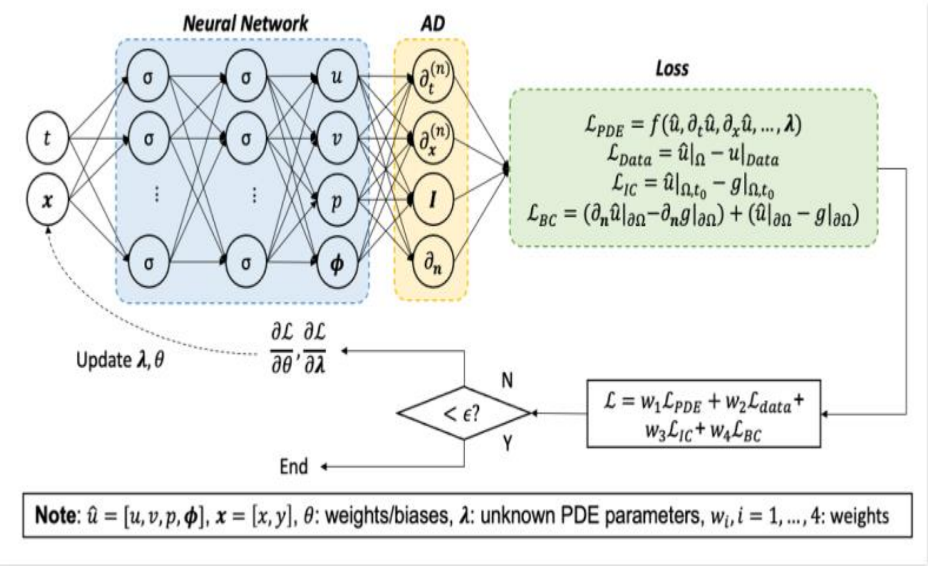
PINNS: Physics Informed Neural Networks – Explainable AI and GAI

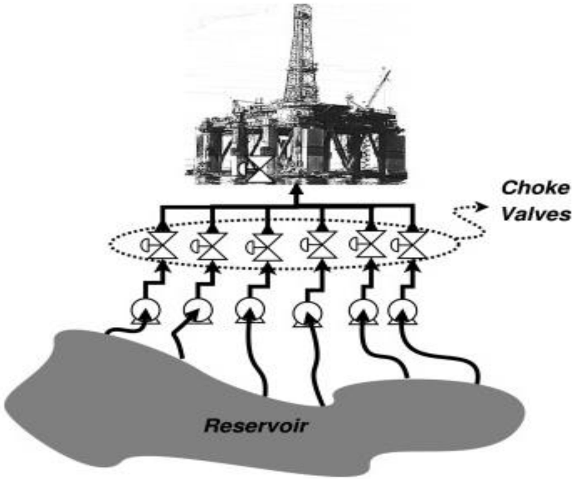
LEARNING OBJECTIVES

- ➢ GOAL01: Physics and Data-Driven Machine Learning
- ➢ GOAL02: Case Study Reservoir Pressure Management
- ➢ GOAL03: Explainable AI in Upstream: An Application to Lithology Prediction
- ➢ GOAL04: Generative AI in Upstream

PINNS

Physics-Based against Data-Driven Based Models





Simplified representation of an offshore oil production system

PINNS

Physics-Based against Data-Driven Based Models

Models	Advantages	Disadvantages
Physics driven	Strong basics, based on existing solid knowledge Easy to interpret Can detect errors and uncertainties and avoid them Lower probability of bias Easy to be generalized to other problems Fundamental relationships give insight and help in understanding Valid prediction at a full range of model coverage	Hard to integrate historical or archived data with the models Prone to numerical instability as a result of having complex boundary conditions and inputs uncertainties Vast physics knowledge in the domain is required High computational power requirement, so it suffers if used for real time Assumptions need to be set in advance
Data-driven	Considers the historical data and experiences into the model Able to stably make predictions after training Does not require knowledge of the domain as it depends mainly on data Deals with heterogeneous data Able to enhance performance over time Can detect complicated relationships and patterns	Black box nature and interpretability issues Cannot detect errors or uncertainties Affected by bias in data Not easy to generalize Data availability is the main concern It is an approximation Lower performance outside the scope of the training data Hard to predict critical conditions or extremes

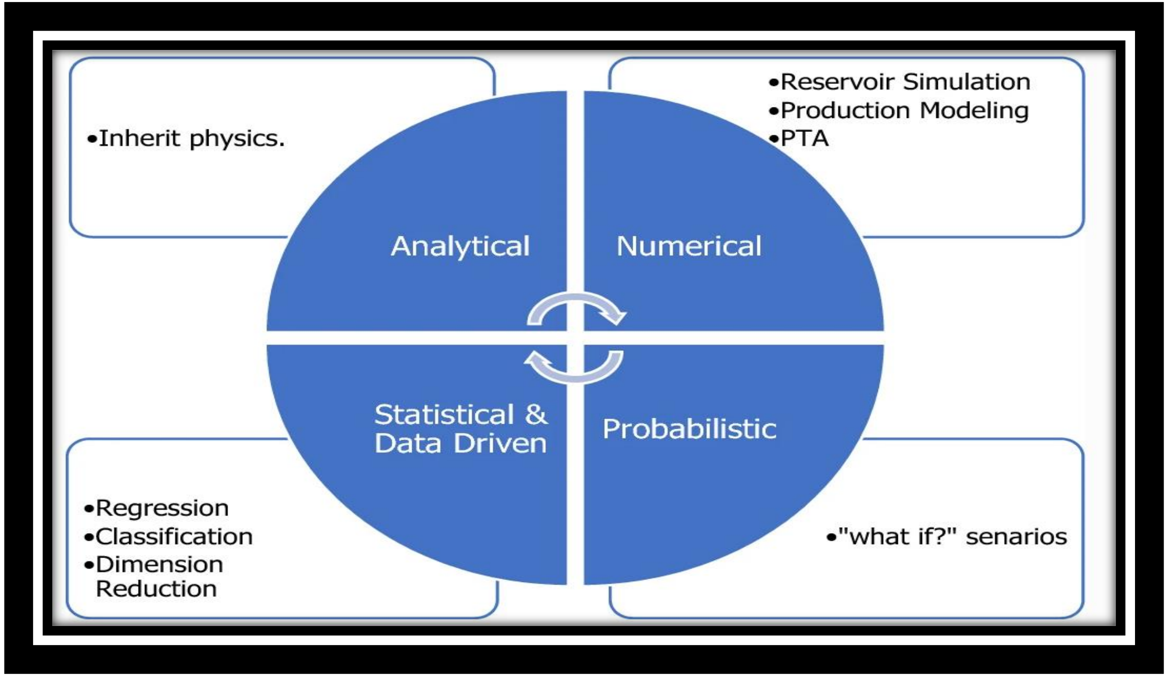
PINNS

Hybrid Models

Model	Application	Components
Digital twin	Drilling engineering	Sensor data in near-real time (Data).Synthetic data generated from simulators (Physics) Humans to interact using avatar (Expert) Digital siblings for "what if?" scenarios
ML & probabilistic approach	Oil and gas production	Calculated input parameters using existing principles (Physics) Classification using ML models (Data) A probabilistic model to quantify the uncertainty associated with each method Cost model to predict financial impact
ML & digital rock analysis (DRA)	Reservoir characterization	Rock image acquisition Image processing using ML models (Data) Numerical simulation (Physics) Result Analysis
Surrogate reservoir model (SRM)	Reservoir characterization	Multiple neuro-fuzzy systems Numerical simulation model Spatiotemporal database

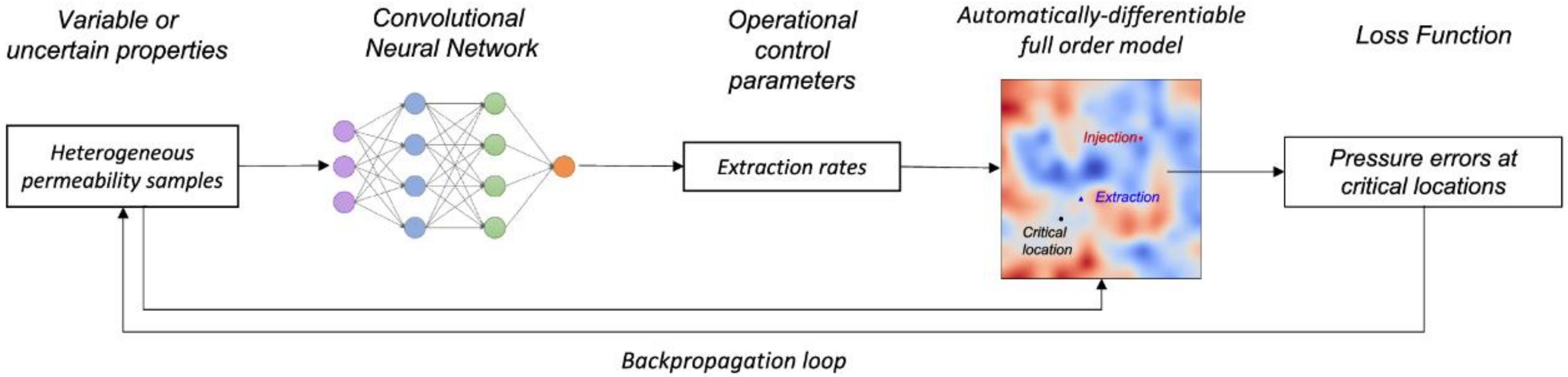
PINNS

Data Science and Machine Learning Applications in O&G



PINNS

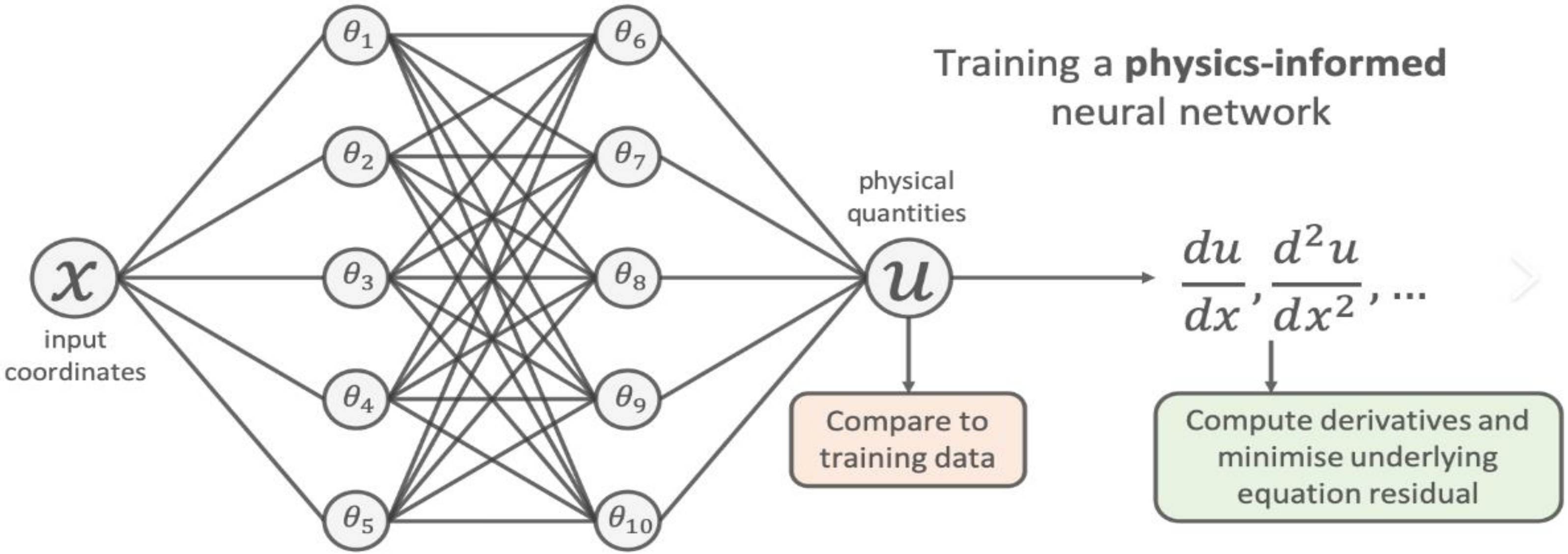
PINN in Upstream: Case Study: Reservoir Pressure Management



Workflow diagram of physics-informed machine learning framework for managing reservoir pressures at a critical location during subsurface fluid injection. The key innovation of the surrogate model is the automatically-differentiable full-order model that allows for heterogeneity.

PINNS

PINNs in Upstream Conclusions



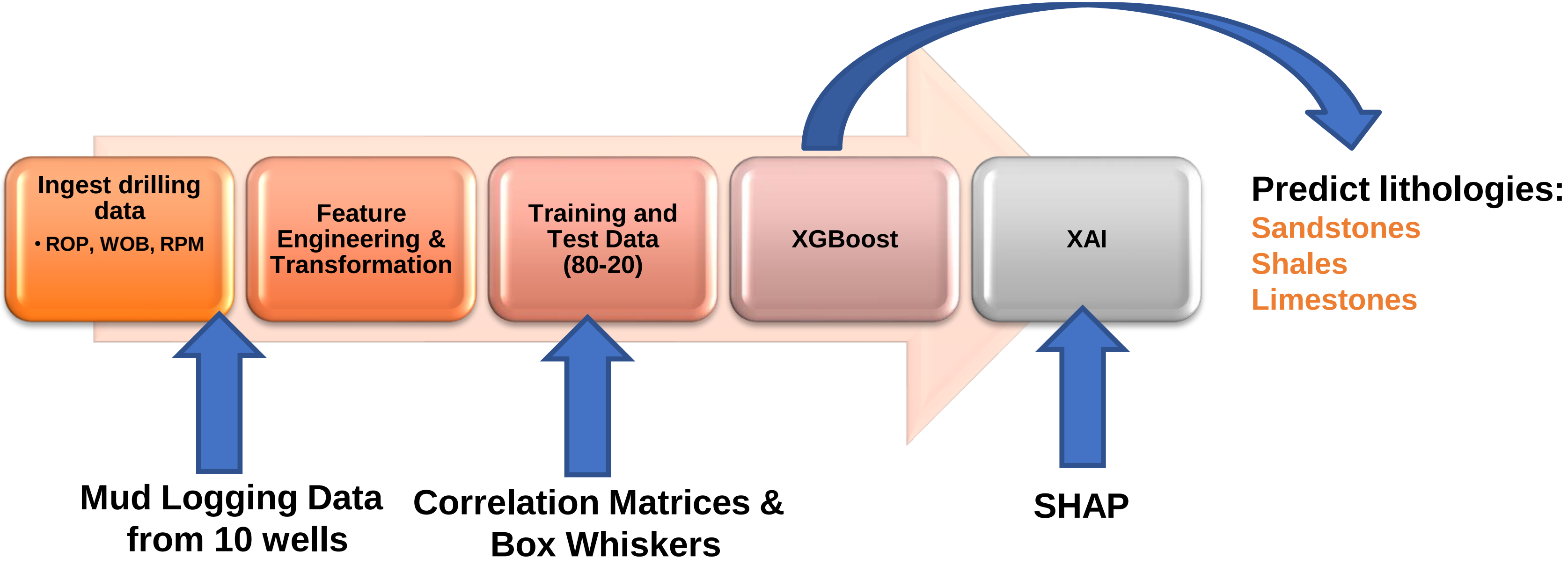
PINNS

Explainable AI (XAI)

Explainable AI is one of the key requirements for implementing responsible AI, a methodology for the large-scale implementation of AI methods in real organizations with fairness, model explainability, and accountability.

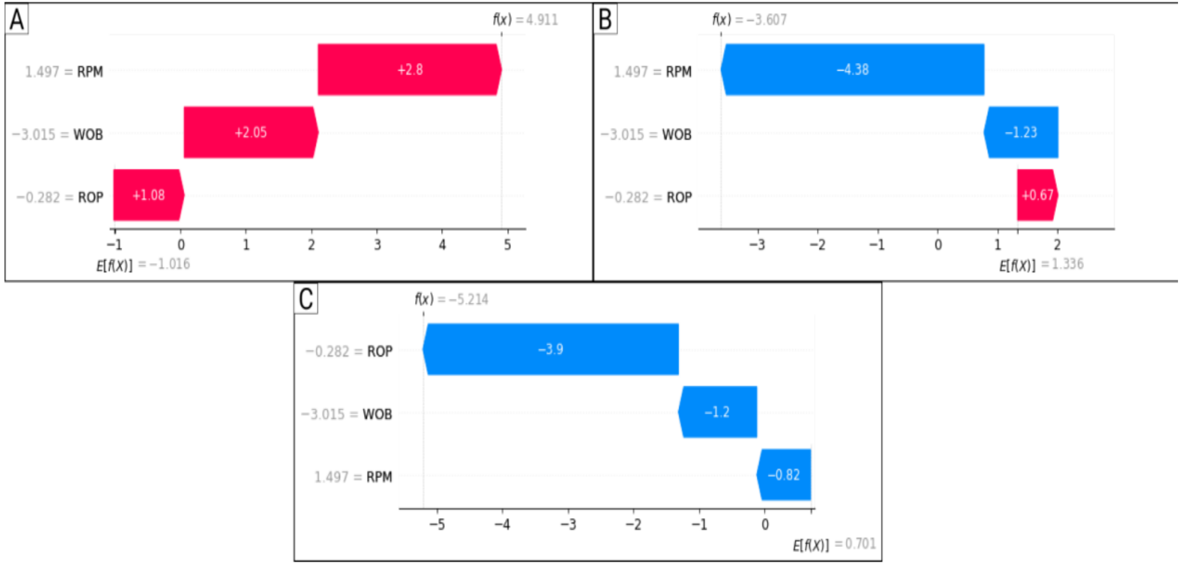
PINNS

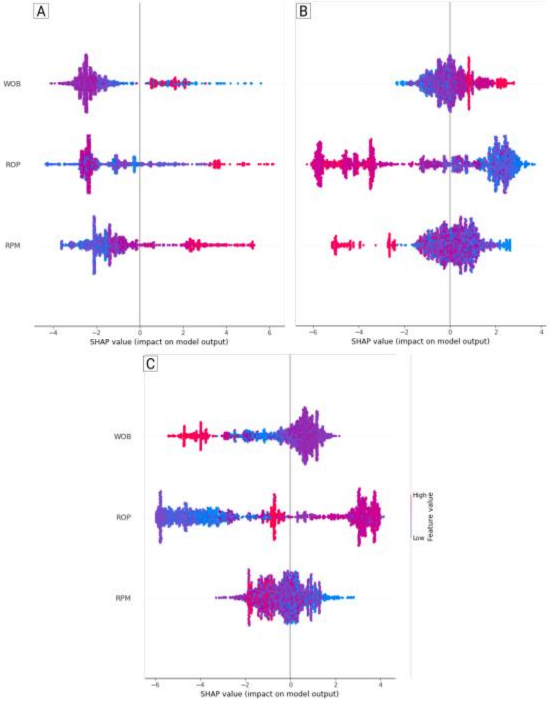
Explainable AI



PINNS

Explainable AI SHAP

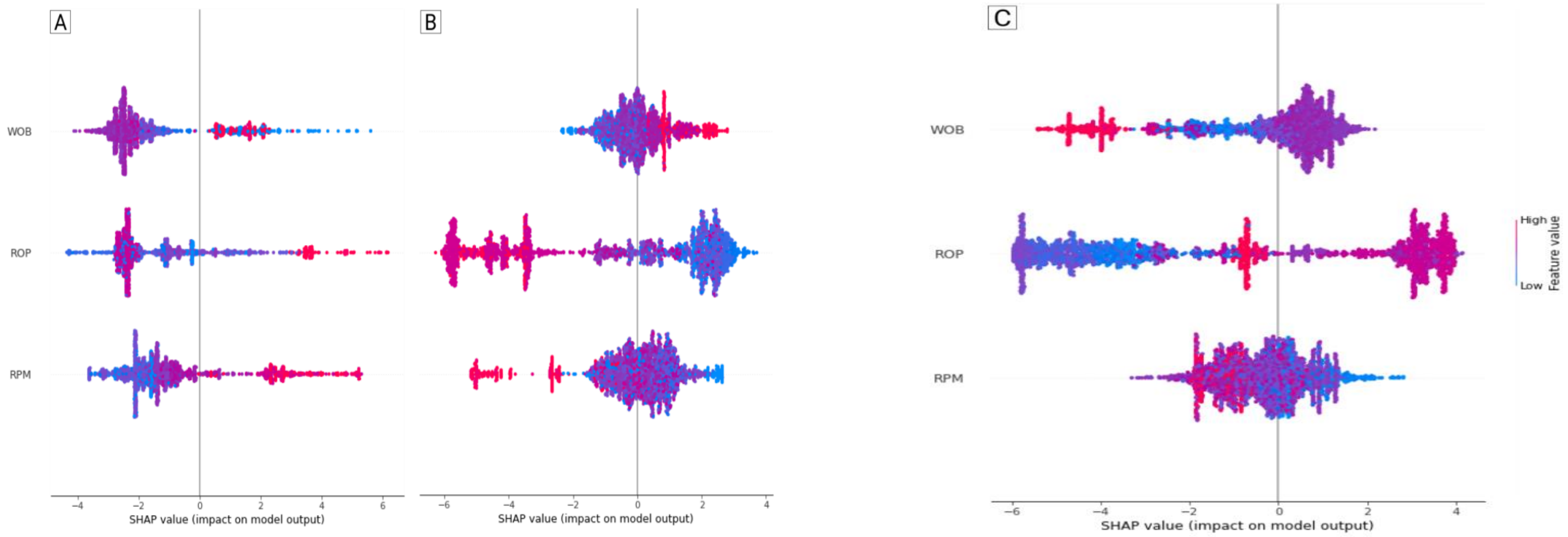




Local explanations for a specific prediction regarding a possible sandstone (A), shale (B), or limestone (C).

PINNS

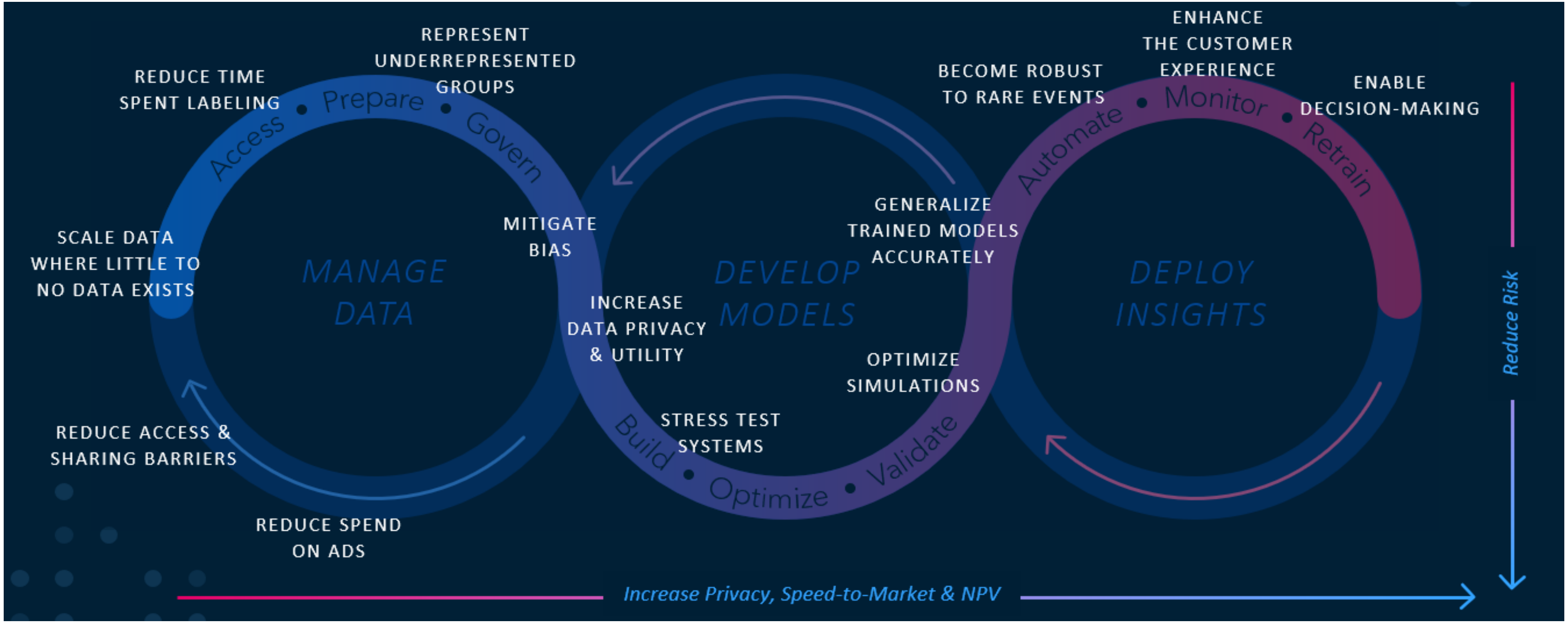
Explainable AI SHAP



Global explanations regarding the XGBoost predictive model, specifically for sandstone (A), shale (B), or limestone (C)

Data Preparation for AI

Generative AI in O&G Upstream



Data Preparation for AI

Generative AI (ChatGPT) in O&G Upstream

ZE

Create a python script for sample data for 50 wells with 10 well bores each and each well bore has 9 stages. Include the following parameters NetH, Phi, Sg, Distance from Peak, Laplacian, Dip, Delta Height, Sum of Prop Vol, Qg100, Water Saturation, Pressure Gradient, EUR, Lateral Length, Stage Spacing. Save the output as a CSV file

```
In [2]: import random
        import pandas [illegible]

        # Set the numb[illegible] oil wells and bores
        num_wells = 50
        num_bores = [illegible]

        # Set the ra[illegible] each parameter
        neth_range = [illegible])
        phi_range = ([illegible]25)
        sg_range = (70, 90)
        distance_range = (400, 800)
        dip_range = (8, 12)
        og100_range = (75, 90)
        lateral_length_range = (5000, 6000)
        stage_spacing_range = (100, 150)

        # Initialize an empty list to store the data
        data = []

        # Generate random data for each well bore
        for well in range(1, num_wells + 1):
            for bore in range(1, num_bores + 1):
                neth = random.uniform(*neth_range)
                phi = random.uniform(*phi_range)
                sg = random.uniform(*sg_range)

                distance = random.uniform(*distance_range)
                laplacian = random.uniform(0.9 * neth, neth)
                dip = random.uniform(*dip_range)
                delta_height = random.uniform(0.1 * neth, 0.5 * neth)

                sum_prop_value = random.uniform(0.8 * neth * distance, 1.2 * neth * distance)
        [illegible]g100 = random.uniform(*og100_range)
        [illegible]ter_saturation = random.uniform(0.5 * sg, 0.8 * sg)
        [illegible]ssure_gradient = random.uniform(0.1 * sg, 0.2 * sg)

        [illegible] = random.uniform(0.1 * neth * phi * sg * distance, 0.3 * neth * phi * sg * distance)
        [illegible]ral_length = random.uniform(*lateral_length_range)
        [illegible]_spacing = random.uniform(*stage_spacing_range)

        # [illegible]end the generated data to the list
        d[illegible]append([well, bore, neth, phi, sg, distance, laplacian, dip, delta_height,
                         sum_prop_value, og100, water_saturation, pressure_gradient, eur,
                         lateral_length, stage_spacing])

        # Create [illegible] DataFrame from the generated data
        columns [illegible] ['Well', 'Well Bore', 'NetH (ft)', 'Phi (%)', 'sg (%)', 'Distance (ft)', 'Laplacian',
                   'Dip (degrees)', 'Delta Height', 'Sum of Prop Value', 'OG100 (%)',
                   'Water Saturation', 'Pressure Gradient', 'EUR (MMbbl)', 'Lateral Length (ft)',
                   'Stage Spacing (ft)']
        [illegible]pd.DataFrame(data, columns=columns)

        # Save the DataFrame as a CSV file
        df.to_csv('oil_wells_data.csv', index=False)
```

ShellGAI
Source

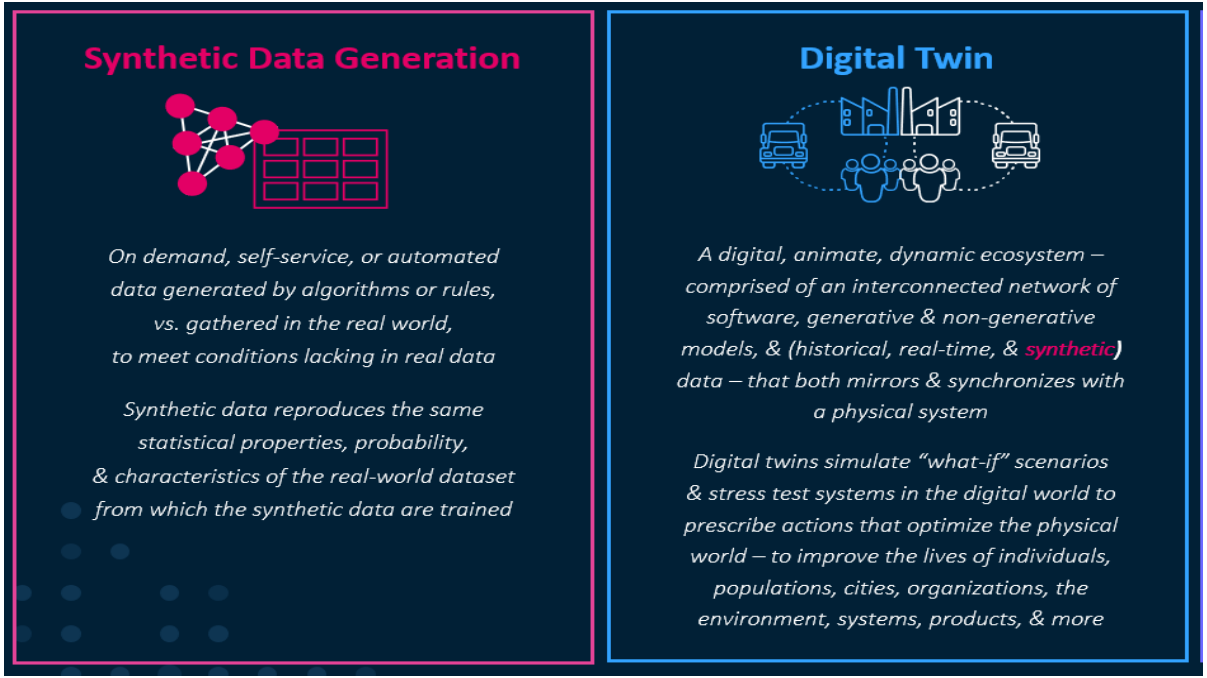
Columns (16/0)
Well
Well Bore
Stage
NetH (ft)
Phi (%)
sg (%)
Distance (ft)
Laplacian
Dip (degrees)
Delta Height
Sum of Prop Value
QG100 (MMscf)
Water Saturation
Pressure Gradient
Lateral Length (ft)
Stage Spacing (ft)

Rows
All rows 9,000
Selected 0
Excluded 0
Hidden 0
Labeled 0

	Well	Well Bore	Stage	NetH (ft)	Phi (%)	sg (%)	Distance (ft)	Laplacian	Dip (degrees)	Delta Height	Sum of Prop Value	QG100 (MMscf)	Water Saturation	Pressure Gradient	Lateral Length (ft)	Stage Spacing (ft)
1	1	1	1	115.3508648	20.08358153	71.28888667	735.7465697	0.062788837	9.911136558	232.1748743	5263.129947	76.61375847	26.50248416	0.483132272	5388.853941	103.6878181
2	1	1	2	105.0189317	24.02831669	81.33902333	423.573941	0.043472298	8.247276797	205.8356519	5318.977328	78.10498506	26.90029601	0.459189435	5617.709328	131.4632348
3	1	1	3	138.3977138	23.41113061	82.68871067	606.4420788	0.046227365	10.58861926	201.4433845	4807.578256	86.43963254	28.15551733	0.47748332	5029.914953	106.5972624
4	1	1	4	140.4557159	23.46314577	76.59016713	459.7448032	0.048349178	9.135097691	186.7799386	5072.927784	77.67585643	30.57331582	0.539333313	5980.226436	123.1853797
5	1	1	5	125.8903279	19.31530594	87.6233268	597.3638469	0.037102747	10.63918077	189.9805739	5197.482462	83.0062503	26.98647243	0.559320403	5828.513858	143.6095486
6	1	1	6	119.5162739	23.21718461	87.71988893	762.6507273	0.036554676	10.74470164	229.4518674	4898.005633	78.44969253	27.30853637	0.580451096	5252.181032	122.44676
7	1	1	7	97.20906954	17.19969355	72.53321818	513.1665839	0.044694138	11.25921038	234.3515789	4885.834209	78.83288036	26.29608003	0.433771404	5655.078406	129.1016393
8	1	1	8	102.0745788	21.30516828	70.1652977	542.7373843	0.03485809	11.03203262	219.1728483	4961.821933	78.59427741	31.3119684	0.572297744	5577.964122	105.9612568
9	1	1	9	103.6536847	20.32381771	85.38531129	515.6614761	0.049509513	11.20040013	232.1266285	4900.857764	79.82483214	31.27624453	0.460682144	5551.231644	117.7593922
10	1	2	1	88.99083624	15.1527661	82.89348549	594.6647664	0.062899338	8.030375157	191.5245002	4767.242659	75.64161327	30.99049797	0.490501534	5118.68205	131.2620026
11	1	2	2	121.8880461	18.58874875	70.96376675	455.7411406	0.049519817	11.57302236	191.9723933	4897.343082	87.78878071	27.34114329	0.554730693	5198.063335	126.0475438
12	1	2	3	112.0205264	24.0248296	72.00510625	719.5253977	0.05825964	8.409833919	219.9204788	5273.218654	87.94201095	27.31707811	0.448529708	5546.439066	105.2553378
13	1	2	4	83.83801623	22.66346313	70.52413641	643.8697317	0.031582124	8.202672315	239.2210635	4702.947052	82.0507185	29.47363061	0.473797334	5288.996678	131.6190649
14	1	2	5	118.9641582	21.03386682	83.44281276	418.0506299	0.067875102	9.557997021	226.3178666	4968.477672	76.00585029	30.73814401	0.488057054	5165.142462	141.0217036
15	1	2	6	133.1444126	20.08490624	84.5375879	718.277311	0.066327263	10.84535757	227.0074627	5351.371511	84.79731378	29.63314012	0.409891307	5607.19527	122.0628389
16	1	2	7	132.1966137	17.33905782	86.63298193	543.9900449	0.060305591	9.183998997	209.5121539	5258.194994	82.10854915	29.24297185	0.479166802	5672.266833	131.4903948
17	1	2	8	145.2566267	15.91405456	86.96015024	760.0182213	0.03202747	11.25107992	223.1798732	5240.286075	76.48555844	30.93426526	0.571131598	5161.090607	133.7235725
18	1	2	9	125.0552036	16.71010495	85.9462936	412.1335315	0.053597387	10.13192191	191.359192	5142.067085	79.7675074	29.44581954	0.498022418	5570.777301	128.9830186
19	1	3	1	126.1628274	19.83638454	72.08728404	607.2932314	0.058504578	8.305994321	235.5471461	4902.442551	86.78355616	30.93559987	0.587109097	5594.818027	131.1035091
20	1	3	2	135.1613267	22.60158266	70.1775458	516.7440375	0.058242948	9.228297874	189.6207491	4744.000539	85.72492377	28.99457088	0.415085663	5847.342904	117.5299751
21	1	3	3	86.87283576	22.30507714	77.37630373	718.3466392	0.056235011	11.21734707	232.8487959	5214.479317	86.38791214	27.42701303	0.446199529	5573.420462	113.100009
22	1	3	4	125.7990166	18.10867779	86.6346381	627.6078429	0.057113259	11.88731187	236.2437496	4902.818943	82.24506257	27.18312838	0.465720759	5379.409773	120.2070077
23	1	3	5	99.88203014	23.95577837	88.88282573	618.4990034	0.032308475	11.9644409	231.2820383	4767.857048	89.37410208	26.67851387	0.498542911	5351.180343	134.0257076
24	1	3	6	133.5092603	16.04184935	84.16821004	692.3761781	0.043207408	8.899922281	227.8504318	5025.317379	80.37511939	27.01031346	0.43159785	5910.684672	141.5355221
25	1	3	7	128.4974324	21.00222222	73.08407515	464.305989	0.030320362	10.04532884	203.510717	5072.6171	77.03509767	29.37913877	0.448825267	5513.795937	127.4380259
26	1	3	8	94.35383758	22.91490487	85.62531236	447.6895704	0.062359425	10.87994381	196.0292856	4854.016346	88.0641691	29.86551321	0.443368304	5452.927259	108.0495195
27	1	3	9	102.0318764	17.78286543	76.67546216	421.000981	0.055162634	10.33675138	211.0570795	5182.54427	87.57521555	30.07696771	0.513935793	5102.266834	117.4331795
28	1	4	1	143.7515905	15.45491795	73.27219287	470.8278446	0.051622743	9.107476766	222.8555991	5392.940879	83.24351111	27.12471357	0.596382274	5626.265291	102.0901141
29	1	4	2	96.17674928	24.57048388	82.44885426	404.1984616	0.036807422	8.381622926	195.8167549	4756.216883	85.49070022	29.48941139	0.508101781	5807.75989	147.4511016
30	1	4	3	137.4993542	21.73757411	84.98663579	700.2873434	0.060447049	10.13495028	204.640599	5139.711974	79.33748998	31.40568952	0.442999438	5600.840122	101.8501011
31	1	4	4	114.7706073	22.33395683	85.90010699	437.9623787	0.04355526	9.258742724	239.3434272	5265.228719	76.40263188	29.92700355	0.578524596	5815.669506	132.4351939
32	1	4	5	118.6350632	20.97346128	79.19714022	454.7468517	0.044949515	9.668737106	193.0773861	5182.49945	83.77260787	30.93219067	0.590103412	5649.256846	124.9416537
33	1	4	6	135.4081107	15.79949141	74.07104509	670.8748923	0.061297203	9.297919189	193.0607722	5383.808504	88.14850236	28.77347417	0.552975979	5858.757024	102.6624758
34	1	4	7	133.3067885	15.31742447	72.71425614	591.9509869	0.036370817	11.19165349	214.7605168	5020.863838	87.17237534	31.19110959	0.482155562	5436.652427	[illegible]
35	1	4	8	101.4975012	22.69593393	72.96419976	798.1883854	0.04206839	8.219584611	236.9064094	5240.636367	75.93851205	28.66738102	0.534763073	[illegible]	[illegible]
36	1	4	9	116.7992965	20.31185872	87.60941833	515.0876363	0.049487648	10.407549	232.8346185	5097.661808	83.51474644	31.03968978	0.484644958	[illegible]	[illegible]
37	1	5	1	128.7529134	18.52693269	75.82153802	528.270574	0.053261144	9.13674309	226.2352173	5169.051076	78.80407066	27.27444439	0.548208934	[illegible]	[illegible]
38	1	5	2	86.49613186	20.07774036	88.82383116	731.7281643	0.058578695	10.32255421	237.0707945	4779.655217	89.71849435	30.52147	0.516333513	[illegible]	[illegible]
39	1	5	3	137.614093	23.86514486	75.96103595	505.6100129	0.063184549	11.636322	201.1982657	4838.584884	86.24739771	28.33353816	0.433199212	[illegible]	[illegible]
40	1	5	4	142.2939046	15.38447461	83.23681777	790.8387557	0.039175009	9.87957714	239.605726	4808.41016	87.58087293	31.3032063	0.411824092	[illegible]	[illegible]
41	1	5	5	137.4645209	23.10251831	80.75522354	503.6986091	0.049832772	11.45414779	212.2410086	5098.660206	78.28039366	27.43459612	0.573422973	[illegible]	104.7581828
42	1	5	6	132.2561779	20.71331297	75.31440685	558.6029301	0.039666673	10.24897734	231.0594793	5324.313303	78.41486314	31.9598974	0.466774265	5101.241423	136.8319507
43	1	5	7	146.3522795	22.79723204	70.00015153	503.5015395	0.030877889	9.306648981	214.3241148	5175.295924	89.47191474	27.73243521	0.441607513	5134.93889	135.7335823

Data Preparation for AI

Synthetic Data Generation using Digital Twins



BAUERBERG KLEIN

TRAINING & CONSULTING

BAUERBERG KLEIN
TRAINING & CONSULTING