Capstone Project 2: Rocks Lithofacies Classification Using Machine Learning

Honglin Yuan, Oct 2022

What is lithofacies?

litho-facies

Lithology - type of rocks

- sandstone,
- shale,
- limestone
- etc.

Depositional environments:

- Fluvial vs. marine
- shallow water vs. deep water

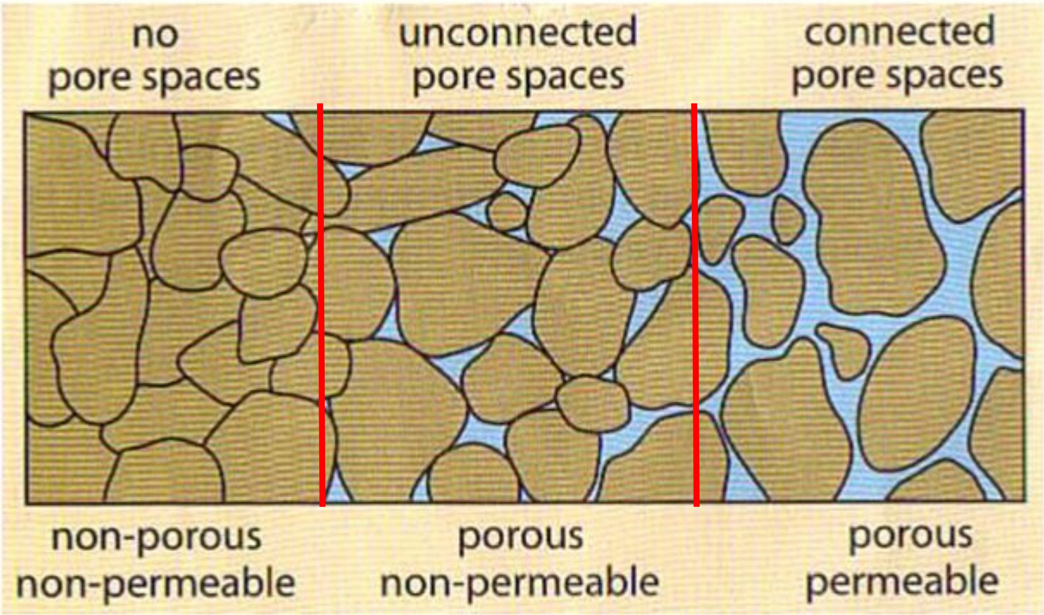
deep marine shale

Why is lithofacies important?

Different lithofacies have different:

- Grains size & sorting
- Pore size
- Connectivity between pores

Reservoir quality



UC Denver

How to obtain lithofacies?

- Core sampling tool
  - accurate,
  - but extremely expensive
- Geologist’s interpretation from wireline logs
  - highly skilled geologists
  - slow
- Supervised machine learning
  - quick and cheap
  - accuracy to be investigated

Dataset

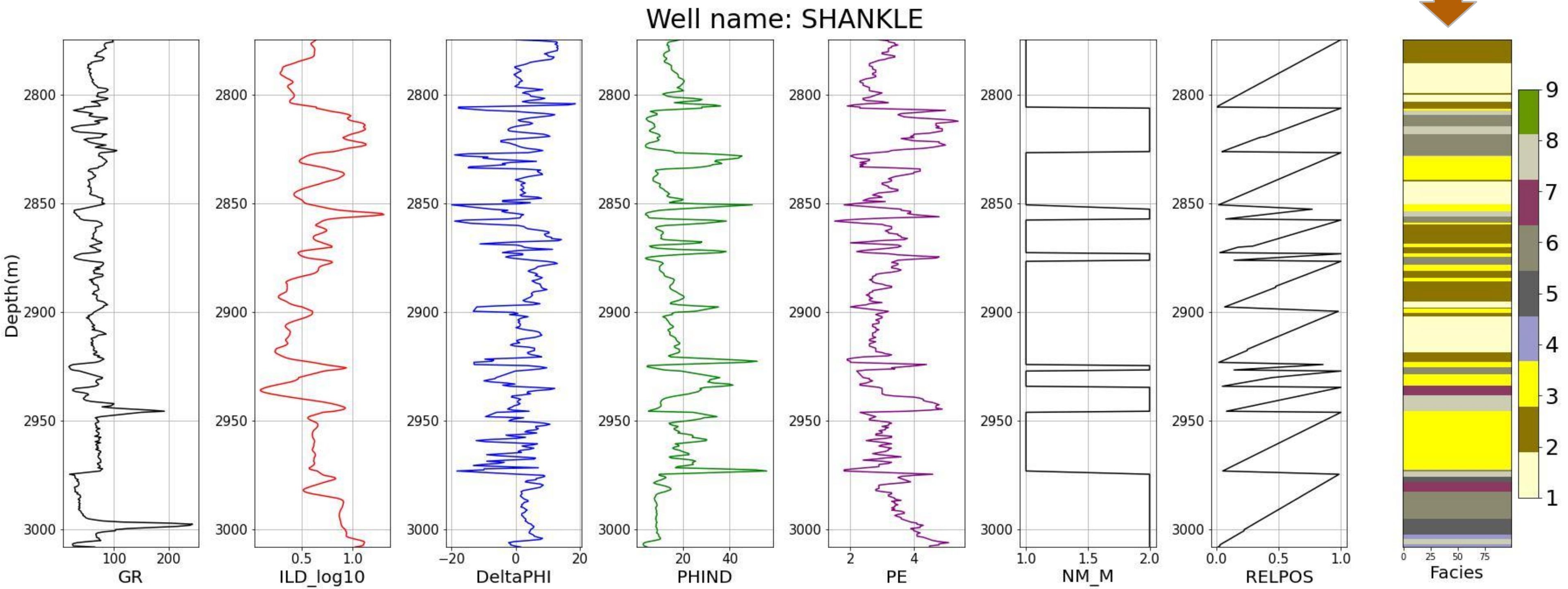
- 10 wells in the Hugoton and Panoma Fields in North America
- 11 features including target feature “Facies”
- Organized by “Well Name”, order by ‘Depth”

Target feature **Wireline measurements or derived logs**

	Facies	Formation	Well Name	Depth	GR	ILD_log10	DeltaPHI	PHIND	PE	NM_M	RELPOS
0	3	A1 SH	SHRIMPLIN	2793.0	77.45	0.664	9.9	11.915	4.6	1	1.000
1	3	A1 SH	SHRIMPLIN	2793.5	78.26	0.661	14.2	12.565	4.1	1	0.979
2	3	A1 SH	SHRIMPLIN	2794.0	79.05	0.658	14.8	13.050	3.6	1	0.957
3	3	A1 SH	SHRIMPLIN	2794.5	86.10	0.655	13.9	13.115	3.5	1	0.936
4	3	A1 SH	SHRIMPLIN	2795.0	74.58	0.647	13.5	13.300	3.4	1	0.915

Visualization - SHANKLE well

Target feature



Target feature - Facies

9 Facies values:

1 - SS: Nonmarine sandstone

2 - CSiS: Nonmarine coarse siltstone

3 - FSiS: Nonmarine fine siltstone

4 - SiSH: Marine siltstone and shale

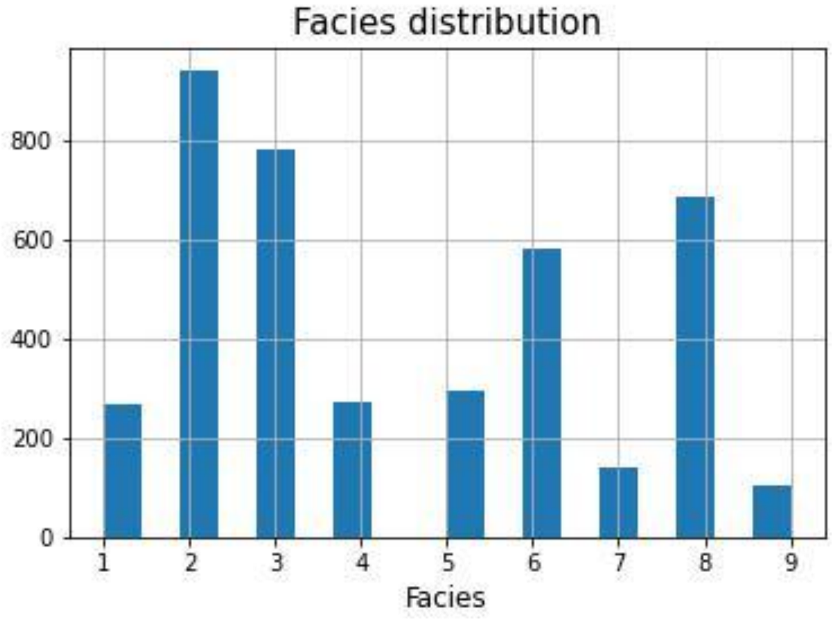
5 - MS: Mudstone (limestone)

6 - WS: Wackestone (limestone)

7 - D: Dolomite

8 - PS: Packstone-grainstone (limestone)

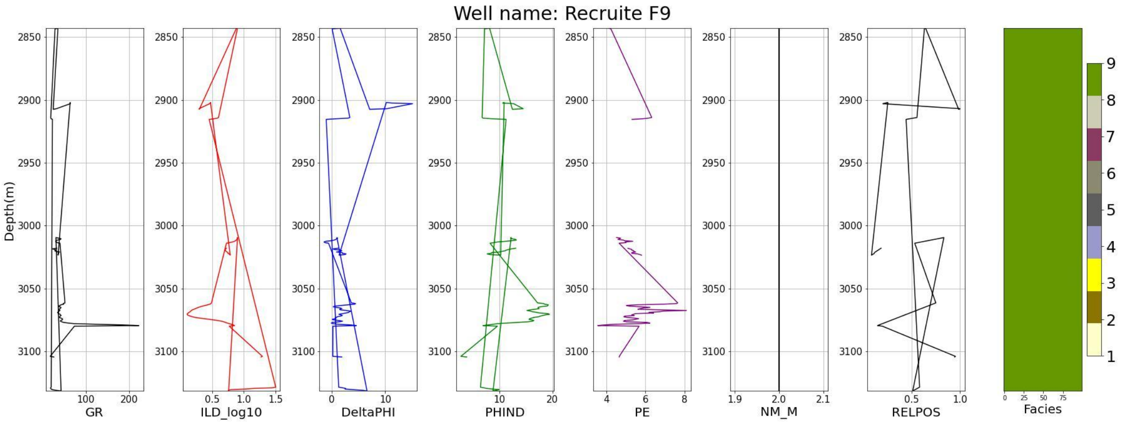
9 - BS: Phylloid-algal bafflestone (limestone)



Data quality

Two issues with the dataset:

- Bad data in one well - Removed
- Missing data in two wells



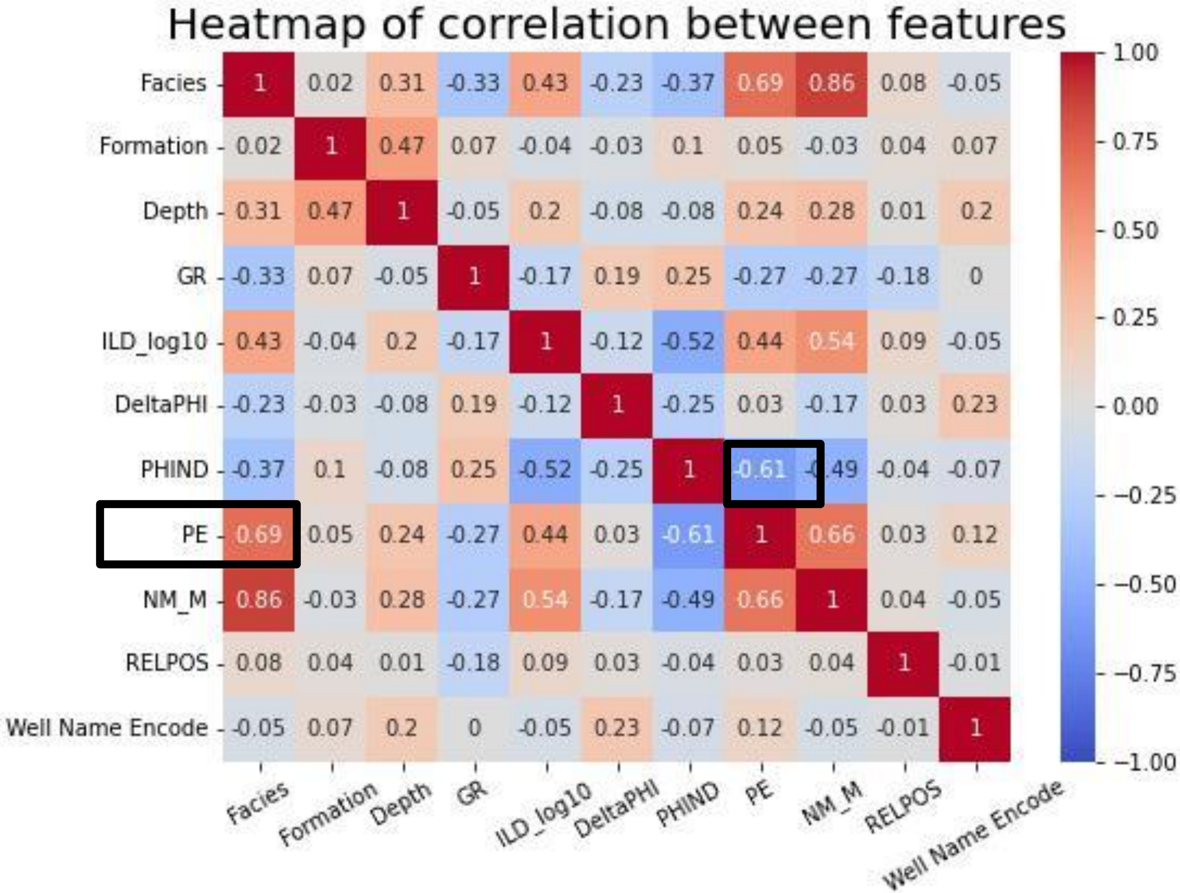
Missing data - how should we handle it?

- Three options:
    - Drop PE feature
    - Drop these two wells
    - Impute PE for these two wells

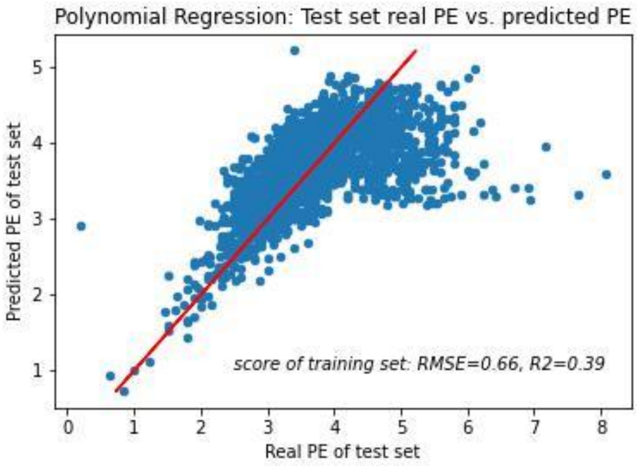
Is PE important?

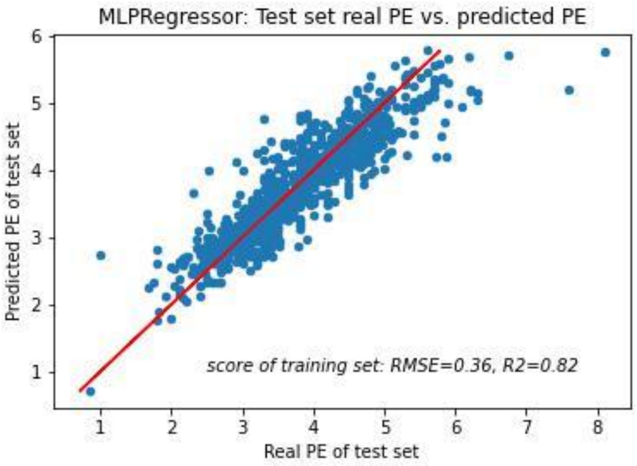
- **YES** - PE is highly correlated with target feature Facies

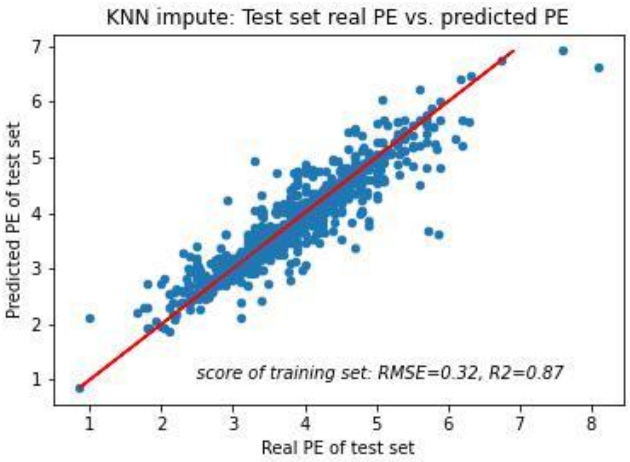
- PE is highly correlated with PHIND



PE Imputation - three approaches







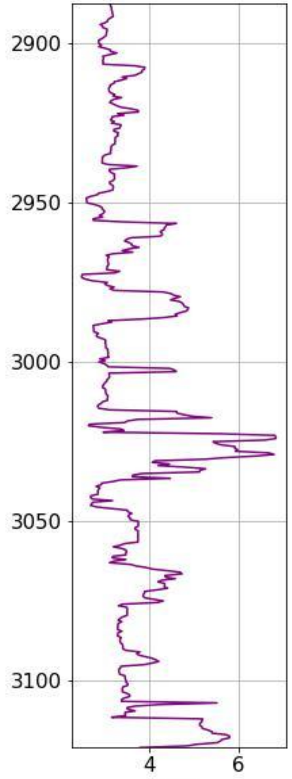
Polynomial Regression using PHIND

Multi-layer Neural Network using multiple features

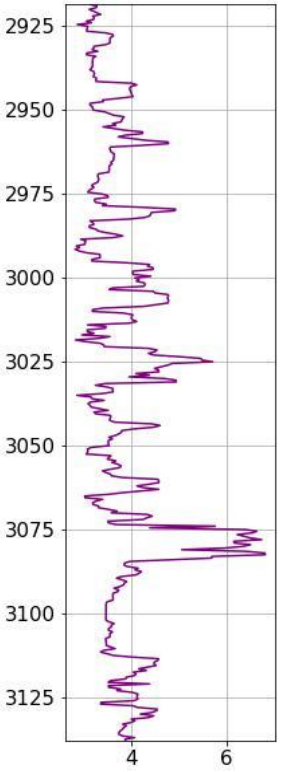
KNN fancyimpute using multiple features

Imputed PE

ALEXANDER D



KIMZEY A



Training and test data split

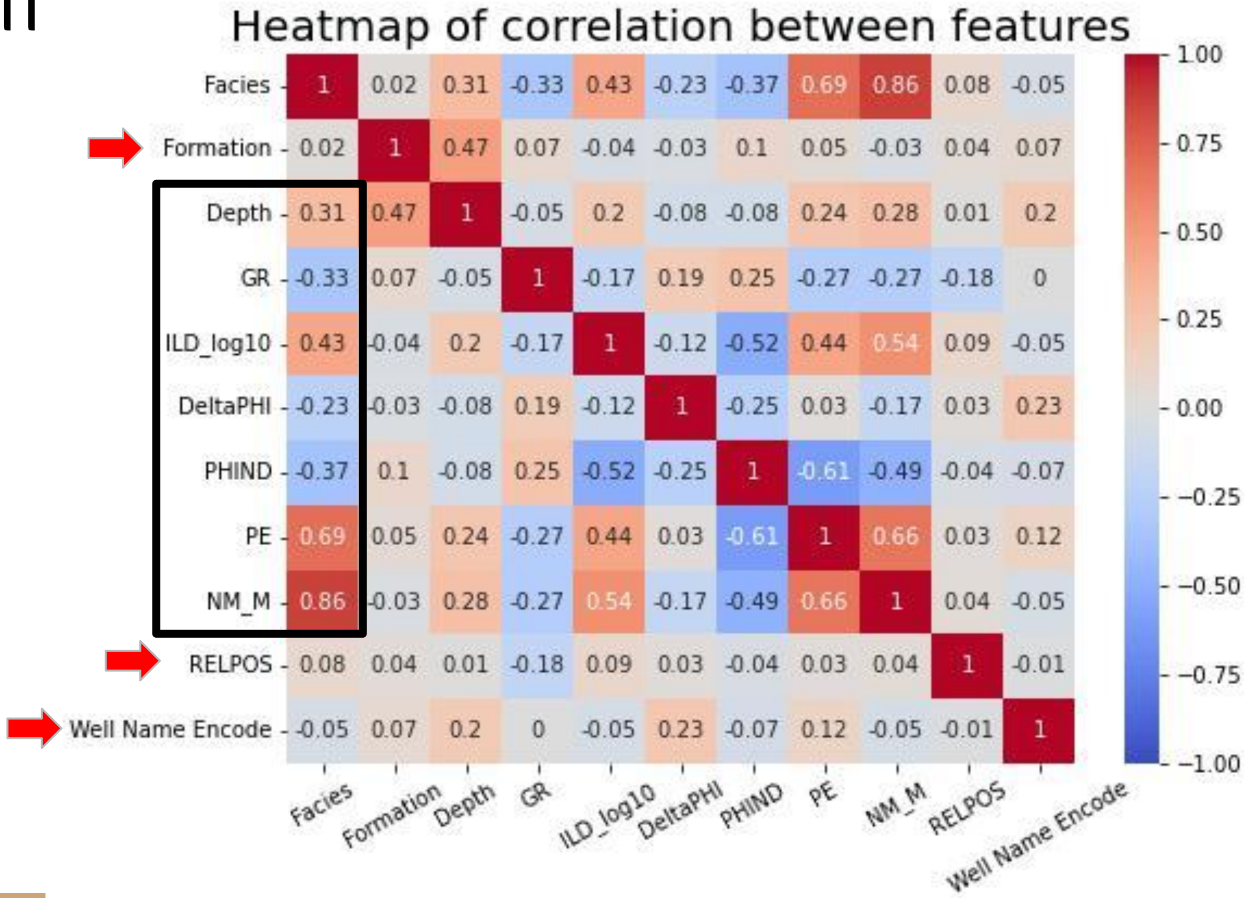
Test well: SHANKLE

Training dataset: the other 8 wells

Features selection

Dropped 3 features:

- Formation
- Well Name
- RELPOS



Modeling

- Three models
  - Random Forest
  - KNN
  - SVM
- Cross-validation & hyperparameters grid search

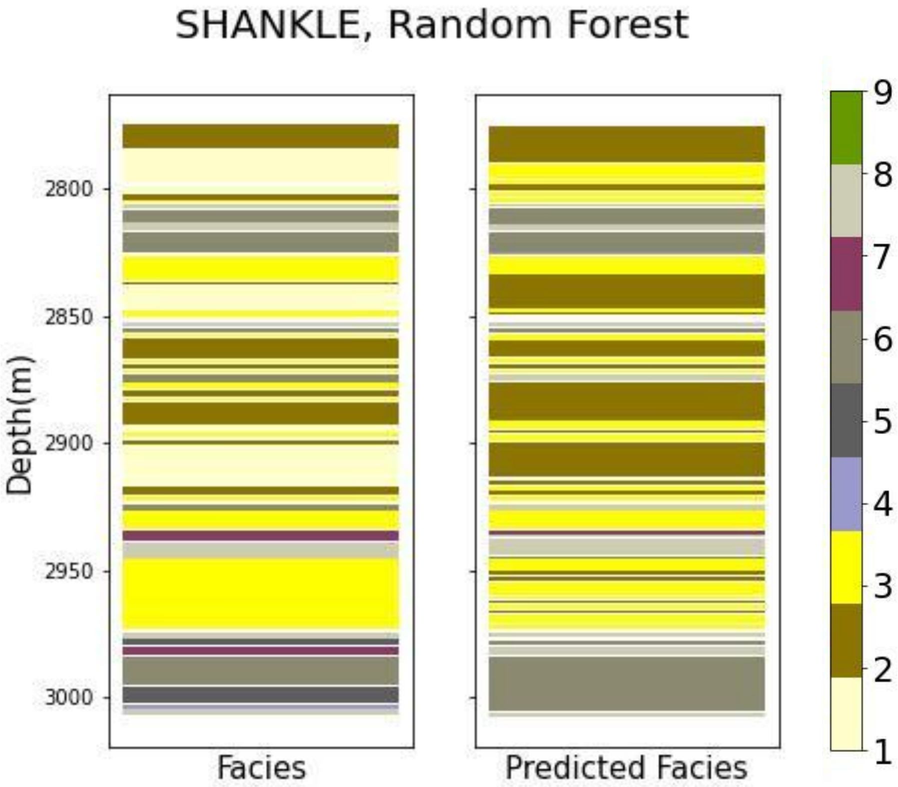
Model accuracy (cross-validation)

Final model

	Random Forest	KNN	SVM
Cross-validation accuracy	0.53	0.45	0.5

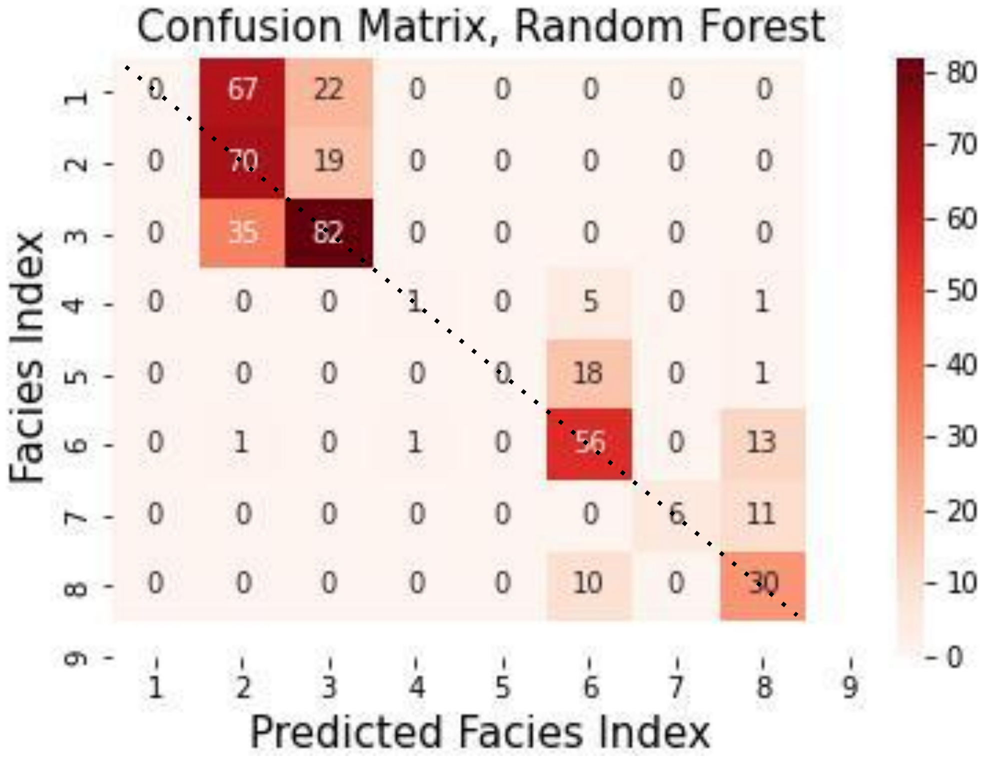
Test well prediction result

Test well prediction accuracy: 0.55



Mis-classification analysis

- Facies mislabeling lie mostly in the neighboring facies
- In reality, facies often changes gradually
- If mis-classification with neighboring facies is tolerated, accuracy increases from 0.55 to **0.91**



Summary

Random Forest model performs better than KNN and SVM for this particular dataset.

Model accuracy is ~0.53 under cross-validation.

If mislabeling with neighboring facies is tolerated (considering the subtle difference between them), model accuracy increases to 0.91.

Machine learning may not be able to accurately classify adjacent lithofacies, but still can be used to separate lithofacies with larger difference.