Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 6, No. 2, 2023

Production Capacity Prediction of Fractured Horizontal Wells Based on XGBoost

Yu Fu, Mingzhou Zhang

Petroleum Engineering School, Southwest Petroleum University, Chengdu, China

Abstract: Accurate prediction of the productivity of tight gas fracturing horizontal wells has important practical significance for optimizing development strategies and improving production efficiency. In view of the limitations of traditional methods in terms of assumptions and lack of historical data, as well as the complex nonlinear relationship between geological and engineering parameters, this paper takes the tight gas fracturing horizontal wells in Block S of Ordos Basin as an example, and uses Spearman correlation analysis to determine the main control factors of post fracturing production; Combining geological and engineering parameters, a horizontal well fracturing productivity prediction model is established based on XGBoost algorithm. The results show that the average error is 11.47%, which can better realize the accurate prediction of post fracturing productivity and provide an important scientific basis for the economic development of tight gas fields.

Keywords: XGBoost, Spearman coefficient, Fracturing horizontal wells, Production capacity prediction, Tight gas.

1. Introduction

The Sulige gas field in the Ordos Basin is the largest onshore gas field in China, and is a typical "four low" tight sandstone gas reservoir with low porosity, low pressure, low permeability, and low abundance. Fracking is an effective means to improve the production of a single well in tight sandstone gas reservoirs. Productivity prediction is a key scientific issue to realize the economic development of Tight gas reservoirs. Accurate prediction of the productivity of tight gas wells in horizontal wells after Fracking is an important prerequisite for determining reasonable development decisions, which is of great significance to the exploration and development process of tight gas fields in the Ordos Basin. At present, the productivity evaluation methods of low permeability Tight gas reservoirs at home and abroad mainly include two types: one is to derive productivity equations based on complex Formula, that is, analytical models, and the other is to predict productivity through numerical simulation using production performance data, that is, numerical models. The above methods are mainly based on theoretical models, requiring idealized assumptions and difficult to obtain parameters, and in the early stages of production testing, there is a lack of historical fitting data, making it impossible to apply theoretical models for yield prediction. At the same time, due to the comprehensive influence of geological and engineering parameters, there is a complex nonlinear relationship between geological parameters and fracturing engineering parameters and the production of Tight gas horizontal wells.

As a powerful Ensemble learning algorithm, XGBoost has obvious advantages in prediction ability and flexibility. Compared to simple models such as linear regression, XGBoost can better cope with complex datasets and feature interactions. XGBoost, based on the gradient lifting framework, trains multiple regression Tree model iteratively and optimizes the Loss function using gradient descent to provide more accurate prediction results. In this paper, firstly, Spearman correlation analysis is used to calculate the weight of geological and engineering factors that affect the productivity after tight gas pressure, and then XGBoost algorithm is used to directly start from geological parameters and engineering parameters through data mining technology, break through the limitations of traditional theoretical models, establish productivity prediction models for Tight gas horizontal wells in Sulige region, and improve productivity prediction efficiency and accuracy.

2. Data Collection and Preprocessing

2.1. Data sources

The original data collected includes 287 fracturing horizontal wells in Sulige Gas Field. The productivity impact parameters include six geological parameters, such as porosity, permeability, gas saturation, shale content, and six fracturing construction parameters, such as flowback rate, displacement, sand ratio, total fluid volume, total sand volume, and liquid nitrogen volume. The target parameter is the cumulative gas production in one year after fracturing.

2.2. Data preprocessing

The data collected in this article comes from the actual production of the Sulige gas field. Because the data records of different blocks are different, and there are missing values or Outlier in the actual production data, direct training is not possible. Therefore, Data cleansing and other operations must be carried out first to obtain higher prediction accuracy before predicting the post pressure production capacity through machine learning.

There are currently two main methods for handling missing values: directly deleting sample groups with missing values or filling in sample groups or features with missing values. Delete the features whose missing values account for more than half of the original data. Porosity, reservoir pressure, gas saturation and other characteristics are missing to varying degrees. The characteristics whose missing values account for more than half of the original data are deleted. For the blank values of other geological parameters, this paper uses the average value of the corresponding characteristics to fill.

If there are Outlier in the data set, the prediction accuracy of the model will be affected. In this paper, Outlier detection method based on box and line chart is used to determine

Outlier. The Laida Code. Assuming equal precision measurement of variables, if the residual error of a measurement value satisfies, it is considered a bad value with a large error value and deleted.

After Outlier and missing value processing, 267 groups of available samples were obtained to establish the productivity data set after fracturing horizontal well pressure in Sulige tight gas field.

3. Identification of Factors Affecting Post Compression Production Capacity Based on Spearman Correlation Analysis

Perform Spearman correlation analysis on the identified q influencing factors, and Pearson correlation is used to evaluate the linear correlation strength between two continuous variables. The purpose is to eliminate parameters with high linear relationships between each other and reduce the dimensionality of the data. Let X and Y be the sample data, and the calculation formula is:

$$\rho = \frac{\sum_{n=1}^{N}\left(X_n - \bar{X}\right)\left(Y_n - \bar{Y}\right)}{\sqrt{\sum_{n=1}^{N}\left(X_n - \bar{X}\right)^2 \sum_{n=1}^{N}\left(Y_n - \bar{Y}\right)^2}} \tag{1}$$

After calculating the Spearman correlation coefficient matrix, the size of the correlation coefficients between each parameter can be intuitively seen.

Parameters	Coefficient	Parameters	Coefficient
Porosity	0.573	Permeability	0.453
Gas Saturation	0.820	Effective Thickness	0.137
Average Total Hydrocarbon	0.734	Reservoir Pressure	0.314
Return Ratio	0.597	Displacement	0.379
Sand Ratio	0.612	Total Fluid Volume	0.624
Total Sand Volume	0.476	Liquid Nitrogen Usage	0.127

4. Construction of A Post Press Production Capacity Prediction Model Based on XGBoost

4.1. Basic Principles

XGBoost regression is a powerful regression algorithm based on gradient lifting framework. It predicts target variables by iteratively training multiple regression Tree model. Each regression tree learns the importance and interaction of features by fitting the nonlinear relationship of data, and optimizes the Loss function through gradient descent. In each iteration, a new regression tree is added to the model to minimize the gradient of the Loss function. To prevent overfitting, XGBoost uses regularization terms to control the complexity of the model. By combining the prediction results of multiple regression trees, a more accurate regression prediction was ultimately obtained. The advantage of XGBoost lies in its robustness to Outlier and missing data, as well as the accuracy evaluation of feature importance, making it a powerful tool for dealing with regression problems.

$$Obj_k = \sum_{i=1}^{M} l\left(y_i, \hat{y}_i\right) + \Omega\left(f_k\right) \tag{2}$$

4.2. Model evaluation indicators

In order to evaluate the effect of XGBoost model in predicting post pressure capacity, the average absolute percentage error, Root-mean-square deviation and correlation coefficient are selected as the criteria for evaluating model quality:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=0}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=0}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{4}$$

4.3. Model Establishment

4.3.1. Model hyperparameter setting

Model parameters are internal configuration variables within the model, and their values can be estimated using sample data. Model hyperparameters are external configurations of the model, and their values cannot be estimated from sample data. Therefore, the optimal values need to be determined through repeated experiments for the target problem. XGBoost has many hyperparameters that can control the scale of the model and characterize its complexity. This article adopts a cross validation method to determine the model complexity corresponding to the data size. Using cross validation, the Loss function is used as the evaluation index to select the optimal super parameter value. The summary of the determined hyperparameter values is shown in Table 1.

Table 1. XGBoost algorithm hyperparameters

Parameters	Value
Number of iterations	80
Characteristic sampling ratio: before node branching	0.69
Characteristic sampling ratio: before tree construction	0.4
Learning Rate	0.6
Regularization coefficient	135
Maximum depth	7
Samples allowed on any node	4
Proportion of sample to be sampled	0.6

4.3.2. Model feature screening

After preprocessing the horizontal well fracturing dataset and optimizing the algorithm hyperparameters, XGBoost was used to model and predict the production capacity after horizontal well fracturing. Considering the impact of different

number of features on the quality of the model, the Coefficient of determination of 50 fold cross validation is used as the evaluation index to determine the optimal parameter of the number of input features of the model.

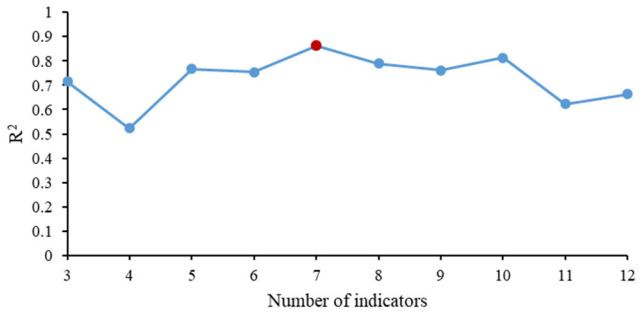

Figure 1. R2 value of 5-fold cross validation for different feature numbers

From Figure 1, it can be seen that as the number of features changes, the accuracy of XGBoost algorithm in predicting post compaction production capacity is constantly changing. When the number of features is 7, the model performs best.

4.4. Model Application

In order to further verify the applicability of the prediction model for volume fracturing production of horizontal wells, geological and engineering parameters of 10 horizontal wells in Sulige area are input into the software to carry out test production prediction, and the prediction results are compared with the field measured data, as shown in the figure.

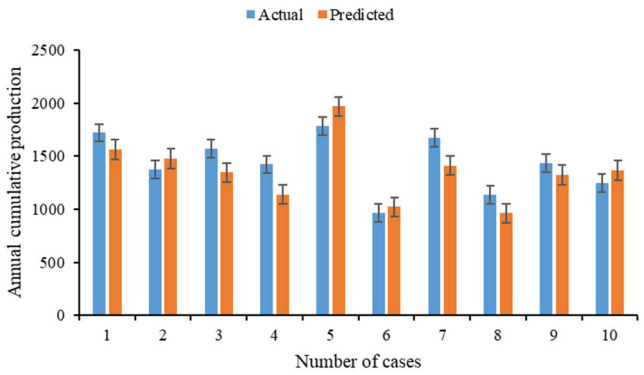

Figure 2. Comparison between predicted data and actual data

It can be seen from Table 2 that the average error of the trained production prediction model for tight gas fracturing horizontal wells is 11.54%, which indicates that the Tight gas production prediction model based on BPXGBoost can well express the internal law and relationship between the test production and various influencing factors, and the error rate of prediction accuracy is small, providing an efficient, feasible and accurate method for predicting tight fracturing production.

5. Conclusion

(1) The XGBoost model is selected. According to the factors affecting the production of Tight gas, the field actual data and the embedded Feature selection method are used to establish a productivity prediction model with 7 geological and engineering parameters, such as gas saturation, total sand volume, porosity, average total hydrocarbon, and effective thickness, as the input layer, and one-year cumulative gas production as the output layer.

(2) Taking the data of 267 actual fractured wells in Sulige area as training samples, the productivity prediction model of Tight gas horizontal wells is established using XGBoost algorithm with high accuracy. The model's generalization ability was validated using data from 10 actual wells, and the results showed an average error of 11.54%. The model has the characteristics of flexible operation and high prediction accuracy. This data mining based analysis method provides a new approach for the production capacity prediction of gas wells in the Sulige area, improving the efficiency of production capacity prediction.

References

[1] Li Xianwen, Wang Lili, Wang Wenxiong, et al Innovation of fracturing key technology and efficient development practice based on slim hole completion -- taking Tight gas reservoir of Sulige gas field as an example [J] Natural Gas Industry, 2022, 42 (09): 76-83.

[2] Fu Suotang, Fei Shixiang, Ye Zhen, etc Optimization of Horizontal Well Parameters in Tight Sandstone Gas Reservoirs [J] Natural Gas Industry, 2018,38 (4): 10.

[3] Anderson Roger N... ' Petroleum Analytics Learning Machine 'for optimizing the Internet of Things of today's digital oil field to refine petroleum system [C]//IEEE: IEEE, 2017.

[4] Temizel Cenk, Canbaz Celal Hakan, Palabiyik Yildiray, et al. A Comprehensive Review of Smart/Intelligent Oilfield Technologies and Applications in the Oil and Gas Industry [C]//SPE: SPE, 2019.

[5] Yu Hongyan, Ding Shuaiwei, Gao Yanfang, et al The application of artificial intelligence in improving the effectiveness of oil and gas field exploration and development [J] Journal of Northwest University (Natural Science Edition), 2022, 52 (6): 1086-1099.

[6] Liu He Prospects for the Application of Artificial Intelligence in Petroleum Exploration and Development [J] Journal of Intelligent Systems, 2021, 16 (6): 985.

[7] Li Yanzun, Bai Yuhu, Chen Guihua, et al New technology for Shale oil and gas production prediction based on artificial XGBoost method -- take Eagle Ford Shale oil gas field in the United States as an example [J] China Offshore Oil and Gas, 2020, 32 (4): 104-110.