Original Article

KNN IN ACTION: MAPPING THE FUTURE OF OIL WITH PREDICTIVE CLASSIFICATION TECHNIQUES

[1]Jia Wei Sun and Li Min Zhou

[1]College of Energy, China University of Geosciences, Beijing, China and [2]The Eighth Oil Production Plant of China National Petroleum Changqing Oilfield Branch, Xi'an, China

Abstract: The development of a robust reservoir potential evaluation system is a pivotal task in the advanced stages of reservoir exploitation. Such a system aids researchers in the precise identification of remaining oil potential areas and the implementation of targeted development strategies. Several factors have been proposed by previous researchers for the establishment of an accurate evaluation system. In this context, Siqi Ouyang et al. have favorably suggested porosity, permeability, and contained saturation as primary evaluation indicators, derived from the division of flow unit parameters [1]. Yichao Zhang et al. have introduced a novel quantitative characterization method for remaining oil, which involves three dimensions: the remaining oil reservoir field, hydrodynamic field, and flow relationship field [2]. Lijie Liu et al. have devised a comprehensive index for the classification and evaluation of remaining oil, utilizing criteria such as geological reserve abundance, remaining oil geological reserve abundance, and movable remaining oil drive efficiency [3]. It is worth noting that the current research trend largely revolves around obtaining geological static evaluations from initial geological data. There is a notable absence of development dynamic indicators within the evaluation system.

Keywords: Reservoir Potential, Remaining Oil, Evaluation System, Geological Data, Development Dynamics

1. Introduction

The establishment of a potential areas evaluation system is an important research task of the later stages of reservoir development, which can help researchers to better identify the remaining oil potential areas. And implement precise development of the remaining oil. In order to establish an accurate evaluation system, various influencing factors have been proposed by previous authors. Siqi Ouyang et al preferentially selected porosity, permeability and contained saturation as evaluation indicators by dividing flow unit parameters[1]. Yichao Zhang et al proposed a new quantitative characterization method for remaining oil through three dimensions: remaining oil reservoir field, hydrodynamic field and flow relationship field[2]. Lijie Liu et al constructs a comprehensive index of remaining oil classification and evaluation by means of remaining oil geological reserve abundance,

Original Article

remaining oil geological reserve abundance and movable remaining oil drive efficiency[3]. It can be seen that researchers are currently more likely to obtain geological static evaluation from preliminary geological data, without introducing development dynamic indicators to build an evaluation system and lacking research on development dynamics.

In the study of algorithms for identifying reservoir potential areas based on evaluation indicators. There are unsupervised cluster analysis algorithms for identifying potential areas. For example, Lijie Liu et al used an improved FCM algorithm for unsupervised cluster analysis of regional remaining oil to identify potential areas[3]. Zhenpeng Wang et al applied a fuzzy C-mean clustering algorithm to identify reservoir potential areas[4]. Shuaiwei Ding et al applied a combination of fuzzy C-mean clustering algorithm and Bayesian discriminant function to identify deep-water reservoir potential areas[5]. It can be seen that these algorithms are unsupervised learning algorithms to classify the categories by evaluation indicators and applying professional knowledge to interpret the categories. There is a certain randomness in the classification. The interpretation of categories doesn't conform to the regularity of the oil reservoir engineering.

In summary, the author combines static and dynamic indicators to establish evaluation indexes and apply field reservoir engineering experience to establish a remaining oil potential areas system. Then uses K-nearest neighbour supervised learning algorithms to establish system, classifying and rating remaining oil potential areas. So that the classification results are more achieving convenient and fast identification of remaining oil potential areas.

2. Characteristics of fault block reservoir development

2.1 Geological development characteristics of the research area

The fault block reservoir has complex geological structure, many types, small fault blocks and many faults. The distribution of remaining oil is characterized by "general distribution and local enrichment" and the reservoir has stronger anisotropism. Especially in the late stage of water injection development. It is influenced by the planes and layers anisotropism. The remaining oil is distributed in different locations and difficult to be determined. Therefore, it is necessary to establish a method to identify the remaining oil potential area of the fault block reservoir, so as to achieve accurate and efficient exploration of the remaining oil in the block reservoir.

The lithology of the DX field is dominated by fine sandstone and siltstone. The field average porosity is 27.1% and average permeability is $498 \times 10^{-3}\mu\mu\mu\mu^2$. The highest permeability grade difference is 61.3 times. Because of long-term water injection, the field enters a period of the ultra-high water cut. The average water content is as high as 87%.

2.2 Remaining oil distribution characteristics

The oil-bearing reservoirs of the DX field can vertically be divided into two intervals, Es2-9-1 and Es2-9-4, which can be further subdivided into 18 sub-layers. It has many oil-bearing layers and long oilbearing well sections. The depth of the reservoirs ranges from 1800 to 2700 m. The reserves of the oilbearing sections account for 65.98% of the total reserves. The strong anisotropism of the fault block reservoir is mostly characterised by large differences in reservoir physical properties, fluid properties and geological layers. It results in different degrees of water drive in each area of the field. The distribution of remaining oil in the first plane of its layer is shown in Figure 1. From the diagram, it can be seen that most of the remaining oil is distributed in the edge or pinch area. Water-driven waves are more difficult to reach and the corresponding development potential is high. Therefore, the use of a single remaining oil saturation index to qualitatively find potential areas is one-sided. More comprehensive indexes and scientific methods are needed to identify remaining oil potential areas.

Original Article

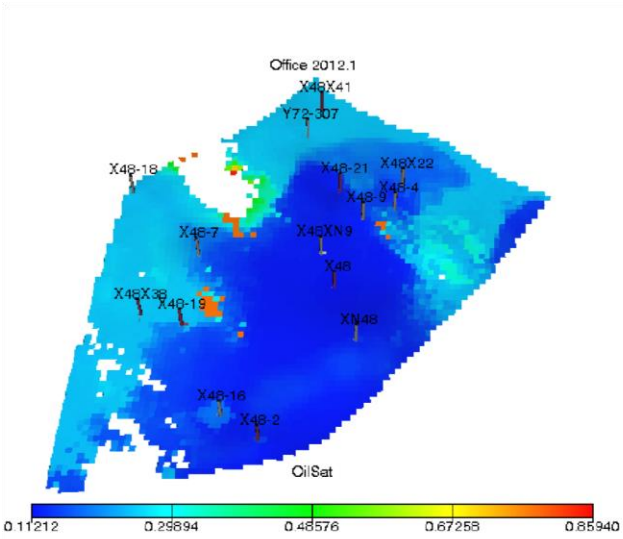

Figure 1: Distribution of oil-bearing saturation in the layer 1 of research area

3. Reservoir potential areas evaluation indicators

Through a combination of reservoir engineering knowledge and field experience, 3 static and 3 dynamic reservoir potential evaluation indicators from the perspective of reservoir development. Among them are flow field strength, residual oil moveable saturation and water content indicators as dynamic evaluation indicators, relative reservoir flow capacity indicators, reservoir permeability coefficient of variation and distance of potential zone from well and as static evaluation indicators.

3.1 Flow Field Strength Indicators

Under the production model of strong injection and production, the flow field distribution of the reservoir also changes with the change of physical properties. For each grid of fluid strength characterization parameters are inversely correlated with the remaining oil potential areas. Therefore, the flow field strength indicator is a quite effective indicator for potential areas evaluation[6]. Calculation of the flow field strength characterization parameters will use the two main parameters of the dynamic production parameters, the cumulative and the instantaneous categories. The overwater multiplier, fluid flow rate and water saturation, which have the greatest influence on the reservoir flow field. Which selected as the defining characterization parameters of the flow field strength. And then uses the hierarchical analysis method to determine the weights of different influencing factors. The expression of which is as follows.

$$R_w(i) = \left(\frac{K_{rw}\mu_0}{K_{ro}\mu_w}\frac{\int_0^{t_D} Q_{IN}dt}{V_i\emptyset_i}\right) \quad (1)$$

Where i is the grid number of the numerical simulation; $R(ii)$ is the repulsion multiplier of the grid i; KK_{rrww} is the relative permeability of the formation water; KK_{rrrr} is the relative permeability of the crude oil; $\mu\mu_{ww}$ is the viscosity of the formation water, $\mu\mu mmmm \cdot ss$; $\mu\mu_0$ is the viscosity of the crude oil, $\mu\mu mmmm \cdot ss$; QQ_{III} is the injection flow rate of the grid $_i$, $\mu\mu^3/ss$; VV_{ii} is the volume of the grid $_i$, $\mu\mu^3$; $\emptyset_{ii}$is the porosity of the grid $_i$.

$$v_i(i) = \left(\frac{K_{rw}\mu_0}{K_{ro}\mu_w}\frac{Q_{IN}^t}{S_i}\right) \quad (2)$$

Where $v(ii)$ is the fluid flow rate of the grid $_i$., $\mu\mu/ss$; $QQ_{III}{}^{dd}$ is the flow rate at moment t; SS_{ii} is the cross-sectional area of the grid $_i$ through which the fluid passes.

The ascending trapezoid method was used to determine the affiliation function.

Original Article

$$A(x) = \frac{x-a}{b-a} \quad (3)$$

Where a and b are the minimum and maximum values of a single indicator; x is the value of an indicator; and A(x) is the affiliation function.

$$WW_{ii} = RR_{dd} \cdot ww_{dd} = (FR_W(i), FQQ_{ll}(ii), FF_W(ii)) \cdot (0.5085,0.2934,0.1982)^{TT} \quad (4)$$

Where W_i is the flow field intensity of grid i; RR_{dd} is the affiliation matrix; WW_{dd} is the weight vector; $FR_W(ii)$ is the affiliation function of the overwater multiplier of the grid $_i$; $FQQ_{ll}(ii)$ is the affiliation function of the fluid velocity of the grid i; $FF_W(ii)$ is the saturation of water content of the grid i.

3.2 Remaining oil moveable saturation index

The remaining oil moveable saturation is a key evaluation factor for the remaining oil potential area in a certain extent. The remaining oil of some area that can be developed. It depends largely on the saturation of the remaining moveable oil. The expression of which is as follows.

$$SS_{rroo} = SS_{rrii} - SS_{rrrr} \quad (5)$$

WhereSS_{rroo} is the remaining oil movable saturation; SS_{rrii} is the original remaining oil saturation. $SS_{rrrr}$is the remaining oil saturation.

3.3 Water content indicators

The water content indicator of each oil recovery well reflects the water condition of oil well. According to the production status of the well, the oilfield will make corresponding adjustments. The expression of which is.

$$K_{rw}(S_w) = K_{ro}^{o}(\frac{S_w - S_{wir}}{1 - S_{wir} - S_{or}})^{\beta_w} \quad (6)$$

$$K_{ro}(S_o) = K_{rw}^{o}(\frac{1 - S_w - S_{or}}{1 - S_{wir} - S_{or}})^{\beta_o} \quad (7)$$

$$f_w = \frac{1}{1 + \frac{\mu_w K_o}{\mu_o K_w}} \quad (8)$$

Where $KK_{rrww}(SS_{ww})$ is the water phase relative permeability; $KK_{rrrr}(SS_{rr})$ is the oil phase relative permeability; K_{rw}^{o} is the endpoint water phase relative permeability; K_{ro}^{o} is the endpoint oil phase relative permeability; SS_{ww} is the water content saturation; SS_{wwiirr} is the bound water water content saturation; SS_{rrrr} is the remaining oil saturation; $\beta\beta_{ww}$, $\beta\beta_{rr}$ are the correlation coefficients; ff_{ww} is the water content.

3.4 Relative reservoir flow capacity index

Relative reservoir flow capacity affects the flowability of water-driven remaining oil. The relative reservoir flow capacity is a parameter that characterizes relative to the average flow capacity. The expression is as follows.

$$K_r h_r = \frac{K_i h_i}{\bar{K}\bar{h}} \quad (9)$$

Where $KK_{rr}h_{rr}$ is the relative reservoir flow capacity; $KK_{ii}h_{ii}$ is the reservoir flow capacity of the grid $_i$; and kkh is the average reservoir flow capacity.

3.5 Reservoir permeability coefficient of variation index

The variation coefficient of reservoir permeability can measure the strength of non-homogeneity in the formation longitudinal direction. The higher the permeability variation coefficient of the reservoir can be greater difference of the permeability within the layer. The expression of which is as follows.

$$v_k = \frac{1}{\bar{k}}\sqrt{(k_i - \bar{k})^2} \quad (10)$$

Where vv_{kk} is the reservoir permeability of variation coefficient; kk_{ii} is the permeability of grid $_i$, $10^{-3}\mu\mu\mu\mu^2$; k is the average permeability of a layer of the reservoir, $10^{-3}\mu\mu\mu\mu^2$.

3.6 Potential areas distance from wells

The distance of the potential area from the well is the straight line distance from the geometric centre of the remaining oil potential area to the injection and extraction well. This indicator parameter can measure the effect

Original Article

of the remaining oil development. The closer the location of the well will be smaller the remaining oil saturation. Its expression is.

$$(xx_{ii}, yy_{ii}) \in \Omega \quad i=1,2,..m;\ j = 1,2 \cdots m \tag{11}$$

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{12}$$

Where xx_{ii}, yy_{ii}, xx_{ii}, yy_{ii} are the x and y directional coordinates of the grid ii,, m; Ω is the reservoir boundary; dd_{iiii} is the distance of the potential areas from the well.

4. Classification of potential areas rating

The DX fault block reservoir is the research areas. A numerical simulation of the reservoir was constructed using eclipse software. The grid number of the model was 106 × 78 × 18 . The first, second, third, 10th, 15th and 18th layers are the main oil-bearing layers. Every layer has an effective grid of about 4,200. According to reservoir engineering and field experience, the research area is divided into four categories of potential areas. Remaining oil potential areas rating classification is shown in Table 1. The data are the minimum and maximum values of the indicator data parameters in the table.

Table 1: Remaining oil potential areas rating classification

Potential Areas Category		Flow field strength indicators	Remaining oil moveable saturation	Relative reservoir flow capability	Reservoir permeability coefficient of variation	Distance of the potential area from the well (m)	Moisture content
Category I	Minimum value	0.60	0.25	1	1	50	60%
Category I	Maximum value	0.70	0.30	20	15	60	75%
Category II	Minimum value	0.70	0.20	20	15	40	75%
	Maximum value	0.80	0.25	30	30	50	85%
Category III	Minimum value	0.80	0.15	30	30	30	85%
	Maximum value	0.90	0.20	40	45	40	95%
Category IV	Minimum value	0.90	0.10	40	45	20	95%
	Maximum value	1.00	0.15	50	60	10	100%

5. KNN Predictive Classification Model

5.1 Basic concepts of the algorithm

K-Nearest Neighbor algorithm is a supervised lazy machine learning classification algorithm. It has the characteristics of simple computing mechanism, fast model training time and applicable to class domain cross samples. It was first proposed by cover et al[7]. Later Bicego M proposed weighted nearest neighbour algorithm

Original Article

allows the KNN algorithm to be improved in terms of distance matching scores [8].The algorithm theory has been mature and widely used in various fields[9-11].

The algorithm calculates the spatial distance between the test samples and the training sample points through the classified training samples set. It finds the K nearest sample points of the cluster centre to the test data. The algorithm adopts the principle of majority rule to determine the category of the test data so as to classify the test set data into categories.

The KNN algorithm has the advantage of being very accurate in the relatively significant division of variability between samples. However, it also has shortcomings. When there are too many sample categories, the calculation of the distance between the data object and the cluster centre will also increase exponentially. Which is computationally heavy and runs slowly. The same number of votes may occur in the classification process and resulting in misclassification. Secondly, when the number of training sample categories is uneven. The discrimination is very likely to favor the sample type with a large number.

5.2 Mathematical model of the algorithm

The data of each indicator has different magnitudes and different magnitudes will have different effects on the results. Firstly, the evaluative indicators mentioned above are standardized and their standard score formula is as follows.

$$X_i^* = \frac{X_i - \bar{X}_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2}} \tag{13}$$

Chebyshev distance between data object and cluster centre formula.

$$\mathrm{d}(\mathrm{x}, C_i) = \lim_{k\to\infty}\left(\sum_{i=1}^{n}\left|x_j - c_{ij}\right|^k\right)^{1/k} \tag{14}$$

Where X_{ii} is the data object; CC_{ii} is the i the cluster centre; n is the number of sample data; K cosine similarity calculation formula (15).

$$\cos\theta = \frac{x_j \cdot c_{ij}}{\|x_j\| * \|c_{ij}\|} \tag{15}$$

Where θ is the angle between the two data objects to the origin.

K-nearest neighbor algorithm process.

(1) The Chebyshev distance formula is used to calculate the distance between the data objects in the training set and each test sample.

(2) Based on the calculated distances, the maximum distance among the current K nearest proximity samples is obtained.

(3) If the distance is less than the maximum distance of a certain K-cluster, the test sample is used as the K-nearest neighbor sample.

(4) Repeat steps 1, 2 and 3 until the test samples are all clustered.

6. Example application for identifying areas of dominant potential

6.1 The calculation of evaluation index parameters and results of predictive classification

Through the calculation of each evaluation index parameter formula, the evaluation indexes of the remaining potential area for DX oil field were obtained as shown in Table 2.

Table 2: Partial display data of evaluation index parameters for the DX oil field

Grid (ii, jj, zz)	Flow field strength indicators	Remaining oil moveable saturation	Relative reservoir flow capability	Reservoir permeability coefficient of variation	Distance of the potential area from the well (m)	Moisture content

Original Article

(28,29,1)	0.74	0.183	10.3	26.5	34.3	76.1%
(28,30,1)	0.77	0.165	13.5	28.7	36.2	78.2%
(28,31,1)	0.78	0.158	16.1	30.2	35.0	77.9%
……						
(28,46,1)	0.92	0.124	36.4	41.3	19.6	92.3%
(28,47,1)	0.90	0.115	37.5	43.7	17.2	91.5%
(28,48,1)	1.00	0.106	39.2	45.2	15.5	93.5%
……						
(94,32,1)	0.68	0.223	28.5	36.0	51.2	70.4%
(94,33,1)	0.70	0.236	30.2	38.2	52.3	71.3%
(94,34,1)	0.71	0.241	31.6	38.9	52.8	72.9%
……						
(14,52,1)	0.57	0.263	7.6	11.2	58.6	64.5%
(14,53,1)	0.59	0.281	8.1	10.6	59.1	64.8%
(14,54,1)	0.60	0.298	8.9	13.7	59.6	65.7%

The evaluation indicator for each grid were calculated. According to the classification rating criteria, they were classified. The principle of data category balancing was used to keep the data of the four categories balanced. The KNN predictive classification algorithm was applied to train 70% of the data to build a predictive model and 30% of the data to validate the classification results. The results of the predictive test set classification is obtained as shown in Figure 2 and confusion matrix is obtained as shown in Figure 3.

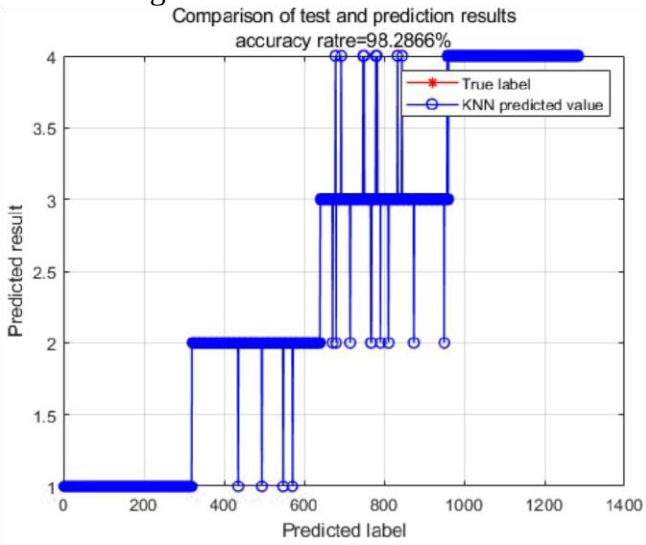

Figure 2: Graph of the results of the KNN predictive classification algorithm

Original Article

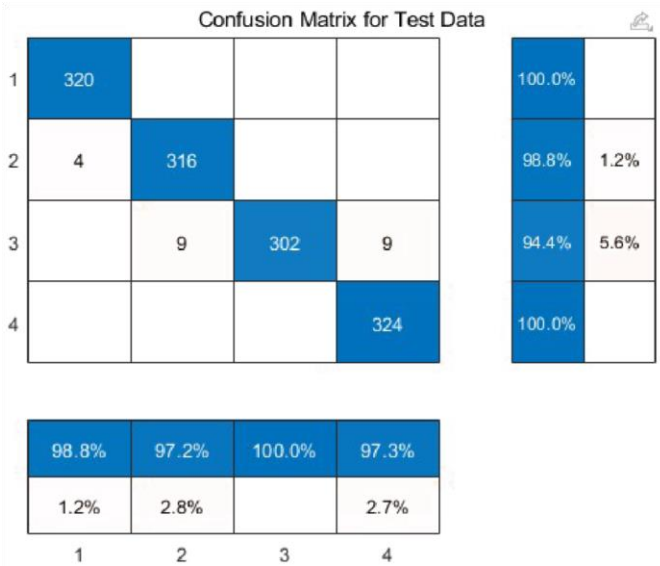

Figure 3: Confusion matrix for predicted data

The predictive classification statistics in Figures 2 and 3 show that the KNN predictive classification algorithm tested out with an accuracy of approximately 98.3%. Its algorithm has a high level of classification accuracy, as well as good recognition of the four classifications. For the incorrectly predicted categories, due to the calculated indicator parameters are too close to the criteria of two categories and incorrectly classified into other categories. The confusion matrix plot shows that there is more data incorrectly classified into the second and fourth categories. Because the algorithm adopts the principle of minority rule and will classify the sample into the category with more samples near it.

6.2 Suggestions for the exploitation adjustment scheme

The DX fault block reservoir is divided into four different rated potential areas by the KNN algorithm. We combined the field reservoir development technology and experience. According to the main problems of different potential areas, corresponding adjustment countermeasures are formulated.

For Class I potential areas, where the reservoir physical properties are poor but the remaining oil reserves are large. It is recommended that reservoir modification measures such as acid fracturing. For Class II potential areas, the reservoir is strong anisotropism in plan and longitudinal direction. There are also low permeability zones where it is difficult for water drive to reach. It is advised that adopt profile control and water shutoff to reduce the water drive ripple in the high permeability zones. For Class III potential areas, the reservoir has good physical properties, high water drive efficiency and medium reserves. But some marginal areas are less affected. It is recommended to adjust the working mechanism of injection and extraction wells, adjust the direction of fluid flow and change the original flow field to improve the ripple efficiency of the marginal areas. For Class IV potential areas, low remaining oil reserves and low development value in the high permeability area. it is recommended to develop in layers.

7. Conclusion

(1) In view of the low degree of water drive development and utilization in the late development of the fault block reservoir. A potential area evaluation system was established for using three dynamic and three static evaluation indicators. The reservoir evaluation system is divided into four major categories.

Original Article

(2) The KNN-based predictive classification algorithm was used to predict the test grid classification. And its prediction accuracy is 98.3%. The algorithm has a high predictive classification accuracy and can be used to guide the classification of potential areas in the research area.

(3) Based on the identification method of remaining oil potential areas, corresponding development adjustment measures are proposed for the rating four development potential areas in the DX oilfield. It provides guidance suggestions for water-flooding oil development technology in the fault block reservoirs.

References

Siqi Ouyang. Study on Flow Unit and Remaining Oil Distribution of Chang 81 Reservoir in Baibao Area, Ordos Basin[D]. Northwest University, 2019.

Yichao Zhang. Study on the adjustment strategy of the horizontal well development in Bohai BZ oilfield [D]. China University of petroleum, Beijing, 2020.

Lijie Liu. Classification and evaluation method of remaining oil in ultra-high water cut stage[J]. Petroleum Geology and Recovery Efficiency, 2022, 29(05):83-90.

Zhenpeng Wang. Development potential classification evaluation for water-flooding in conglomerate reservoir[J]. Lithologic reservoirs, 2018, 30(5):109-115.

Shuaiwei Ding, Hanqiao Jiang, et al. Classification and evaluation of deepwater oil reservoirs by combining clustering algorithm based on fuzzy C-mean with Bayesian discrimination function[J]. Journal of Xi'an Shiyou University(Natural Science Edition), 2014, 29(02):42-49.

Fuzhen Chen, Hanqiao Jiang, et al. A quantitative description of reservoir flow fields and its application[J]. Journal of Oil and Gas Technology, 2011, 33(12):110-115.

COVER T, HART P, et al. Nearest neighbor pattem classi6cation[J]. IEEE Transactions on Information Theory, 1967, 13(1):21-27.

Bicego M, Loog M. Weighted K-nearest neighbor revisited[C]. International Conference on Pattern Recognition(ICPR), 2016:1642-1647.

Cheng Huang, Wenjin Pan, et al. Research and application of oil multi-peak model based on machine learing[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2020, 42(6):7581.

Kui Sun. Logging Identification Method of Complex Lithology in Buried Hill Based on the Improved KNN Algorithm[J]. Special Oil & Gas Reservoirs, 2022, 29(03):18-27.

JIANG J H, CHEN Y J, MENG X Q, et al. A novel density peaks clustering algorithm based on k nearest neighbors for improving assignment process[J]. Physica A:Statistical Mechanics and its Applications, 2019, 523:702-713.