Using Machine-Learning for Depositional Facies Prediction in a Complex Fan Delta Reservoir, El Morgan Field

GPC Workshop, Oct, 2023

Dr. Alaa Hassan, Mr. Mohamed Abuelmajd GUPCO

Outline

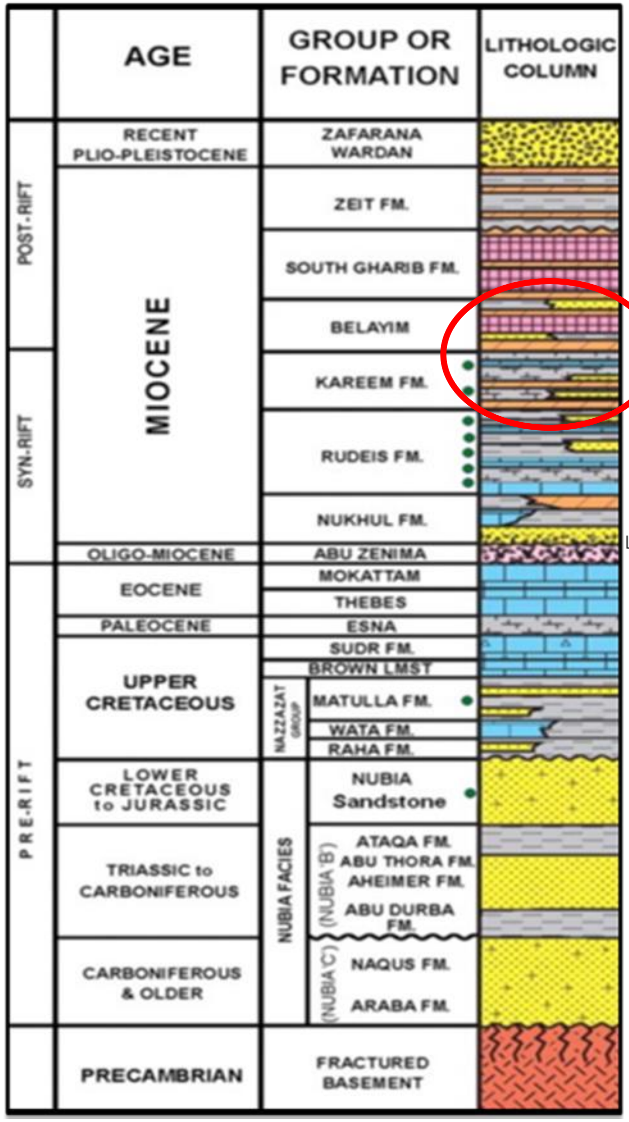
- Introduction
- ML methods and performance metrics
- Facies classification results
- 3D Facies Models comparison
- Conclusion & Recommendations

Introduction

- El Morgan field discovered in 1965 and started production in 1967

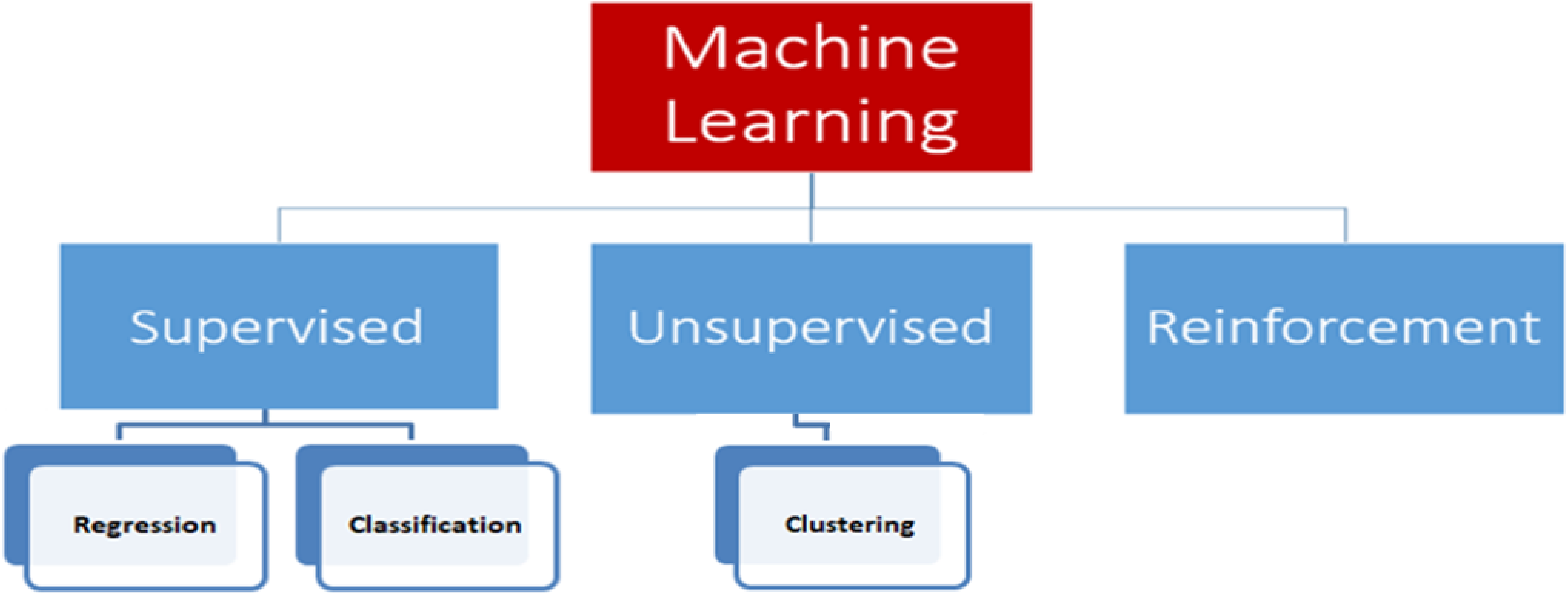
- Two reservoirs, Belayim and Kareem, porosity (15-25%) and permeability (<1-10000 mD)

- Reliable facies prediction is critical to exploit areas of bypassed oil and water flooding optimization



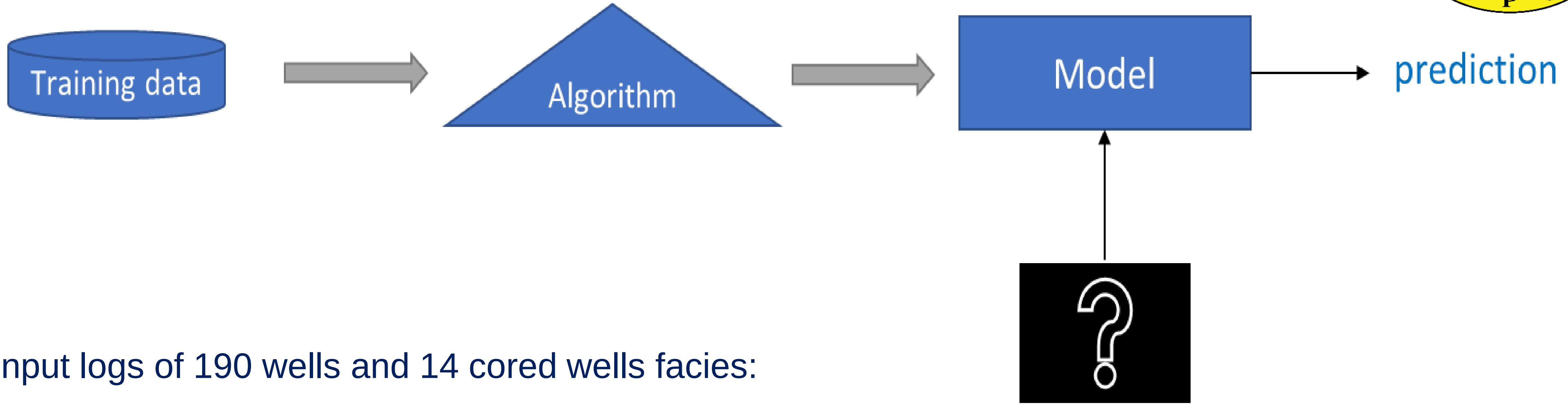
Gulf of Suez Stratigraphy

Machine Learning Methods



- ML gives computers the ability to learn automatically without being explicitly programmed (Arthur, 1959)
- ML applied on the Morgan reservoirs utilizing python-based ML (*sickit-learn* libraries)

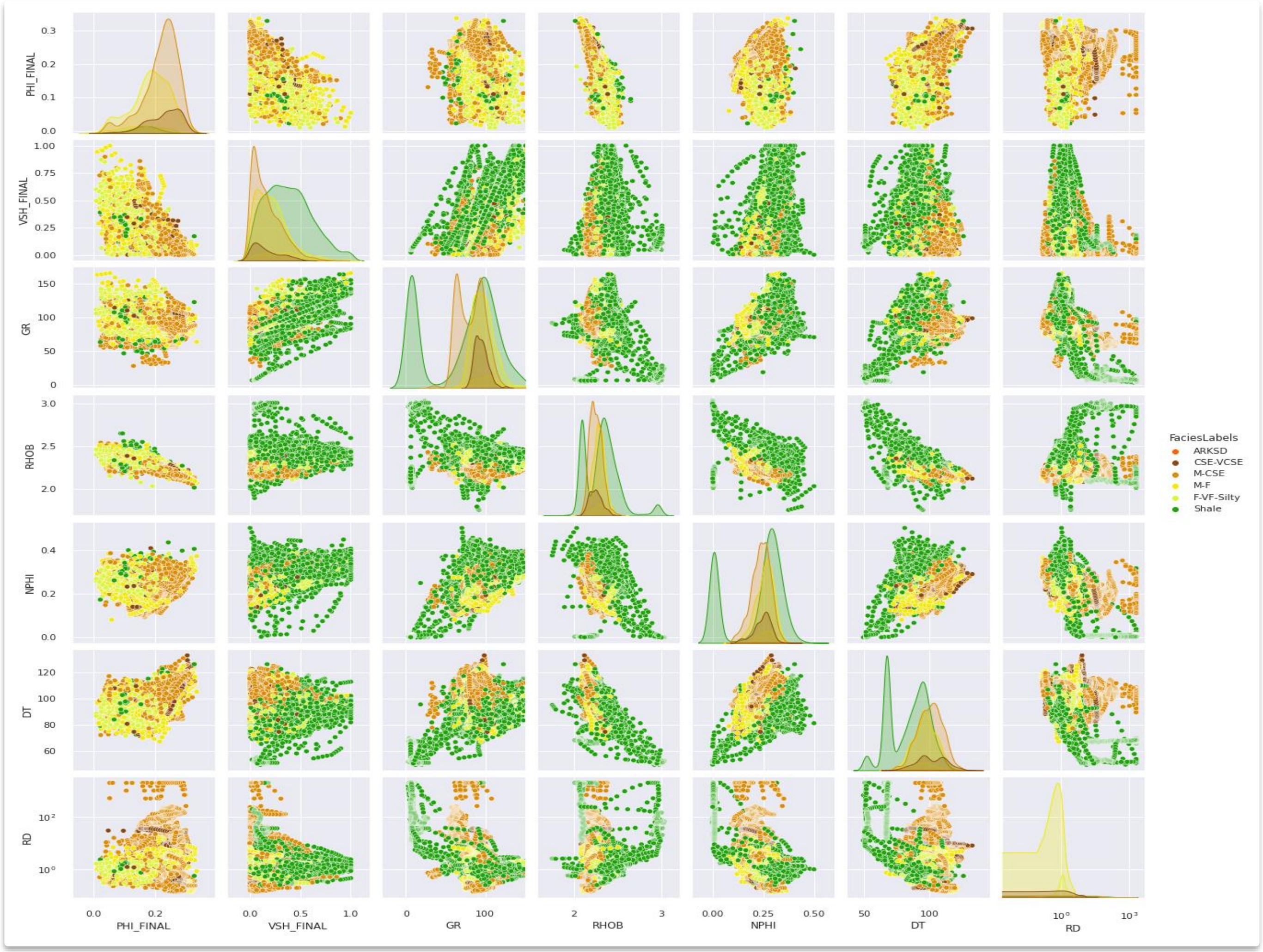
Machine Learning Procedure



- Input logs of 190 wells and 14 cored wells facies:
  1) Gamma ray (GR)
  2) Resistivity (RD)
  3) Sonic log (DT)
  4) Neutron-density (NPHI-RHOB)
  5) Porosity (PHI_Final)
  6) Shale volume (VSH_Final)
- Data preparation, model choosing, scatter plots, training (XGBoost, KNN, NNC, SVM, etc…) and evaluating

Facies	Description	Label
1	Shale, silty sandstone, dolomite and mudstone with ichno fossils	0
2	Fine sandstone, mature, silty sandstone laminated with high silt intercalations , well sorted, Distal fan delta	1
3	Medium to fine sandstone, mature, moderately to well sorted, massive bedded, fining upward Mid fan delta	2
4	Medium to coarse sandstone, sub mature, moderately to pore sorted, coarsening upward , Prograding Mid fan	3
5	Coarse to very coarse grading to conglomerates , poorly sorted, immature, coarsening upward, thick bedded, structreless, Upper fan	4
6	Arkosic Sandstone, Immature, massively bedded, structure less, Prograding Upper fan	5

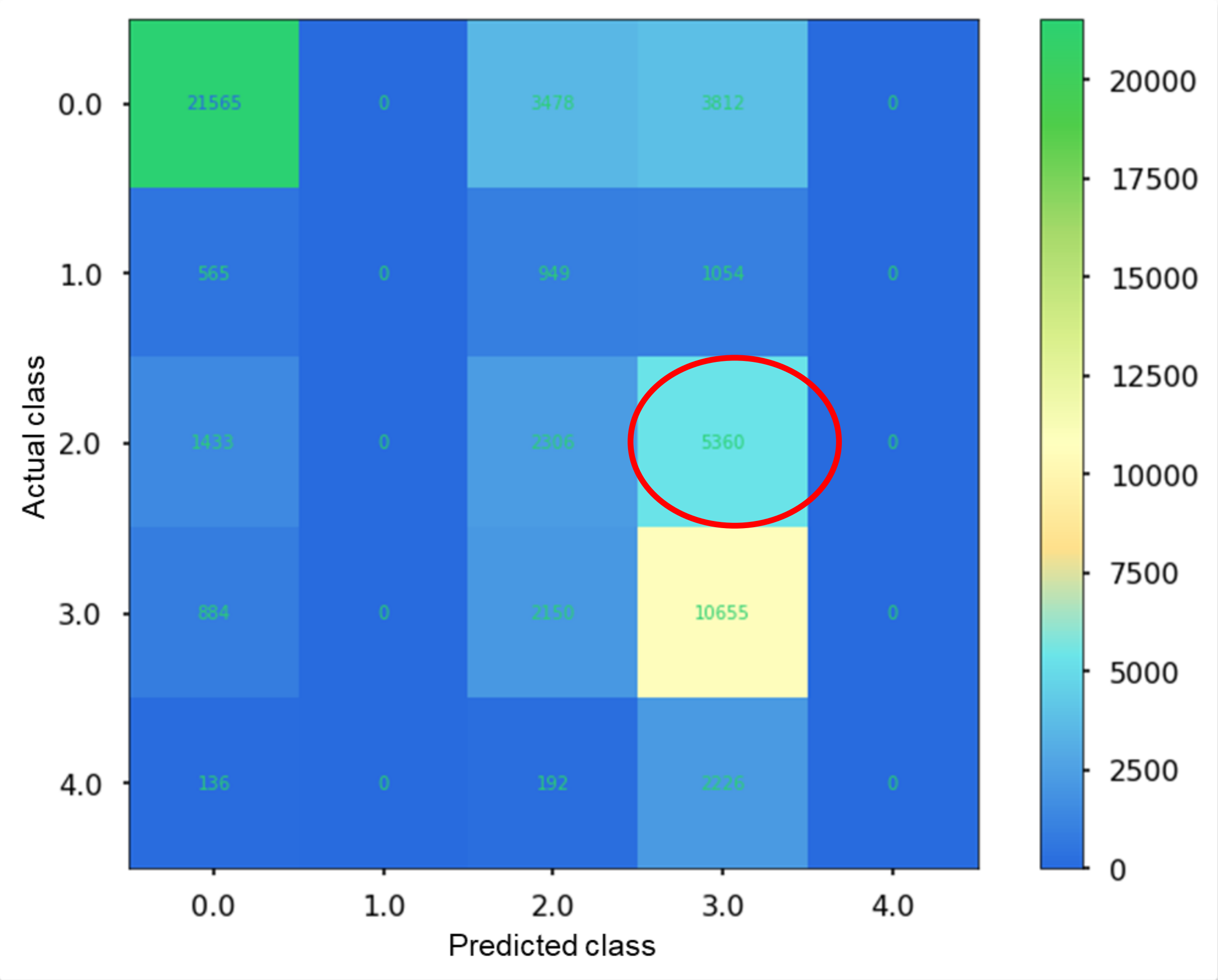
Scatter plot of input variables



Scatter plot of input variables prior to data conditioning

Visualize the pattern between different logs and facies. The diagonal shows the statistical distribution of each well log with different facies classes. The off-diagonal shows relationships between different well logs

Performance of a classification algorithm



Confusion matrix provides examine the performance of a classification algorithm

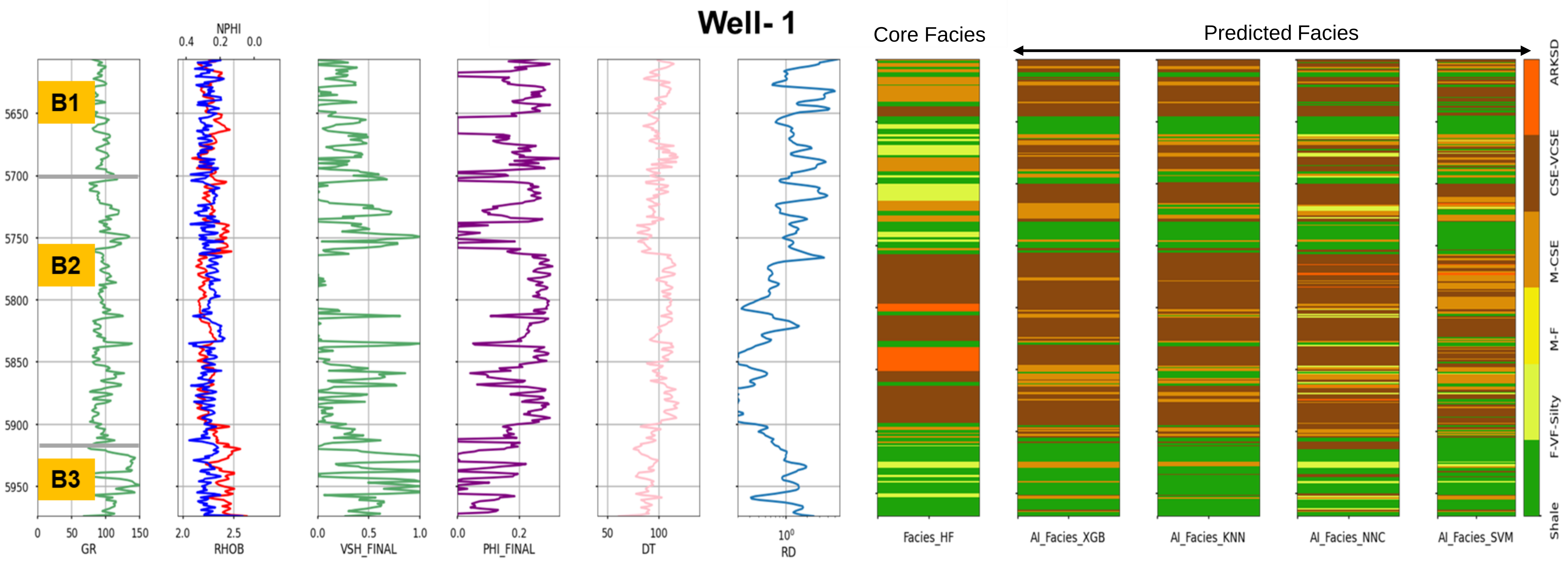
The principal diagonal of the matrix gives a better visualization of the prediction behavior. The off-diagonal elements of the matrix tell the ratio of misclassification for each facies class

Performance Metrics

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

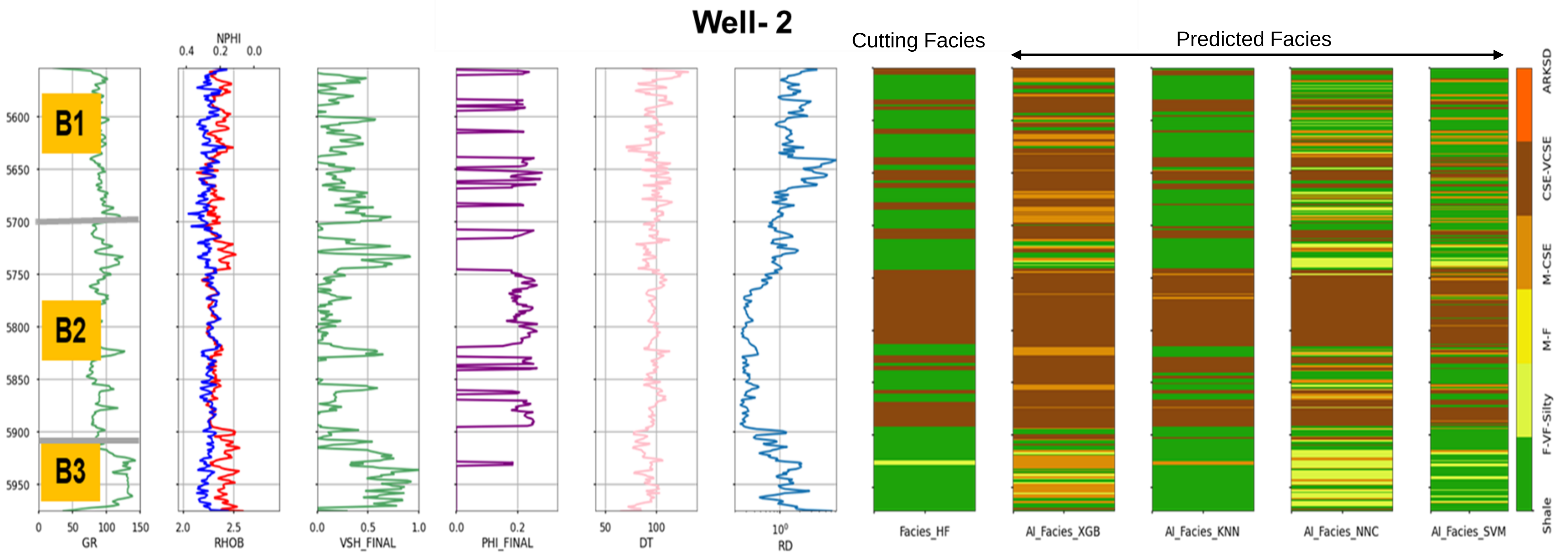
- **Accuracy:** The most commonly used metric and represents the percentage of correct predictions
  TP +TN / TP + FP +TN + FN
- **F1 score:** weighted average of recall and precision that measures a model's accuracy
  F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Cored Well Training Vs. Test Facies Prediction (BEL)



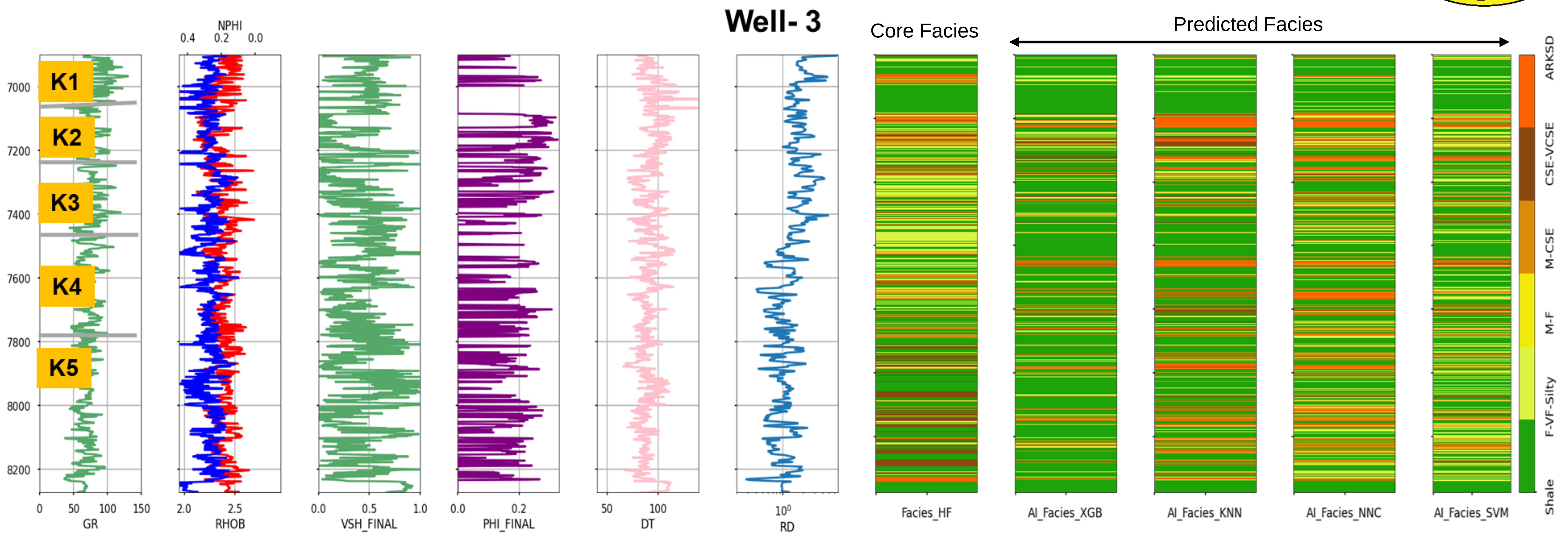
Well- 1 Correlation of well logs with facies prediction Vs. core facies

None Cored Wells Facies Prediction Different Models (BEL)



Well- 2 Correlation of well logs with facies prediction Vs. cutting facies

Cored Well Training Vs. Test Facies Prediction (KAR)

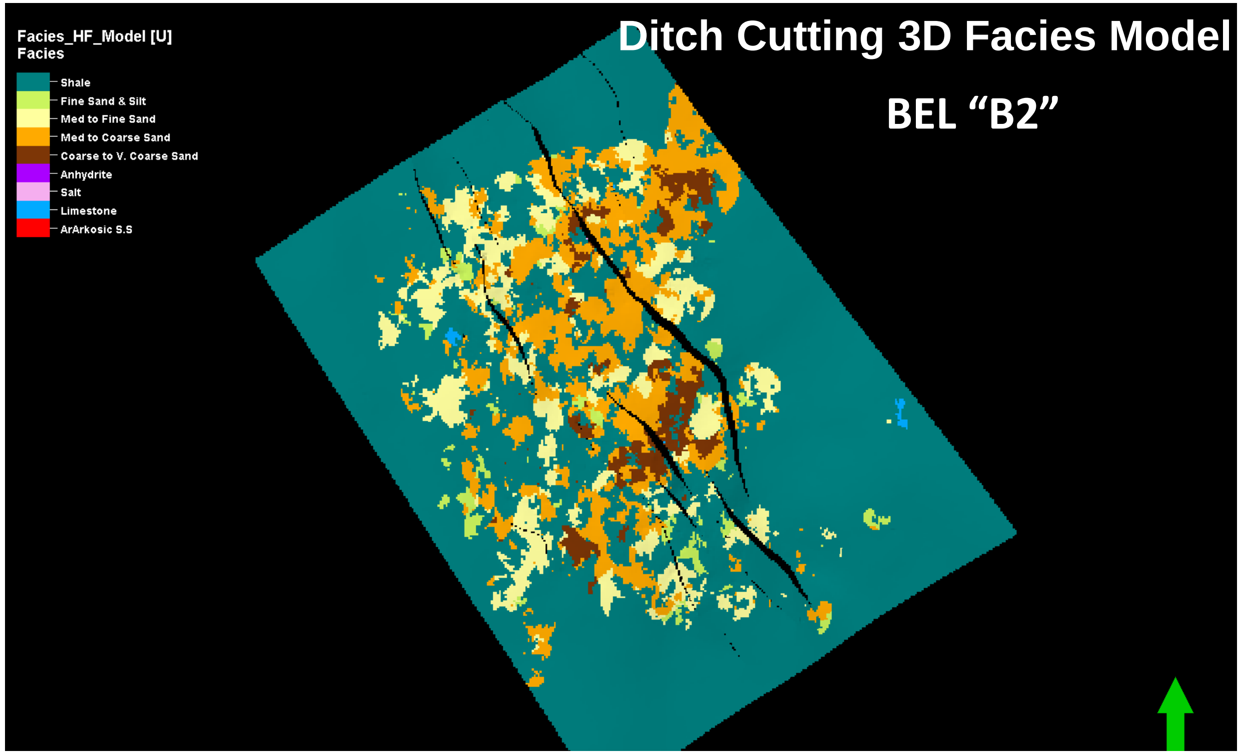


Well- 3 Correlation of well logs with facies prediction Vs. core facies

Non Cored Well Training Vs. Test Facies Prediction (KAR)

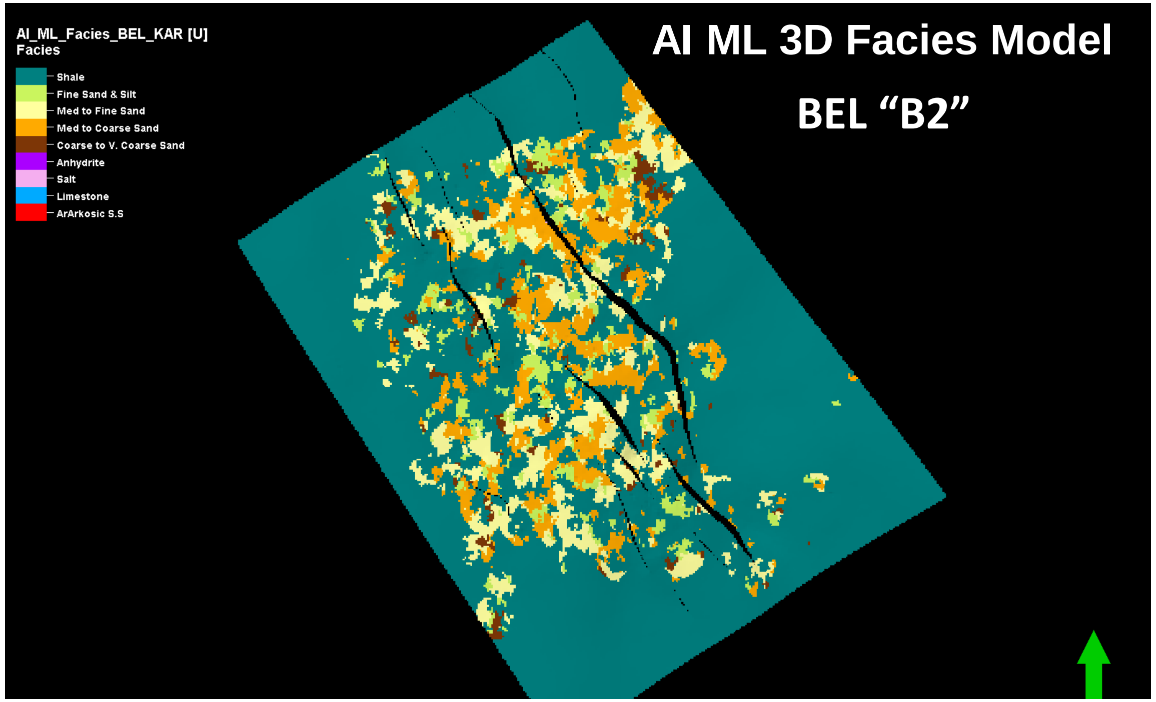
Well- 4 Correlation of well logs with facies prediction Vs. cutting facies

3D Facies Model Results (BEL)

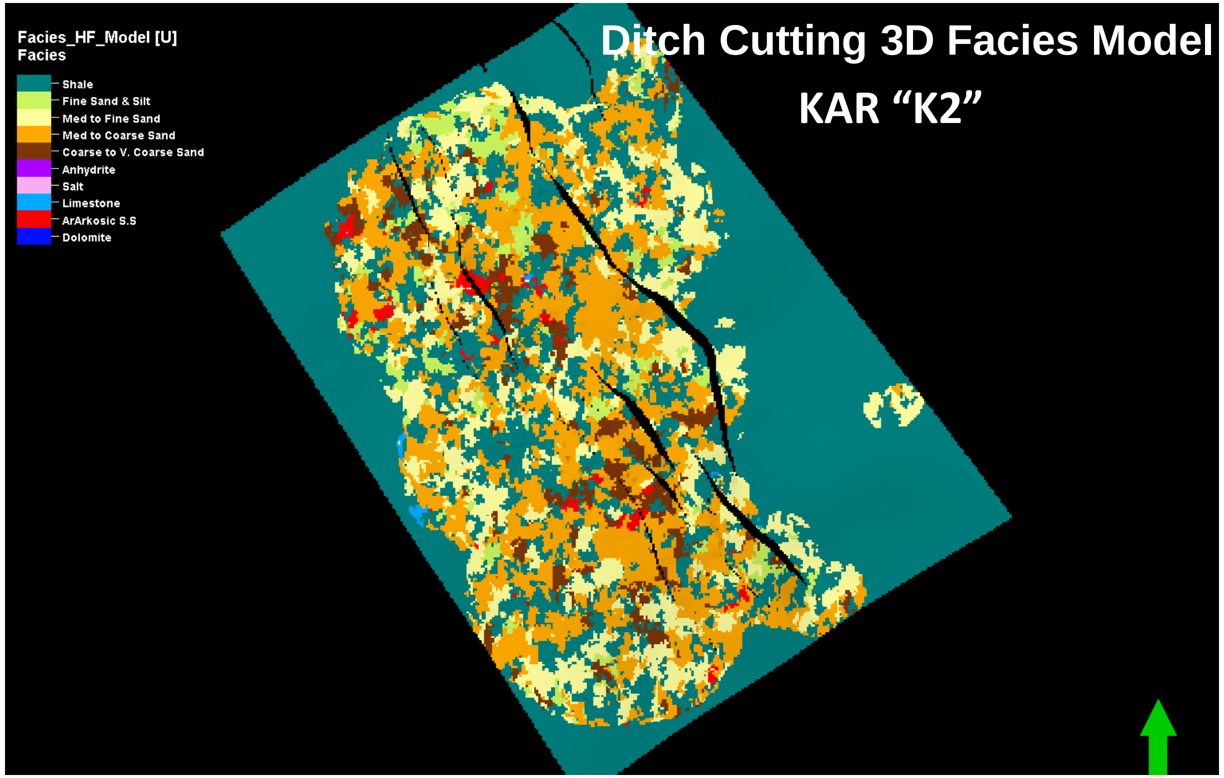


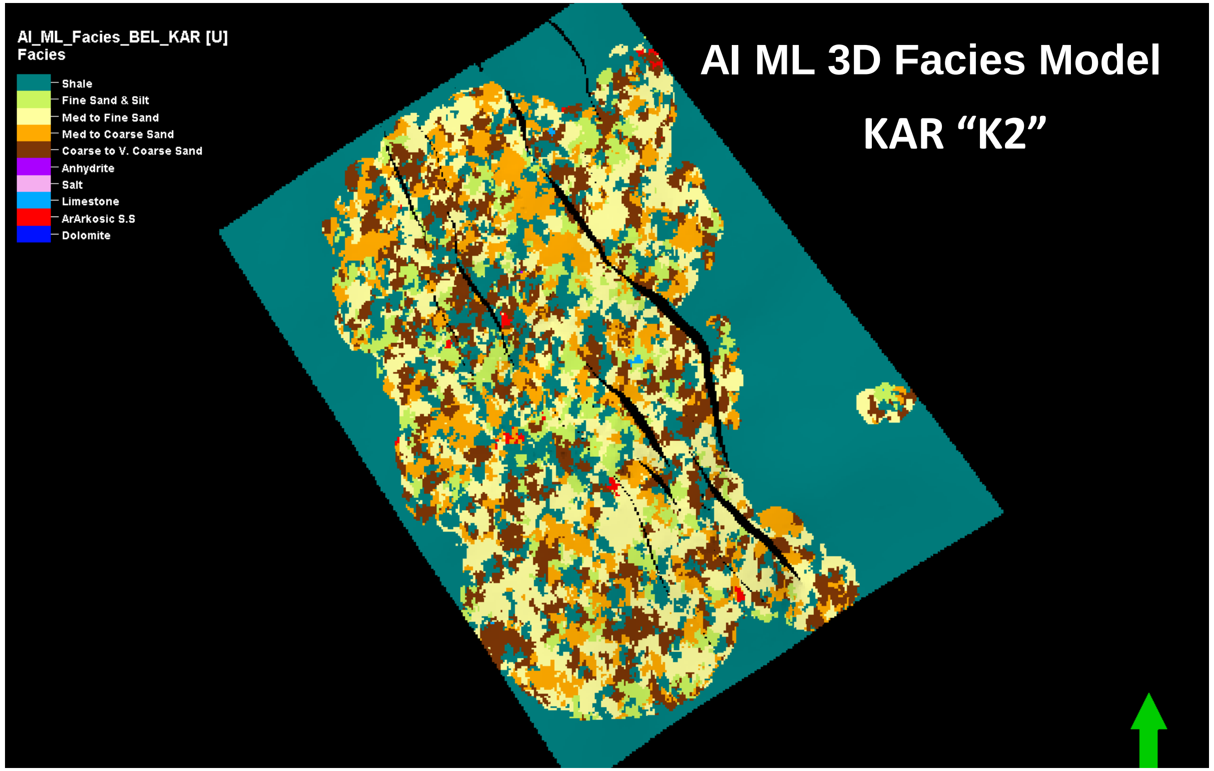
- **Ditch Cutting Facies Interpretation**
  - Advantages:
    1. Routinely have drill cuttings
    2. More control of wells facies analysis
    3. Variations in facies type and abundance
  - Disadvantages:
    1. Time consuming
    2. Possible contamination (representative sampling)
    3. Cuttings description accuracy

- **ML Facies Predictions**
  - Advantages:
    1. ML automate analyzing and interpretation of data
    2. Improved accuracy (accurate predictions)
    3. Scalability (handle larger datasets)
  - Disadvantages:
    1. Overfitting and underfitting
    2. Data dependency (depend on input data quality)



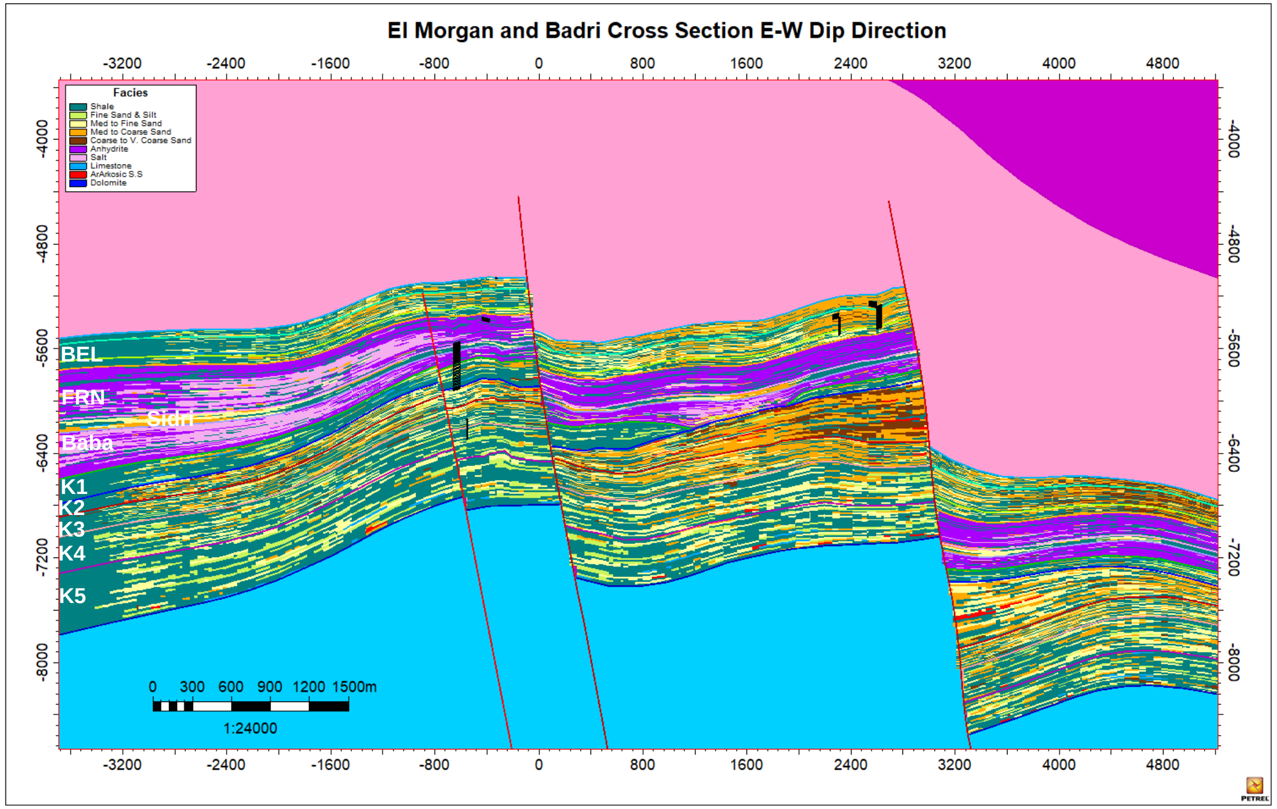
3D Facies Model Results (KAR)

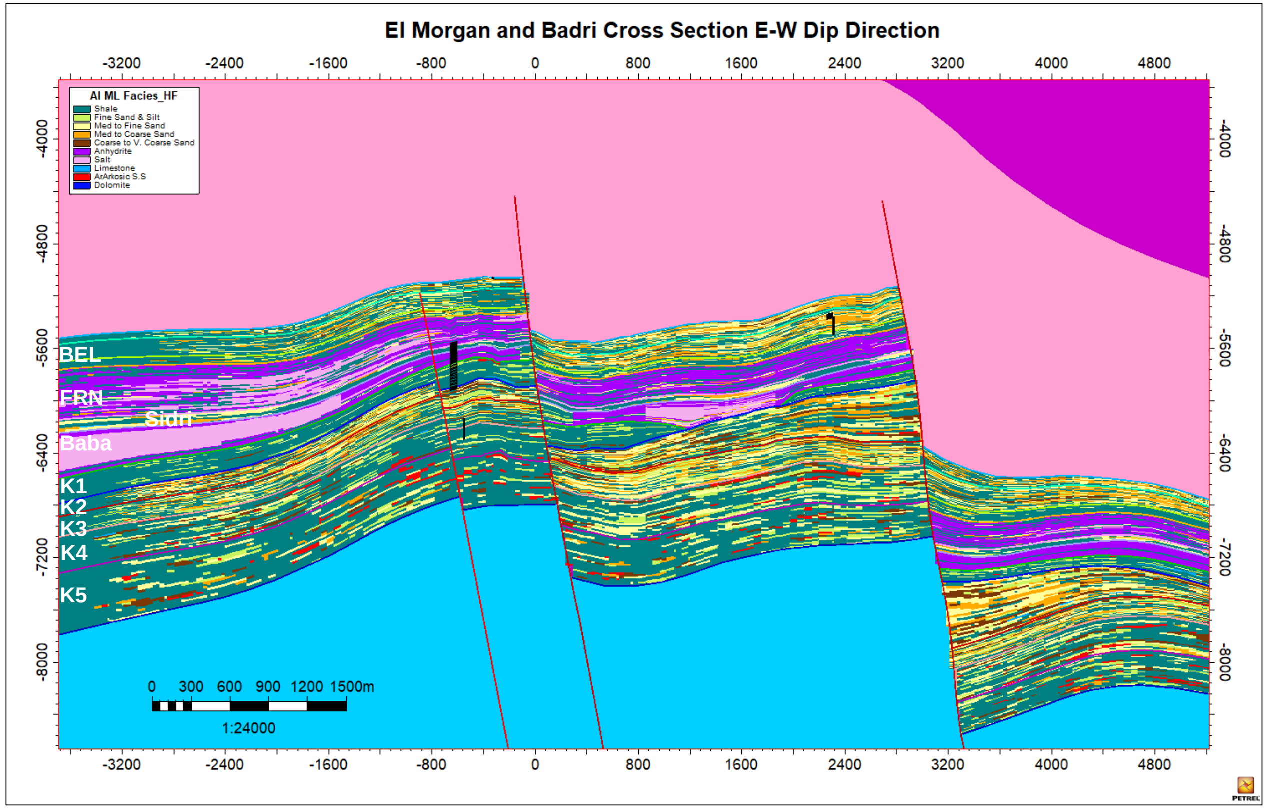




ML classifiers provide reasonable prediction on the facies classes of sand, while perform poorly when facies classes are mostly overlapped with other facies, which makes it very difficult to identify by the classifiers

3D Facies Model Results (BEL & KAR)





ML facies prediction provides acceptable estimate on the facies classes of sand in Belayim and Kareem zones, except for K5, where facies classes (Arkosic sand, red color & light green) are mostly overlapped with other facies

Conclusion & Recommendations

- ML supervised algorithms were applied on a large dataset of Morgan complex fan delta reservoirs

- The experimental results demonstrated the high efficiency of the developed workflow for automatic facies classification with reasonable prediction accuracy

- To improve the ML models' accuracy, introduce more training data samples to models with different facies classes and optimize model parameters

Thank You

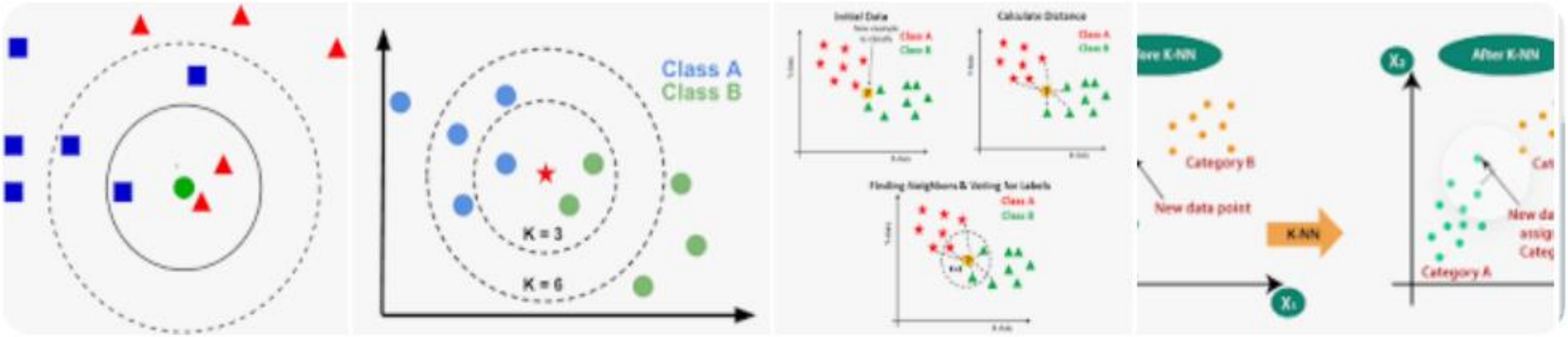
Backup

eXtreme Gradient Boosting (XGB)

XGBoost (eXtreme Gradient Boosting) is a popular supervised-learning algorithm used for regression and classification on large datasets. It uses sequentially-built shallow decision trees and the predictions from each tree are combined to form the final prediction. It is a highly-scalable training method that avoids overfitting.

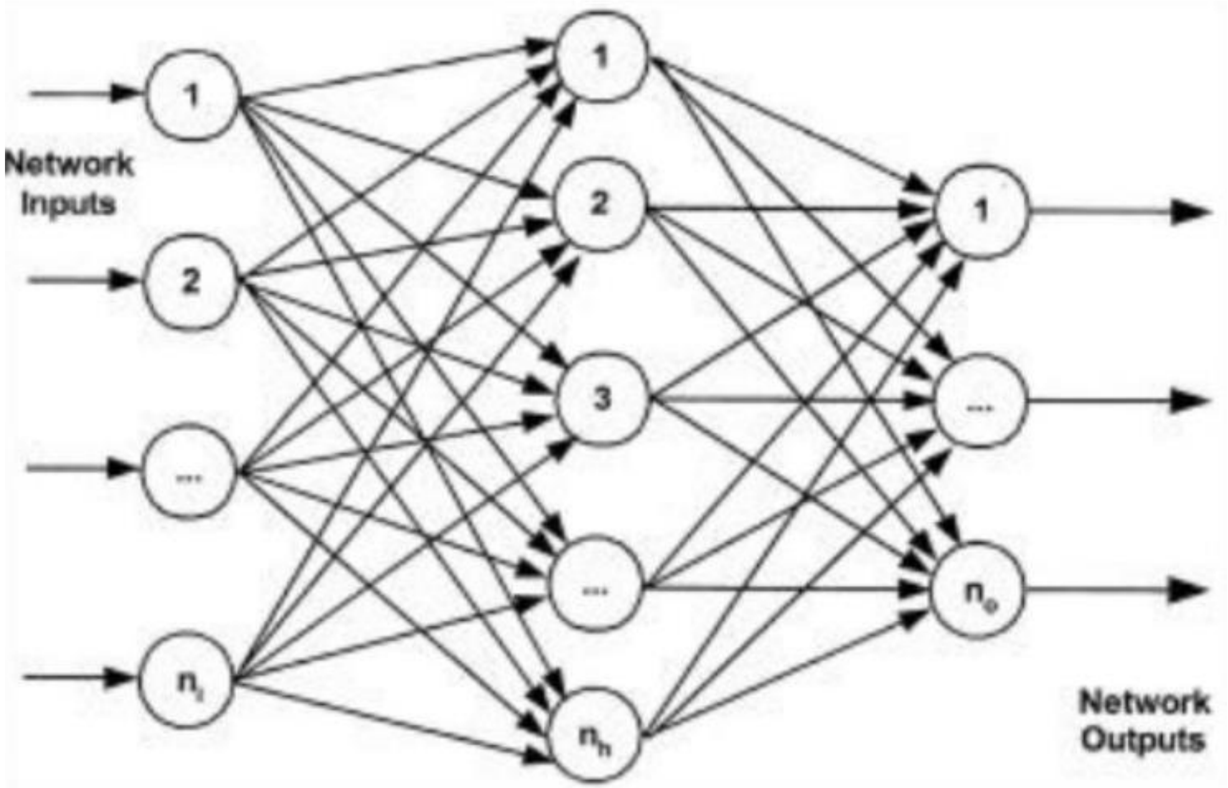
K-Nearest Neighbors

In pattern recognition, the k-Nearest Neighbors algorithm (KNN) is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point (Keller et al. 1985).



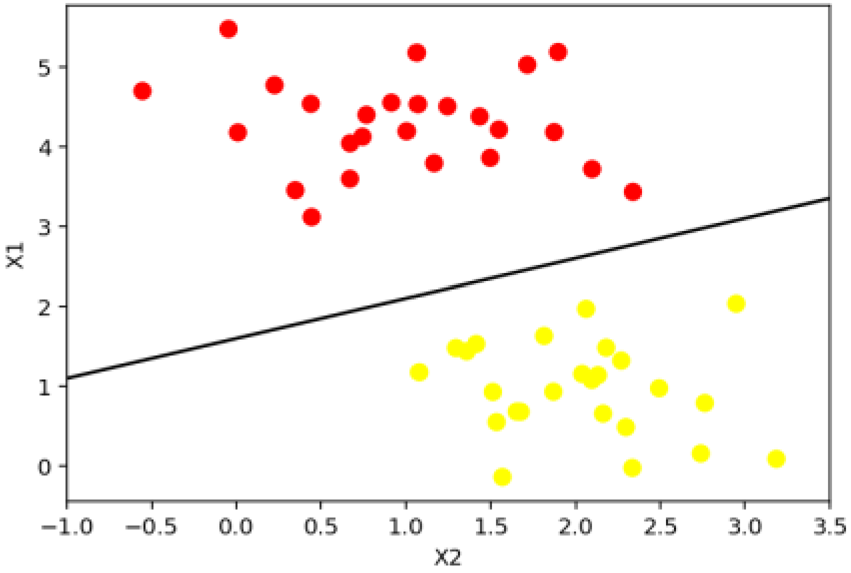
Neural Network

Neural networks in machine learning refer to a set of algorithms designed to help machines recognize patterns without being explicitly programmed. They consist of a group of interconnected nodes. These nodes represent the neurons of the biological brain.

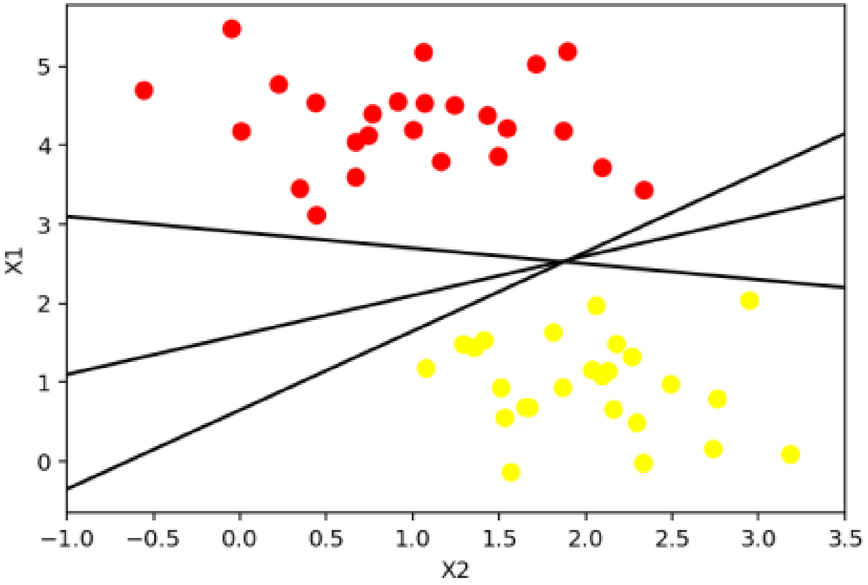


Support Vector Machine

Support Vector Machine (SVM) is a multipurpose machine learning algorithm, capable of performing linear or nonlinear classification, regression, and even outlier detection (Nishitsuji and Exley 2019)



An illustration of a hyperplane separates two types of samples. Yellow and red circles represent different types of samples.



An illustration of a hyperplanes that can separate two types of samples. Yellow and red circles represent different types of samples.

They hyperplane is a decision boundary that divides the input space into two or more regions, each corresponding to a different class or output label. In a 2D space, a hyperplane is a straight line that divides the space into two halves