EGU 2020

Machine Learning Technique for The Quantitative Evaluation of Tight Sandstone Reservoirs using High-Pressure Mercury-Injection Data

Jing-Jing Liu*, Jian-Chao Liu
School of Earth Science and Resources, Chang'an University, Xi'an, China
** Author for correspondence: liujingjing@chd.edu.cn*

Background

Concerns regarding tight sandstone reservoirs have become crucial issues in unconventional oil and gas resources. Because of the heterogeneity of tight sandstone reservoirs, the pore structures are complex and the lithology and electrical properties vary greatly. Achieving ideal results from the evaluation of tight sandstone reservoirs is difficult from the perspective of regional adaptability using the routine analysis methods.

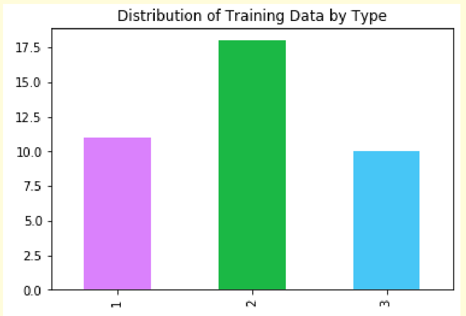
Problem Statement

The microscopic pore structure of the reservoir directly affects the reservoir's seepage capacity, controls oil and gas migration and reservoir distribution, and ultimately determines the distribution of reservoir productivity. This project focuses on prediction of reservoir quality of tight sandstone reservoirs using high-pressure mercury-injection data. We used two popular machine learning methods for the reservoir quality prediction, including K-means and random forest models.

Datasets

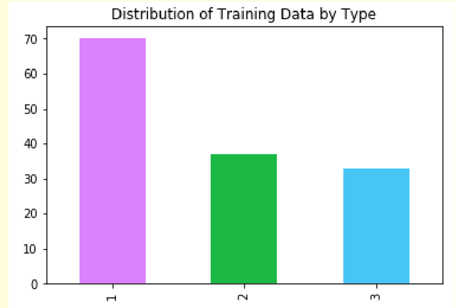
This project collected the high-pressure mercury intrusion data of eight wells, divided into two groups for analysis and prediction. The distributions of the data sets are as follows:

Group 1



	Por	Perm	R	SHgmax	WE	Pd	Type
count	39.000000	39.000000	39.000000	39.000000	39.000000	39.000000	39.000000
mean	10.369487	0.398462	0.486923	79.174615	26.602051	1.482821	1.974359
std	2.115910	0.494600	0.353361	11.386378	4.881262	1.310086	0.742938
min	5.700000	0.030000	0.030000	22.330000	17.670000	0.170000	1.000000
25%	9.105000	0.115000	0.210000	77.005000	23.050000	0.565000	1.000000
50%	10.710000	0.210000	0.420000	81.400000	26.920000	1.260000	2.000000
75%	11.530000	0.525000	0.720000	85.530000	29.885000	2.015000	2.500000
max	17.320000	2.370000	1.540000	90.810000	38.590000	6.200000	3.000000

Group 2



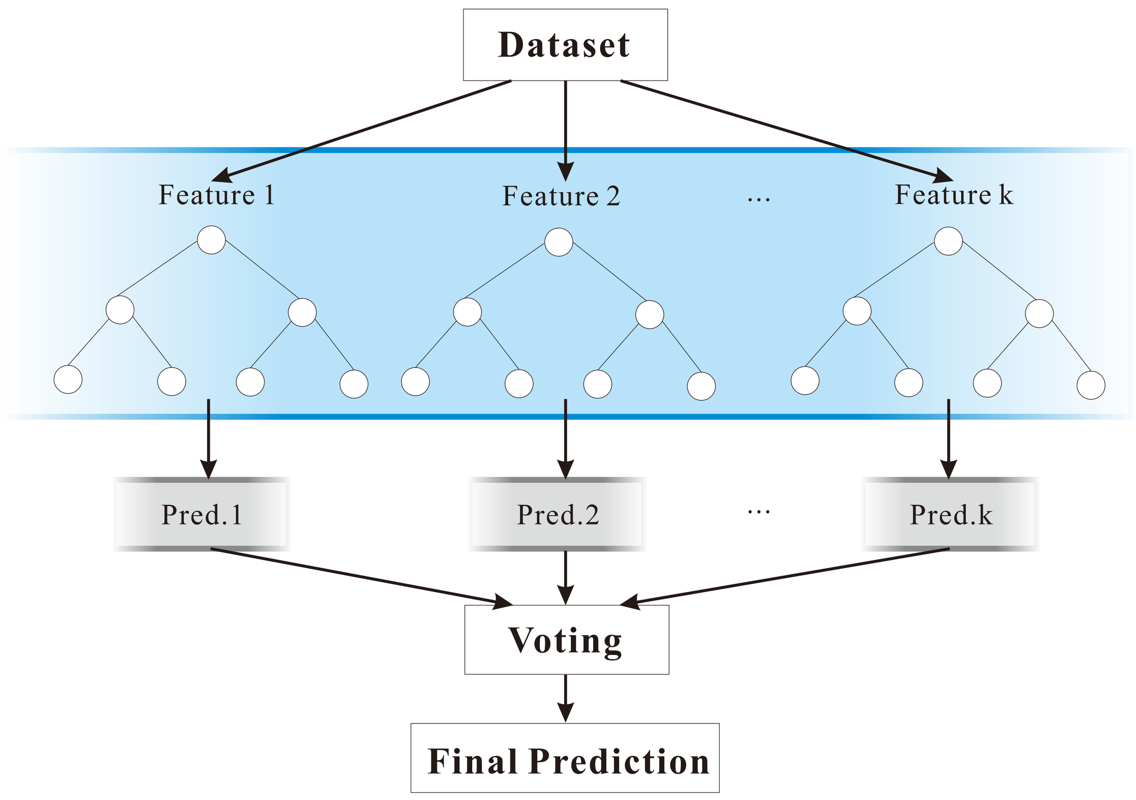
	Depth	Por	Perm	Pd	Pc50	R50	SHgmax	WE	Type
count	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000
mean	777.357857	9.711357	0.525236	1.335257	10.578221	0.116121	68.514571	35.407429	1.735714
std	26.811189	1.567950	0.504717	1.325759	8.744327	0.082982	9.024134	6.146080	0.818937
min	727.300000	1.800000	0.017000	0.103000	1.648000	0.015000	29.860000	15.560000	1.000000
25%	767.600000	8.800000	0.178000	0.517750	4.932750	0.059250	64.227500	33.260000	1.000000
50%	772.050000	9.750000	0.361000	0.928500	7.660000	0.098000	69.525000	35.805000	1.500000
75%	780.725000	10.700000	0.803000	1.569500	12.555500	0.151250	73.945000	39.285000	2.000000
max	908.500000	13.670000	2.770000	7.528000	49.865000	0.455000	91.820000	46.280000	3.000000

Models-K-means

The K-means clustering method is a non-genetic clustering method based on an iterative algorithm that clusters samples into *k* sets. The number *k* of classes can be predetermined or determined during the clustering process. The method has a small calculation cost and fast processing speed, which can be applied to a large data group.

Models-Random Forest

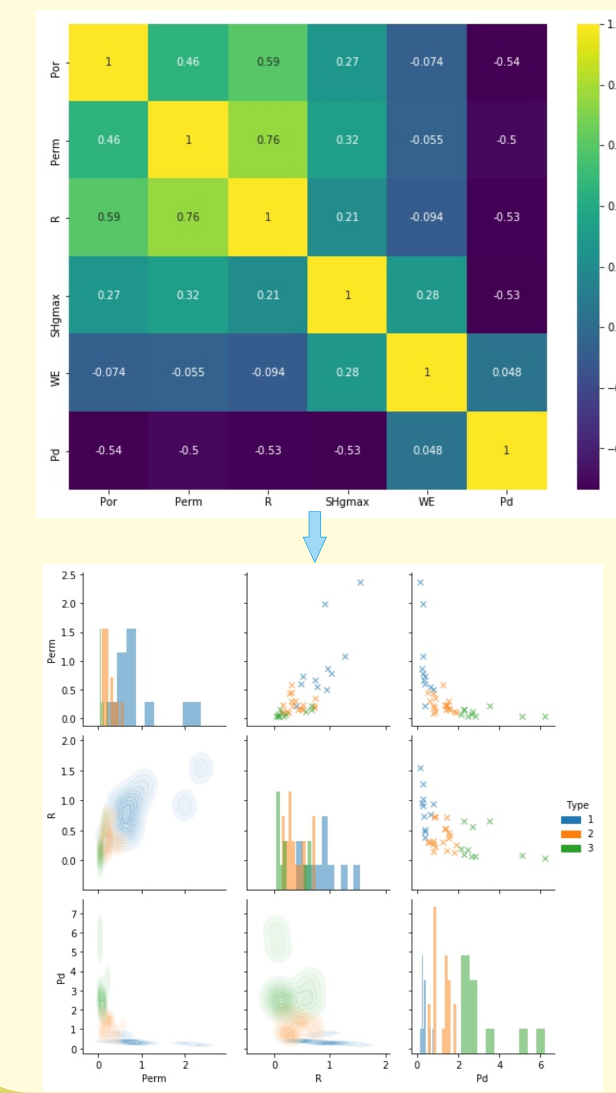
Random forest is an ensemble learning method composed of several decision trees. It contains multiple decision trees trained by Bagging, that is, Bootstrap Aggregating learning technology. When the samples to be classified are input, the final result is output by many decision trees by voting. This method solves the problem of performance bottlenecks of decision tree, and has good parallelism and scalability for high–dimensional data classification problems, and has good tolerance for noise and outliers.



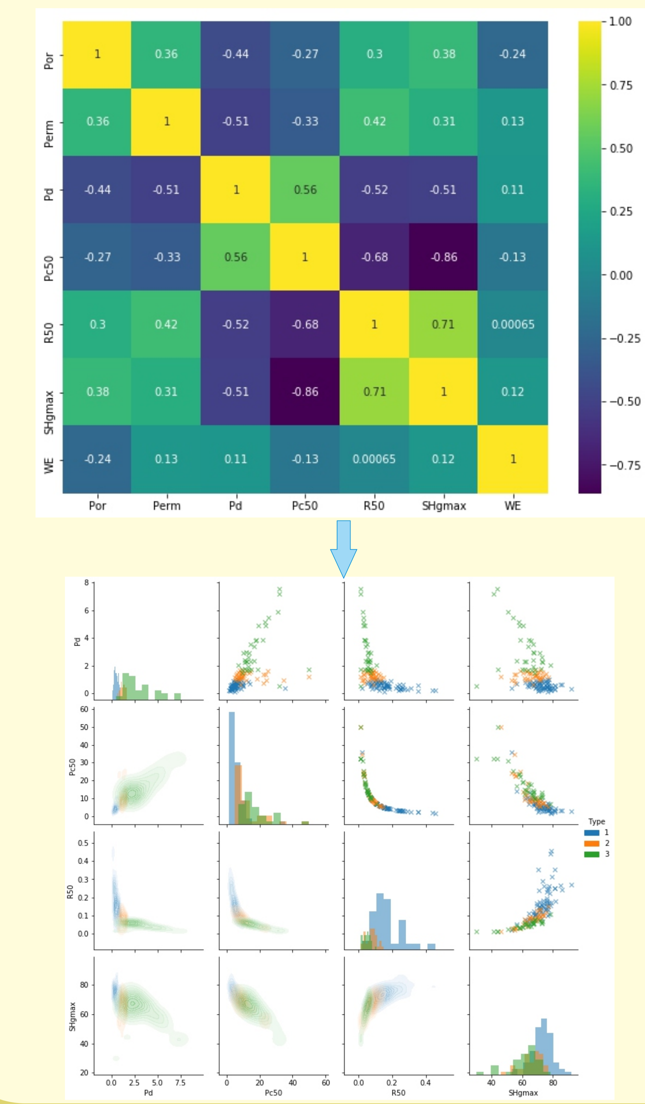
Feature selection

Heatmap is a good way to express the relationship between features, reflecting the importance and correlation of features. When the dimension of the data set is large, it will increase the calculation amount of the classifier and the running time. You can remove some of the insignificant features through the correlation analysis of the features to increase the effectiveness of the analysis task.

Group 1



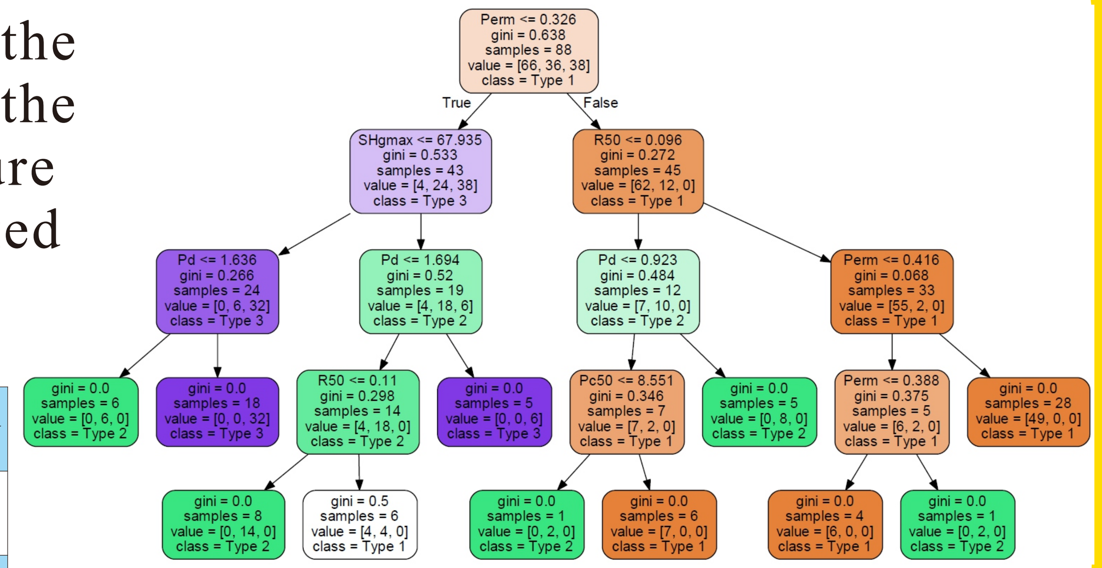
Group 2



Results

The prediction results of the two models are shown in the following table. The figure is a decision tree visualized from a random forest.

Model	F1 Score	Accuracy
K-means	0.53	0.60
Random Forest	0.75	0.89



Conclusions

1.Based on K-means clustering and random forest analysis, the tight sandstone reservoirs in the study area were divided into three types. The prediction model achieved relatively good accuracy.
2. The low computational complexity, low complexity, and low subjective factor interference ensure the quantification and accuracy of the evaluation results.It may apply to the prediction of other parameters in the petroleum field.