Application of Machine Learning Algorithm in Predicting Estimated Ultimate Reserves of Shale Gas

Shaolei Wei, Xuebin Huang, Yinghong Su, Jun Li, and Shuo Peng

Abstract Compared with conventional oil and gas resources, the development of shale gas is restricted by double sweet spot factors, namely geological factors and engineering factors. As a result, it is more difficult to predict the estimated ultimate reserves (EUR) for shale gas wells. Especially for new shale reservoir types or new layers from the exploration and evaluation stage to the early development stage, the uncertainty of EUR for shale gas becomes even more greater, and traditional single factor analyses cannot meet the prediction accuracy. This paper takes comprehensive factors (geological data and fracturing data) into consideration and applies machine learning methods to predict EUR for shale gas wells. Firstly, the integrity and validity of the collected data are analyzed and multiple interpolation algorithm methods are employed to generate the missing values; Then, single factor analyses are employed to qualitatively analyze the correlation of various geological and fracturing factors. Next, the XGBoost method is used to quantitatively characterize the correlation between single well EUR and various factors; Finally, three machine learning methods (XGBoost, RF, and SVR) are compared and selected to predict EUR for shale gas wells.

Keywords Shale gas · EUR · Main effecting factors · Machine learning

1   Introductions

After more than ten years of exploration and breakthrough, China's shale gas resources have successfully achieved commercial development, which means shale gas has become the main direction for increasing gas reserve and promoting gas production in the future. Compared with conventional oil and gas resources, the

S. Wei (✉) · X. Huang · Y. Su · J. Li
Petroleum Exploration and Production Research Institute, SINOPEC, Beijing, China
e-mail: leisurewin@163.com

S. Peng
MOE Key Laboratory of Petroleum Engineering, China University of Petroleum, Beijing 102249, China

development effect of shale gas is restricted by double sweet spot factors, namely geological factors and engineering factors. As a result, it is more difficult to predict the estimated recovery reserve (EUR) for shale gas wells. Especially for new reservoir types or new reservoir layers during the exploration stage, evaluation stage and early development stage, the uncertainty of single well EUR becomes even more greater, and the traditional single factor analysis method cannot meet the prediction accuracy. In recent years, with the development of big data technology and intelligent algorithms, machine learning algorithms have been widely applied in various industries, including field of oil and gas exploration and development.

Oil and gas exploration work involves multiple aspects of data such as geological data, geophysical data, logging, and production data. Machine learning methods can take all the data into consideration, which helps improve the accuracy of analysis results. Yu [1] elaborated on the application of machine learning algorithms and intelligent optimization theory in the fields of reservoir parameter prediction, reservoir fluid identification, lithofacies identification, fracturing fracture identification, automatic well location optimization, advantageous channel classification, and CO_2 oil displacement and storage, respectively. Liu [2] introduced unsupervised machine learning algorithms into the delineation of favorable zone boundaries and developed an intelligent delineation method for complex reservoir favorable zone boundaries. Firstly, the sensitive reservoir data and unsupervised clustering algorithm were used to cluster the reservoir distribution map, and then the favorable zone boundary range of the reservoir was given based on the criteria for delineation of favorable zones.

Compared with oil and gas exploration, machine learning methods are more widely used in the field of oil and gas field development. On the one hand, machine learning can predict development factors such as production capacity, yield, and single well EUR. On the other hand, based on the prediction of development factors, they can optimize the design of development plans and fracturing parameters. Ma et al. [3] proposed a machine learning process based prediction of shale gas productivity; Hirsch Miller et al. [4] used the recursive elimination method to screen the main control factors, and established a gas production prediction model with random forest to optimize engineering parameters; Sheikhi et al. [5] used linear regression, random forest, gradient elevator, tree regression, neural network and other methods to establish a shale gas prediction model, and evaluated engineering parameters from both local and global aspects; Shelley et al. [6] established a shale gas prediction model based on the BP neural network algorithm and used the established model to evaluate the quality of engineering parameters; Mao [7] elaborated on the current research status of artificial intelligence in oil and gas fracturing production, analyzed the key theoretical issues faced by the development of fracturing artificial intelligence, and looked forward to the main research direction and application scenario design of fracturing artificial intelligence.

This paper takes comprehensive factors (geological data and fracturing data) into consideration and uses machine learning method to predict single well EUR for shale gas. Firstly, analyze the integrity and validity of the collected data and use multiple interpolation algorithm methods to generate the missing values; Then, single factor analysis is employed to qualitatively analyze the correlation of various types of data.

Next, the XGBoost method is used to quantitatively characterize the correlation between single well EUR and various types of data; Finally, three machine learning methods (XGBoost, RF, and SVR) are selected to predict single well EUR.

2　Research Method and Realization Process

2.1　Introduction to Machine Learning

Machine learning essentially allows computers to find regulations from data and predict future trends based on the obtained patterns. The basic idea of machine learning is to imitate the process of human learning behaviors. For example, when we encounter new problems in reality, we usually summarize the rules through experience and predict the future process. The checkers program proposed by Turing and developed by Samuel in the 1950s marked the official entry of machine learning into its development period. In the 1980s, the concept of multi parameter linear programming (MLP) trained by neural network back propagation (BP) algorithm brought machine learning into a renaissance period. The decision tree (DT) proposed in the 1990s, along with the later support vector machine (SVM) algorithm, transformed machine learning from a knowledge driven approach to a data driven approach. At the beginning of the twenty-first century, Hinton proposed deep learning (DL), which led machine learning research from a downturn period to a flourishing period. Since 2012, with the improvement of computing power and the support of massive training samples, DL has become a hot research topic in machine learning and has driven widespread applications in the industry.

2.2　Research Process

In this paper, the machine learning algorithm is used to predict the EUR of a single well in JS block of F Shale Gas Field in China. On the basis of clarifying the geological characteristics and production characteristics of JS shale gas block, a series of geological parameters, engineering parameters and production data are put into comprehensive consideration. First, single factor analysis is employed to qualitatively describe the impact of different geological and engineering parameters on single well EUR; Then, the XGBoost method is used to quantify the importance of each parameter by recording the total number of feature splits and the total/average information gain, calculate the weight of different parameters, and screen the main control factors. Based on the selected main control factors, SVR, RF, and XGBoost algorithms are used to predict single well EUR, and the adaptability of the three methods is analyzed and compared.

Table 1 Geological parameters of JS shale gas block

Depth (m)	Pressure coefficient	Effective thickness (m)	Porosity (%)	Initial gas saturation (%)	TOC (%)	Initial temperature (°C)
2885	1.39	71	4.30	64.90	2.66	90
Adsorption gas content (m³/t)	Total gas content (m³/t)	Density (g/cm³)	Brittle mineral content (%)	Ro (%)	B_{gi}	Initial pressure (MPa)
2.40	5.20	2.56	56.50	2.59	1.012	40.22

3 Field Application in JS Shale Gas Block

3.1 Basic Condition and Data Collection

3.1.1 Introduction to JS Shale Gas Block

JS shale gas Block in China is a medium deep high-pressure shale gas reservoir with excellent reservoir physical properties, moderate burial depth and a high gas reservoir pressure. After fracturing, high open flow capacity can be obtained. The physical parameters of the JS shale gas block are as in Table 1.

Since the first shale gas well in JS shale gas Block was put into production in 2012, a total of 257 wells have been drilled in the three-year construction period of which 3 wells are in long-term shut-in condition due to fracturing failure. At present, there are 254 wells in production, with an average well test productivity of 37.9×10^4 m³/d. Since the production time of all shale gas wells in the study area has exceeded 7 years, the decline analysis method is used to predict the EUR of 254 wells, and the average EUR per well is 1.91×10^8 m³.

3.1.2 Data Collection and Diagnose

Many geological, engineering and production parameters of 254 wells in the study area were collected. The geological parameters include burial depth, TOC, porosity, gas content, free gas content, adsorbed gas content, gas saturation, pressure coefficient and brittle mineral content. The engineering parameters include the length of the horizontal well section, the number of fracturing sections, the length in the main layer, sand volume, and the fracturing fluid. Production parameters include open flow rate, cumulative flowback rate, and single well EUR for gas testing.

Based on the degree of data missing, quality evaluation standards for data is established to diagnose the validity and integrity of collected data. According to the standard, Python is used to create an automatic data quality control screening program to achieve batch control of data quality.

3.2 Analyses of Key Factors

The potential factors affecting the EUR of shale gas reservoirs include geological factors and engineering factors. Some factors have a great impact on the shale gas production while others have little impact. Qualitative analysis and quantitative analysis are carried out to evaluate the importance of collected factors.

3.2.1 Qualitative Analysis

The cross plot map between each factor and EUR is a good way the qualitatively show their relationship. The cross plot of single well EUR per kilometer and the above variables is drawn as follows, and the correlation is observed manually.

(1) Geological parameters

From the above cross-plots, it can be seen that reservoir burial depth, pressure coefficient, porosity and gas saturation have a greater impact on single well EUR. The burial depth affects compaction degree, crustal stress, ground temperature, etc. It can be seen from the analysis of the Fig. 1a that the burial depth is negatively related to single well EUR. The deeper the burial depth is, the stronger the shale reservoir is composed and the harder to maintain effective fractures. From Fig. 1b, we can see that formation pressure coefficient is positively correlated to single well EUR. The formation pressure coefficient not only reflects the energy of the gas reservoir, but also affects the hydrocarbon expulsion ability of the shale reservoir and the diffusion degree of shale gas. As can be seen in Fig. 1c, porosity is also an important factor. For conventional oil and gas reservoirs, porosity and permeability are two important characterization parameters of the reservoir to determine the quality of the reservoir. An important enrichment form of shale gas is that it exists in the pores and fractures of shale in free form. As a result, gas saturation has a big effect on EUR as is shown in Fig. 1d. The porosity and permeability of shale reservoirs also determines the amount of shale free gas. The correlations of gas content and TOC to single well EUR can be probed in Fig. 1e–f. Gas content is the most direct factor affecting shale gas enrichment. Shale gas content is determined by many factors, including enrichment and accumulation conditions, burial depth, TOC content, Ro, shale thickness, formation pressure, preservation conditions, etc. Gas content determines the enrichment of shale gas, and the correlation between free gas content and reserves is more obvious. The formation of shale gas reservoir requires source rock to have high total organic carbon content, moderate organic maturity, large thickness and wide distribution of effective source rock. TOC and recoverable reserves indicators are moderately correlated.

(2) Engineering parameters

The complexity of artificial fracturing is the key factors affecting the production of shale gas reservoirs, and horizontal strength, fracturing liquid volume, and sand
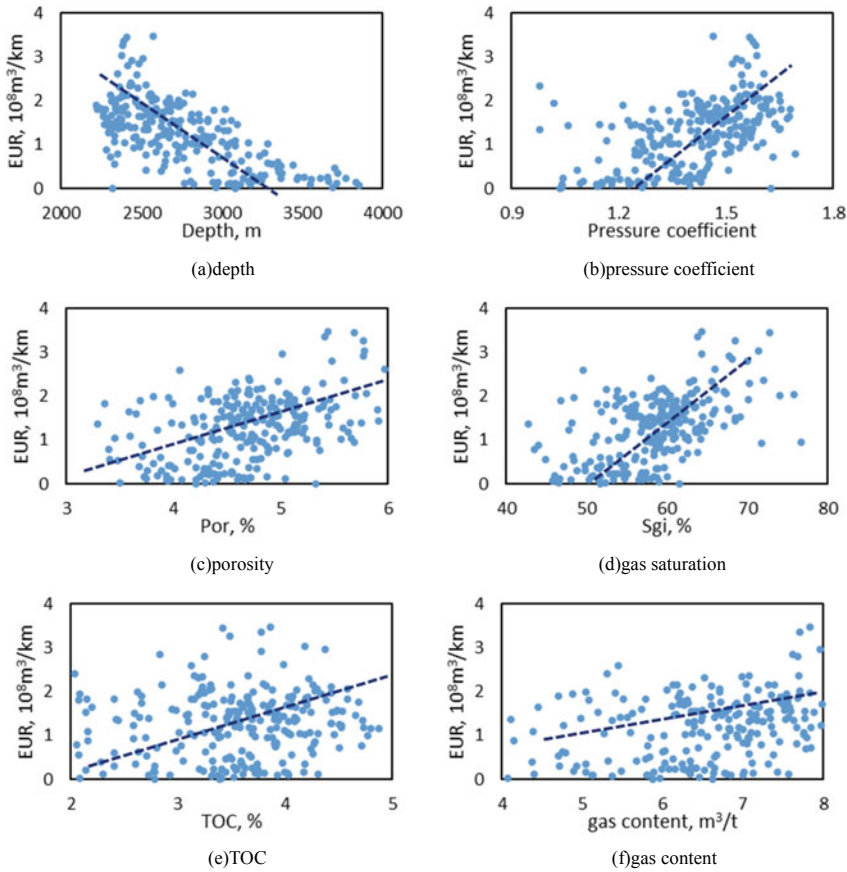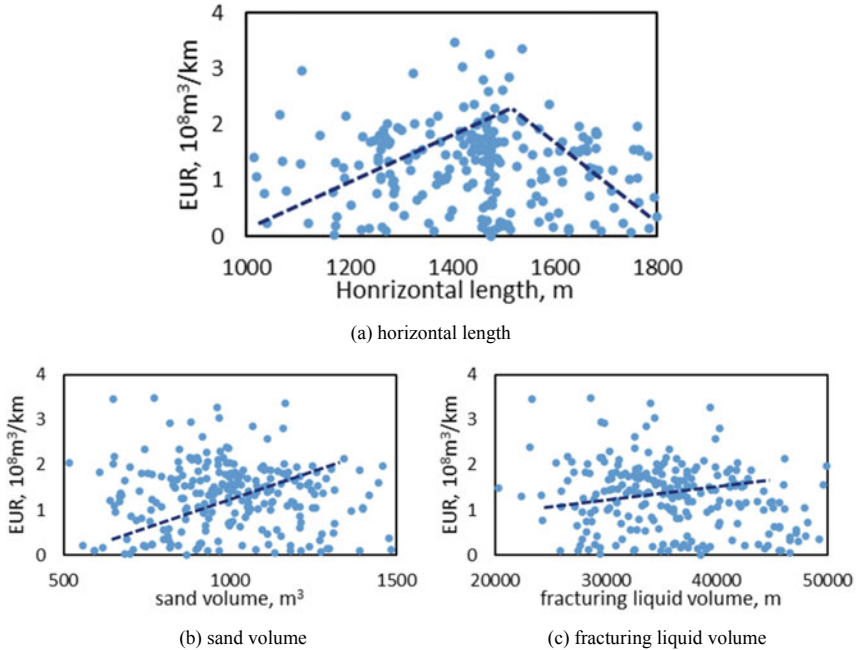
Fig. 1 Cross plot between EUR and different geological parameters

volume are the key parameters to describe the degree of fracturing. As is shown in Fig. 2a, within a certain range, there is a positive relationship between the length of the horizontal section and recoverable reserves. When the length of the horizontal section is greater than around 1500–1600 m, the extension of the horizontal section is not conducive to an increase in EUR per unit length. The amount of fracturing fluid and sand addition reflect the degree of effective fracture opening in the reservoir, and these two parameters show moderate correlation with the unit length EUR.

3.2.2 Quantitative Analysis

The traditional Pearson correlation coefficient method can only measure the linear correlation between two variables and all the variables must obey the assumption of normal distribution. The distance correlation coefficient overcomes the shortcomings

(a) horizontal length

(b) sand volume
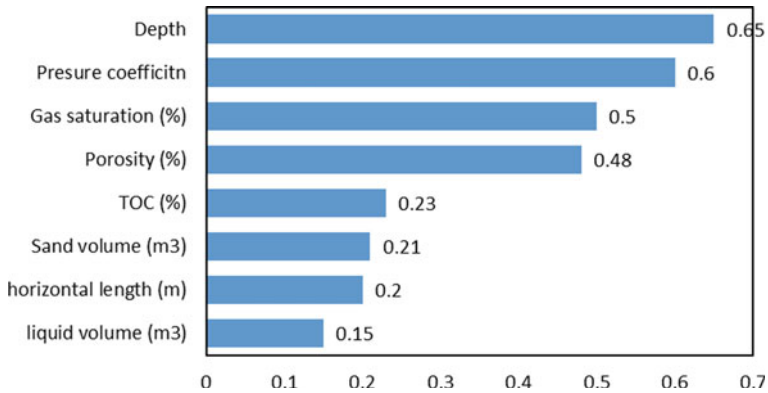
(c) fracturing liquid volume

Fig. 2 Cross plot between EUR and different engineering parameters

of Pearson correlation coefficient method because it measures the nonlinear relationship between two variables and is not limited by the assumption. The distance correlation coefficient studies the independence between two variables. A distance correlation coefficient of 0 indicates that the two variables are independent; The larger the distance correlation coefficient, the stronger the correlation between two variables.

Based on the distance correlation coefficient method, the overall analysis of the whole area shows that geological factors have a greater influence than engineering factors. The top three factors are burial depth, pressure coefficient and free gas content (Fig. 3).

XGBoost algorithm can directly output the importance of every factor, and a single decision tree model actually finds a suitable segmentation point for a certain feature during model establishment. During the training process, the feature importance is quantified by recording the total number of times that the feature is split and the total/ average information is gained. The weight calculation method uses the number of times that the feature is used as a partition attribute in all trees. That is to say the more features are used to construct a decision tree in a subtree model during splitting, the higher their importance.

Based on XGBoost algorithm, JS shale gas Block is evaluated as a whole, and the order of importance of influencing factors is analyzed, including burial depth,

Fig. 3 Ranking of Main Control Factors for EUR per kilometer using Pearson correlation method
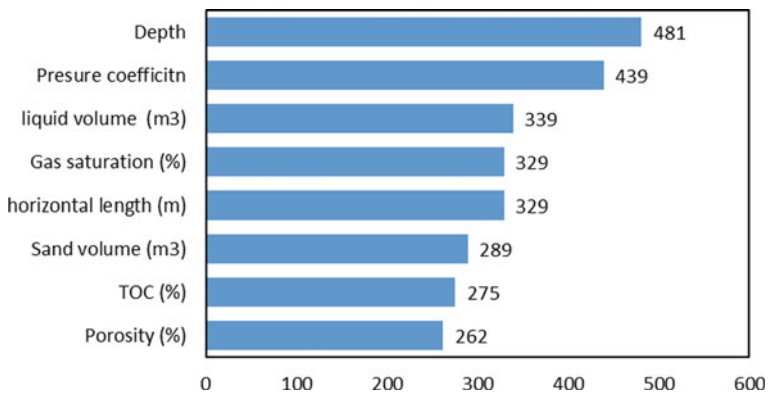
Fig. 4 Ranking of Main Control Factors for EUR per kilometer using XGBoost

pressure coefficient, fracturing fluid volume, free gas content, horizontal well section length, sand addition volume, TOC, porosity, and number of fracturing sections (Fig. 4).

3.3 EUR Prediction Using Machine Learning Methods

The results of three machine learning algorithms, namely SVR, Random Forest, and XGBoost, are compared for the ranking of the main control factors that can be extracted above. The basic component of XGBoost is the decision tree. We refer to these decision trees as weak learners, and with these weak learners combined, XGBoost model is formed. During the prediction process, trees are continuously added to fit the residuals of the previous prediction. After training, the corresponding

scores of each tree add up to the predicted values of the sample, which has the advantages of fast speed, good performance, and being able to process large-scale data. However, the algorithm has too many parameters, complex parameter tuning, and is not suitable for processing ultra-high dimensional feature data (Figs. 5, 6 and 7).

Comparison among predicting results of three algorithmic models show that XGBoost algorithm performs relatively well, with 35% of wells having a relative error of less than 20% in the test set. The total number of wells used in the test set
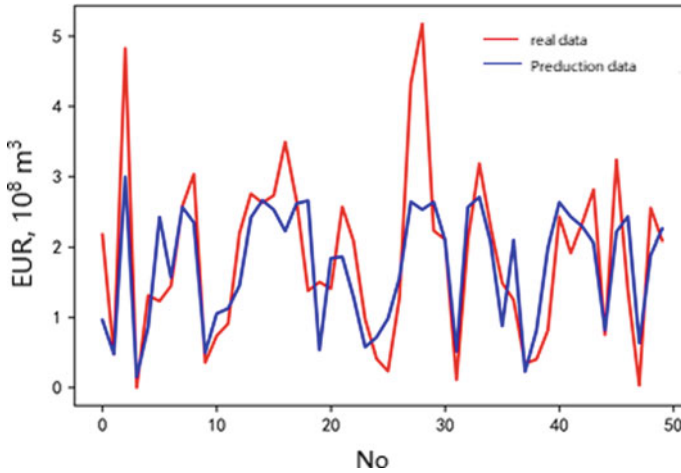

Fig. 5 Comparison between real production data and prediction data with SVR
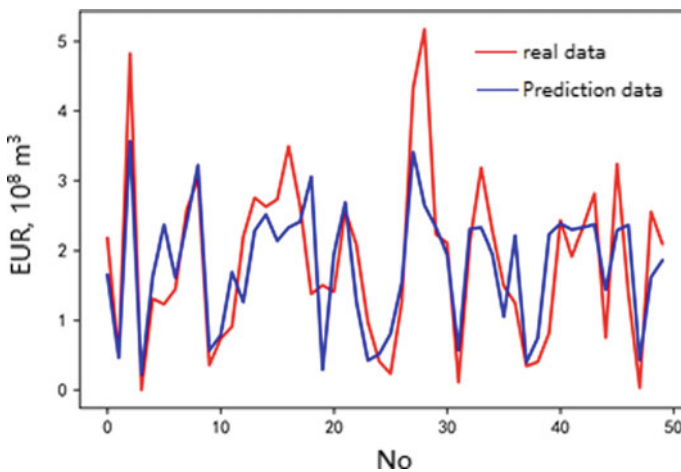

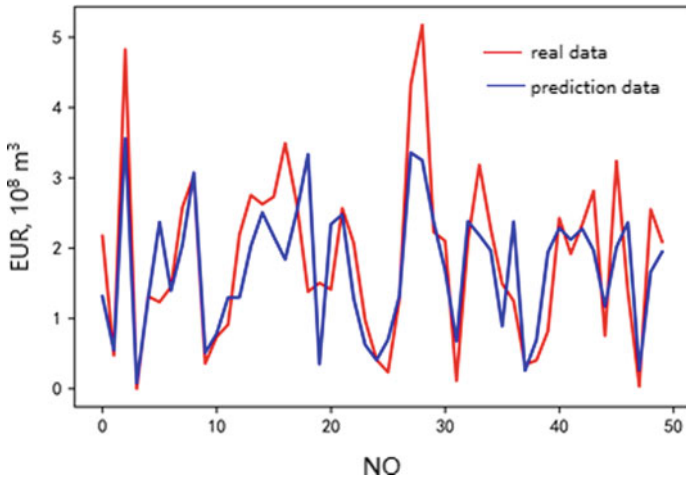**Fig. 6** Comparison between real data and prediction data with RR

Fig. 7 Comparison between real data and prediction data with XGBoost

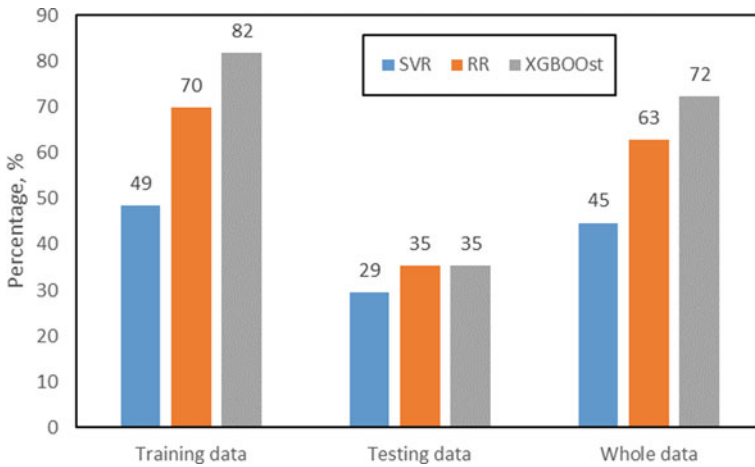is 50, and the number of wells with relative error within a certain range is shown in Fig. 8.

Fig. 8 Relative error comparisons between three methods

4 Conclusions

(1) Compared with conventional oil and gas, EUR prediction for shale gas has more uncertainty due to the comprehensive geological and engineering conditions. It is difficult to quantitatively characterize the correlations between EUR and various parameters simply by traditional methods. This article adopts a combination of traditional single factor analyses and intelligent methods. First, single factor analyses are used to qualitatively analyze the correlation between EUR and various parameters, and then XGBoost method is used to quantitatively characterize the impact of each factor on single well EUR.

(2) Through the combination of qualitative and quantitative analyses, the ranking of the main controlling factors for EUR is obtained. Controlling factors may be different for specific shale gas reserves. In our case, the burial depth of the gas reservoir, pressure coefficient, porosity and gas saturation are the main geological factors that have a key impact on EUR of shale gas wells; and the length of the horizontal section, the amount of sand and fracturing fluid are the main engineering factors that affect the EUR of shale gas single wells.

(3) Three machine learning algorithms (SVR, RF and XGBoost) are employed to predict the EUR of shale gas single wells and comparison is made among the three methods. It is shown that XGBoost method is the best prediction method for the data in this paper.

References

1. Hongyan, Y.U., Shuaiwei, D.I.N.G., Yanfang, G.A.O., et al.: Application of artifical intelligence in improving the effectiveness of oil and gas field exploration and development. J. Northwest Univ. (Nat. Sci. Edn.) **52**(06), 1086–1099 (2022)
2. Liu, E., Huo, S., Zhang, F.: Intelligent characterization method for favorable zone boundaries of complex reservoirs based on unsupervised machine learning (2022)
3. Wenli, M.A., Zhiping, L.I., Yuping, S.U.N., et al.: Non-deterministic shale gas productivity forecast based on machine learning. Spec. Oil Gas Reservoirs **26**(2), 101–105 (2019)
4. Hirschmiller, J., Biryukov, A., Groulx, B., et al.: The importance of integrating subsurface disciplines with machine learning when predicting and optimizing well performance–case study from the spirit river formation. In: Proceedings of the SPE Annual Technical Conference and Exhibition, F (2019)
5. Sheikhi Garjan, Y., Ghaneezabadi. M.: Machine learning interpretability application to optimize well completion in montney. In: Proceedings of the SPE Canada Unconventional Resources Conference, F (2020)
6. Shelley, R., Oduba, O., Melcher, H.: Machine learning and artificial intelligence provides wolfcamp completion design insight. In: Proceedings of the SPE Hydraulic Fracturing Technology Conference and Exhibition, F (2021)
7. Mao, S.M., Gensheng, L.I., Shouceng, T., et al.: Research status and prospect of artificial intelligence in reservoir fracturing stimulation. Drill. Prod. Technol. **45**(04), 1–8 (2022)