Well-Log Based Determination of Rock Porosity Through Machine Learning Algorithms

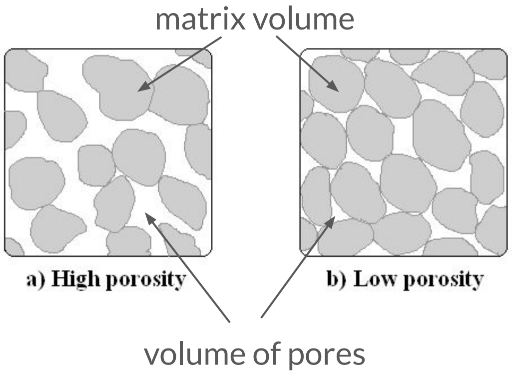
Sharifullina Ramilya
Elizaveta Gladchenko
Ilya Nachinkin

Porosity

$$\phi = \frac{V_p}{V_b}$$

V_p = volume of pores

V_b = bulk (matrix) volume



Classical approaches:

- CT scan
- laboratory tests

Logging - the process of recording and analyzing measurements collected discretely or continuously in a wellbore.

The group of porosity logs includes:

- density log
- neutron log
- sonic log

The main contributions of this project

- performance of several standard machine learning models while solving the task of well-log interpretation was observed
- peculiar properties of the task such as problems with real data, feature importance, adaptation of the model to the data from different wells were investigated
- supervised machine learning algorithms (FNN, RNN) were implemented
- comparison of the results obtained by neural networks and standard machine learning algorithms was performed

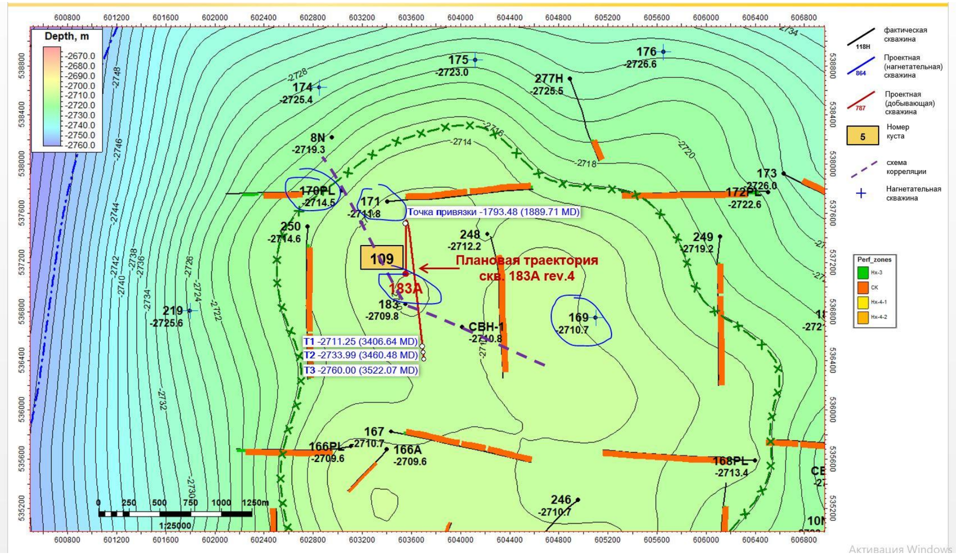
Research Object

As an input we have <u>4 wells</u> : 169, 170, 171, 183

<u>Features:</u>

- gamma ray log
- resistivity log
- caliper
- sonic log (s-wave + p-wave)
- neutron log
- density log

<u>**Target variable:**</u> **porosity**



Techlog
Because every well counts

Machine Learning algorithms

1) Linear Regression
2) kNN (`weights`: 'distance', `n_neighbors`: 3)
3) Random Forest (`n_estimators`: 10, `max_features`: 'sqrt', `max_depth`: None, `bootstrap`: False)
4) XGBoost (`max_depth`: 3, `booster`: 'dart')
5) FNN (3-layer NN with 3 hidden layers)
6) RNN (1-layer LSTM mode with `hidden_size` = 30

ML ALGORITHM	RMSE	MAE	R^2
LINEAR REGRESSION	0.0122	0.0102	0.9024
K-NN	0.0183	0.0128	0.7795
RANDOM FOREST	0.0113	0.0079	0.9152
XGBOOST	0.0094	0.0061	0.9413
FNN	0.0207	0.0156	0.7162
RNN	0.0092	0.0068	0.9454

Accuracy metrics

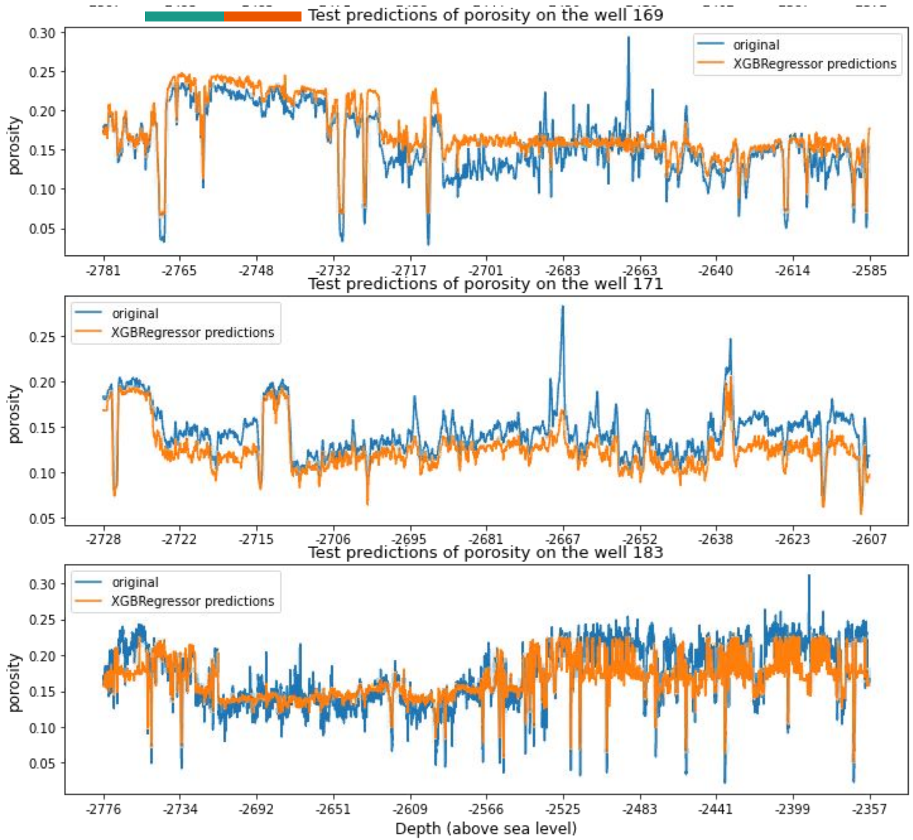
- RMSE - root mean squared error
- MAE - mean accuracy error
- R^2 - coefficient of determination

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\phi_i^t - \phi_i^p)^2}{N}},$$

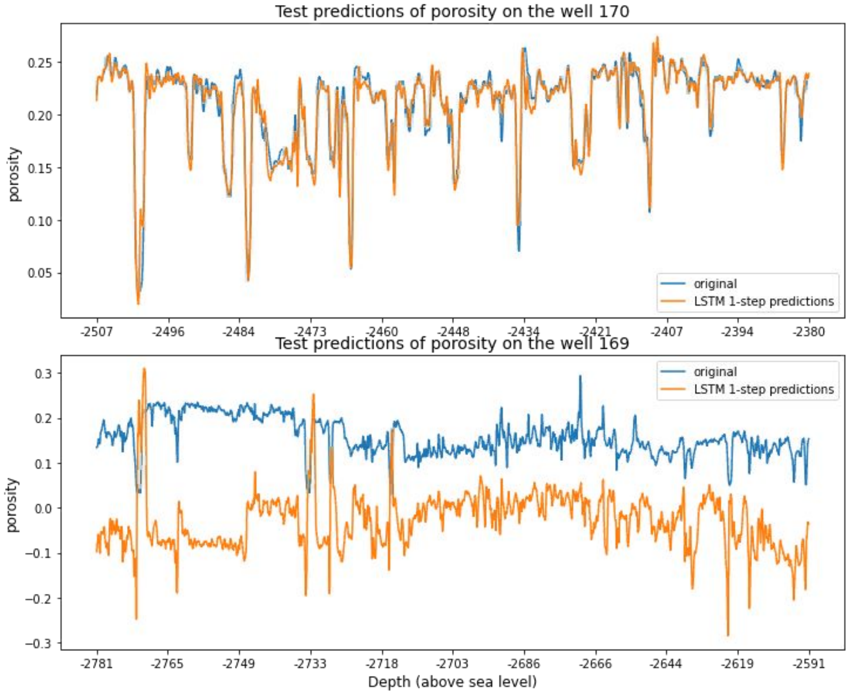
$$MAE = \frac{\sum_{i=1}^{N} |\phi_i^p - \phi_i^t|^2}{N},$$

$$R^2 = 1 - \frac{\sigma_{y|x}^2}{\sigma_y^2},$$

XGBoost for test wells



RNN for test wells



XGBoost for test wells

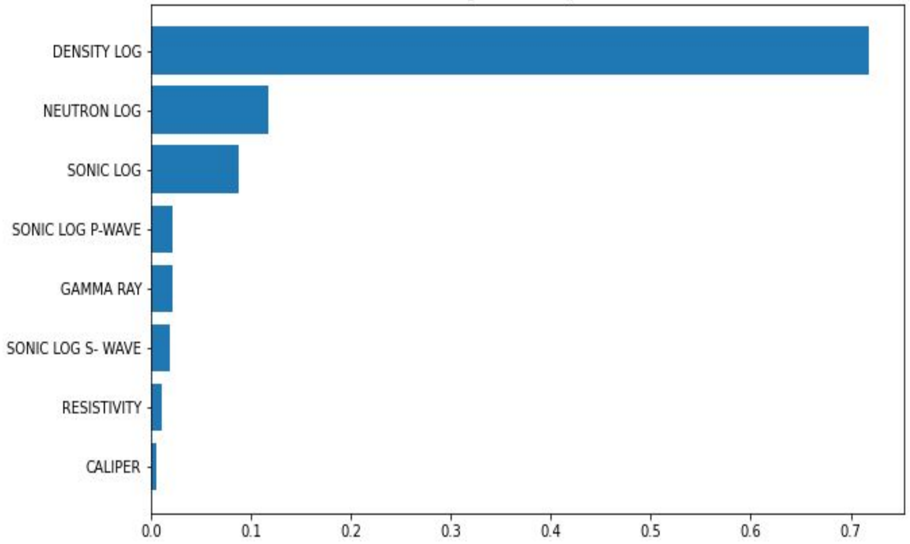
Proved itself as the algorithm with the best performance on other wells

WELL	RMSE	MAE	R^2
169	0.0211	0.01702	0.7393
171	0.02047	0.0173	0.3923
183	0.0290	0.0223	0.5510

RNN for test wells

WELL	RMSE	MAE	R^2
169	0.1308	0.1117	-9.0226
170	0.0092	0.0068	0.9454

Feature importance



The most important features **density, neutron and sonic logs**

Results without Feature Engineering

ML ALGORITHM	RMSE	MAE	R^2
LINEAR REGRESSION	0.0122	0.0102	0.9024
K-NN	0.0183	0.0128	0.7795
RANDOM FOREST	0.0113	0.0079	0.9152
XGBOOST	0.0094	0.0061	0.9413
FNN	0.0207	0.0156	0.7162
RNN	0.0092	0.0068	0.9454

Results with Feature Engineering

ML ALGORITHM	RMSE	MAE	R^2
LINEAR REGRESSION	0.0127	0.0100	0.8933
K-NN	0.0183	0.0128	0.7796
RANDOM FOREST	0.0103	0.0063	0.9303
XGBOOST	0.0089	0.0061	0.9477
FNN	0.0285	0.0255	0.4630
RNN	0.0098	0.0066	0.9372

New feature for regression models: the difference between values (1 m ago)
New features for time-series model: rolling mean and standard deviation (window size of 1 m), and the difference between values on rolling window's borders, lagged basic features (1 m), standard deviation of original series of depth intervals of 1 m

Conclusions

- using machine-learning algorithms is more straightforward comparing to classical approaches
- it does not need any extra procedures and equipment
- it is more robust compared to laboratory-based approaches
- it does not require upscaling procedure