Research Article

Predicting Heavy Oil Production by Hybrid Data-Driven Intelligent Models

Songhai Qin ^{1,2}, Jianyi Liu,¹ Xinping Yang,³ Yiyang Li,⁴ Lifeng Zhang,³ and Zhibin Liu¹

¹State Key Lab of Oil & Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, Sichuan 610500, China

²Development Department, Petro China Southwest Oil & Gasfield Company, CNPC, Chengdu, Sichuan 610051, China

³Research Institute of Exploration and Development, Xinjiang Oilfield Company, CNPC, Karamay, Xinjiang 834000, China

⁴Heavy Oil Development Company, Xinjiang Oilfield Company, CNPC, Karamay, Xinjiang 834000, China

Correspondence should be addressed to Songhai Qin; 505744700@qq.com

Received 28 January 2021; Revised 27 April 2021; Accepted 29 July 2021; Published 27 August 2021

Academic Editor: Kauko Leiviskä

Copyright © 2021 Songhai Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is difficult to determine the main control factors owing to the complex geological conditions of heavy oil reservoirs, including high viscosity, a wide range of variation of crude oil, and the great difference in production between different recovery methods. In this context, main control factors of heavy oil production in different recovery methods are analyzed and obtained based on the Apriori algorithm. The prediction of heavy oil production is faced with problems such as low prediction precision and insufficient data usage. Therefore, a novel intelligent simulation and prediction model of data-driven heavy oil production with time-varying characteristics is established based on differential simulation, machine learning, and intelligent optimization theory, which overcomes the defects of nonlinear, multifactor, and low fitting precision of dynamic data of heavy oil development. The parameters of the heavy oil production time-varying simulation model are identified by the least square support vector machine (LSSVM) to realize the intelligent prediction of the production. Numerical experiments show that the prediction result of the novel intelligent simulation and prediction model is better than the BP neural network model and the GM (1, N) model. This study provides a novel feasible method for data-driven heavy oil production prediction, and it can be helpful in further study of data-driven heavy oil production.

1. Introduction

Production forecasting is of great significance in heavy oil development planning. Because of the complex geological conditions of heavy oil, many oil recovery methods have been attempted, such as cyclic steam stimulation (CSS), steam flooding, steam-assisted gravity drainage (SAGD), in situ combustion, and toe-to-heel air injection (THAI). Many development and exploitation factors influence the prediction of heavy oil production; therefore, it is difficult to establish a heavy oil production prediction model with high accuracy and applicability.

The methods of oilfield development production prediction mainly include the decline curve method and mechanism model prediction method. The Arps declining

method is the earliest research on the analysis of production decline curves, and research on the decline curve is widely used in reservoir gas development dynamic prediction, which is the basis of reservoir dynamic prediction [1, 2]. Agarwal et al. [3] presented novel production decline curves for analyzing well production data from radial and vertically fractured oil and gas wells. Li et al. [4] presented a decline analysis model derived based on fluid flow mechanisms, which was proposed and used to analyze oil production data from naturally fractured reservoirs developed by water flooding. Ling and He [5] derived the governing equations of production decline for different reservoirs by combining static geological and reservoir data with dynamic production data. Jongkittinarukorn et al. [6] presented a new method to improve production forecasts and reserve estimation for a

multilayer well in the early stages of production, using the Arps hyperbolic decline method to model the decline rate of each layer. There are various recovery methods for heavy oil reservoirs, and many factors affect production. The traditional production decline method is mostly an empirical equation, which only considers the relationship between production and time, and has some limitations due to its harsh application conditions and small application scope. The model parameters of the mechanism model prediction method have clear physical significance and are an important means of studying the reservoir fluid flow mechanism. A complex numerical simulation model and historical fitting are required for the production prediction; subsequently, the new parameters are substituted into the modified model to predict the production [7]. However, the solution of the model takes a long time.

Recently, artificial intelligence (AI) technology has been gradually applied in petroleum exploration and development. Research has shown that the data-driven model performs better than the experience-based prediction model [8]. At present, the application of AI technology in the field of oil and gas field exploration and development includes the prediction of reservoir physical properties, oil and gas properties, recoverable reserves, optimization of well layout plans, optimization of hydraulic fracturing design, and production prediction. For example, Ahmadi et al. [9] compared the effectiveness of conventional models with the developed GA-LSSVM and GA-FL models. Al-Marhoun et al. [10] used AI technology to predict the crude oil viscosity curve in Canadian oil fields. Bhattacharya et al. [11] used a random forest and an artificial neural network to establish a supervised data-driven machine learning model and used the support vector machine (SVM) algorithm to understand oil well dynamics and predict daily gas production. Davtyan et al. [12] constructed an expected dynamic regression model based on the machine learning method of sliding window regression and obtained a stable crude oil production prediction model with long-term prediction ability. Most engineers do not have the time to analyze the geological conditions, production system, and production decline of most wells in detail. However, the application of AI plans can quickly and efficiently evaluate oil well performance and predict production [13] and can enable fast and accurate analysis of thousands of wells or even tens of thousands of wells. Compared with conventional reservoir engineering and numerical simulation methods, the data-driven AI model is simple, has strong generalization ability, and can accurately reflect the nonlinear response of the relationship between the production data [14]. AI has self-learning ability, and the production model established has intelligence and adaptability and can satisfy the needs of oil field production and development intelligence.

Machine learning is an important part of AI prediction methods. Machine learning algorithms include neural networks, decision-making tree, k -means, SVM, Apriori algorithm, expectation maximization algorithm, k -nearest neighbor algorithm, and naive Bayes algorithm. The Apriori algorithm [15], as a classical method of association rule

mining, is often used in factor analysis. This algorithm can find the correlation between data items in a large relational dataset and has been applied in many fields. The SVM has high fitting precision and a close relationship between the prediction index and the influencing factor; therefore, it is often used in the dynamic analysis of oil and gas reservoir development and production prediction. For example, Zhong et al. [16] used an SVM to predict the development index of ultrahigh water cut oil fields in eastern China. Elhaj et al. [17] combined a neural network with an SVM to predict the single well flow of a gas reservoir. Peng et al. [18] combined SVM with LSSVM and particle swarm optimization algorithm to predict oil and gas field production. According to the historical data of oil well production, Machado de Almeida Duque et al. [19] analyzed the prediction effect of six machine learning algorithms on crude oil production, and the verification set showed that the SVM and logistic regression models had the best prediction effect. However, the SVM is often used as a nonlinear fitting tool in crude oil production prediction to realize a univariate time sequence prediction or multiple regression prediction of production, without fully considering the variation trend of the oil field development index itself.

Grey theory was proposed by Deng [20] and has been widely applied in recent years [21–23]. Kumar and Jain [21] used the Grey–Markov and GM (1, 1) models with a rolling mechanism to predict the energy consumption in India. In [22], a prediction method using a grey model for cumulative plastic deformation under cyclic loads was demonstrated. Pao and Tsai [23] used the GM (1, 1) model to predict energy consumption in Brazil compared with the ARIMA model. The mainstream grey prediction models include the univariate grey model and multivariate grey prediction model, such as the GM (1, 1) model [20], discrete grey model DGM (1, 1) [24], fractional cumulative grey model FAGM (1, 1) [25], GM (1, N) model [26], and DGM (1, N) model [27]. Grey prediction pays more attention to the variation trend of the prediction index itself, which overcomes the defect that the SVM fails to pay attention to the variation trend of the prediction index itself. The parameters of the grey prediction model are important factors that affect the performance of the grey prediction model. Generally, the least squares method is used to estimate the parameters [28]. Based on the residual sum of squares minimum optimization, the least squares method easily falls into the local minimum. When it is applied to the prediction of heavy oil production with strong nonlinearity, the results obtained will deviate significantly. On the contrary, the least squares method has poor stability and cannot fit well in medium- and long-term production prediction, thus affecting the precision of the prediction model [29].

In this study, the Apriori algorithm is used to determine the main control factors of different heavy oil recovery methods, a time-varying multiple grey prediction model is established, and the LSSVM is used to identify model parameters for production prediction. The application of LSSVM to parameter identification of the grey model not only ensures historical high-precision fitting but also fully considers the variation trend of the state variable itself,

solves the problem of large deviation in the production prediction of the grey model, and realizes intelligent prediction of heavy oil production.

The remainder of this paper is organized as follows. In Section 2, the main control factors for heavy oil production are identified. In Section 3, an intelligent simulation and prediction model for data-driven heavy oil production is established. In Section 4, we conduct a numerical experiment and an analysis of the model application. Finally, the conclusions are drawn in Section 5.

2. Determination for Main Control Factors of Heavy Oil Production

Many factors affect the production of heavy oil. Under different recovery methods, these factors may play different roles, which make it difficult to determine the main control factors of production. In this study, the Apriori algorithm is introduced to conduct association analysis, and strong association rules related to crude oil production are mined. Then, the correlation degree between each factor and heavy oil production is ranked through the Pearson correlation coefficient, and the main control factors of heavy oil production are finally determined.

2.1. K-Means-Based Data Discretization. The numerical values of each column in the dynamic original dataset of heavy oil development are similar and have strong continuity and insufficient discrimination, which is not conducive to using the association rule algorithm to analyze the data. Before the Apriori algorithm was used to analyze the heavy oil production data, these data were discretized and converted into corresponding logical characteristic set data. In this study, the k -means clustering algorithm is used for realizing the discretization processing of the original dataset of oil field production.

Specific steps are as follows:

- (1) Select k centroids, $C_1, C_2, \dots$, from the dataset, and C_k is taken as the initial cluster centroid.
- (2) The Euclidean distance from each sample to the cluster center was calculated, and each sample was assigned to the most similar set according to the nearest distance principle. The average value of all objects in the cluster represents the cluster centroid. For each point V_j , the cluster centroid C_j is found. If the distance $d(V_j, C_j)$ between is minimum, V_j is assigned to the j th set.
- (3) In this way, all data samples are allocated to the corresponding set, and the initial centroid C_j of each cluster is recalculated using the above method.
- (4) Continue to follow Steps (2) and (3) circularly until the division of data no longer changes.
- (5) Obtain the minimum value $SEEC(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - C_k\|^2$.

Figure 1 shows the data discretization flowchart of the k -means-based clustering algorithm.

FIGURE 1: Data discretization flowchart of the k -means algorithm.

2.2. Influence Factor Analysis of Apriori Algorithm-Based Heavy Oil Production. The Apriori algorithm uses an iterative method known as layer-by-layer search to find frequent item sets related to crude oil production based on the given minimum support degree and then obtains strong association rules related to crude oil production based on the minimum confidence degree. Apriori association rule analysis has four basic definitions: frequent item set, association rule, support degree, and confidence. The frequent item set refers to the dataset that often appears, and the frequency is determined according to the support degree, while the support degree refers to the probability of the set appearing in the data. For example, in equation (1), X and Y represent the two individuals to be analyzed, and the support degree is defined as follows:

$$\text{Support}(X, Y) = P(XY) = \frac{\text{number}(XY)}{\text{num}(\text{AllSamples})}. \quad (1)$$

The association rule refers to the association between two individuals, which is measured by confidence, as shown in the following equation:

$$\text{Confidence}(X \rightarrow Y) = P(X|Y) = \frac{P(XY)}{P(X)}. \quad (2)$$

For a given rule $X \rightarrow Y$, the higher the confidence value, the more likely Y is to appear in a transaction involving X . Confidence can also estimate the conditional probability of Y under the condition of the given X , where the conditional probability is the condition of our association rule. The association rule that satisfies the minimum support threshold (min_sup) and minimum confidence threshold (min_conf) is called a strong rule. The two thresholds were between 0% and 100%. The task of association rule mining is

to determine the strong association rules related to crude oil production data.

Specific steps of the Apriori algorithm are as follows:

- (1) Find all 1-item sets first through the iterative updating method, then carry out the judgment according to the corresponding support degree, and eliminate the ones below the minimum support degree, and the rest are the frequent 1-item sets.
- (2) Put all frequent 1-item sets together to form 2-item sets, screen the 2-item sets according to the minimum support degree, and eliminate the ones with low support degree, and the rest are the actual frequent 2-item sets. If the iteration is carried out continuously, frequent $l+1$ -item sets can be obtained, and then, they are eliminated according to the minimum support degree to obtain the final production result of frequent l -item sets. Figure 2 shows the flowchart of the Apriori algorithm.

2.3. Sequence of Influence Factors of Heavy Oil Production. Through the analysis of the factors influencing the heavy oil production using the Apriori algorithm, the association rules of factors that may influence the heavy oil production were obtained. The degree ranking of these factors was conducted based on the Pearson correlation coefficient to determine the main control factors of heavy oil production. With each potential influencing factor influencing heavy oil production as an independent variable $x_1, x_2, \dots, x_p$ and heavy oil production as a dependent variable y , the correlation coefficient is calculated as follows:

$$r_{ij} = \frac{\sum_{a=1}^n (x_{ai} - \bar{x}_i)(x_{aj} - \bar{x}_j)}{\sqrt{\sum_{a=1}^n (x_{ai} - \bar{x}_i)^2 \sum_{a=1}^n (x_{aj} - \bar{x}_j)^2}}, \quad (3)$$

$$r_{iy} = \frac{\sum_{a=1}^n (x_{ai} - \bar{x}_i)(y_a - \bar{y})}{\sqrt{\sum_{a=1}^n (x_{ai} - \bar{x}_i)^2 \sum_{a=1}^n (y_a - \bar{y})^2}},$$

where $i = 1, 2, \dots, p$, $j = 1, 2, \dots, p$, and $\bar{x}_j = (1/n) \sum_{a=1}^n x_{aj}$.

The Pearson correlation coefficient is the ratio of the product of covariance and the standard deviation of two variables, and it is a dimensionless and standardized covariance. The linear variation does not affect the result of the Pearson correlation coefficient, so the unit change of the abscissa or ordinate will not change the value r , that is, the r values of different unit data are comparable. According to equation (3), the correlation coefficient can be calculated to judge the influence of independent variables on dependent variables, to sort the influencing factors and determine the main control factors of heavy oil production.

3. Intelligent Simulation and Prediction Model of Data-Driven Heavy Oil Production

Many factors affect heavy oil production, and the dynamic variation is different; therefore, it is difficult to predict the heavy oil production precisely. In the previous section, the authors analyzed and determined the main control factors

FIGURE 2: Flowchart of the Apriori algorithm.

for heavy oil production using the Apriori algorithm. Based on this, a heavy oil production simulation and prediction model from the perspective of a multifactor time-varying system is established in this section, considering the influence of the main control factors on heavy oil production. Meanwhile, the parameter identification of the model is improved, and a new intelligent simulation and prediction model for heavy oil production based on time-varying is obtained.

3.1. Establishment of Time-Varying Intelligent Simulation Model

3.1.1. Establishment of the Main Control Factor Dataset for Heavy Oil Production. The accurate selection of data reflecting valid information and the elimination of interference data are essential to the predictive ability of the data-driven model. Generally, abnormal data can be eliminated, but for the dynamic oil field development system, considering the oil field development index from the perspective of time sequence can better reflect the dynamic law of the whole oil field development, and the adoption of the elimination process will cause inconvenience to the subsequent modeling. Therefore, for the missing and abnormal data $x(i)$ in this study, the geometric mean of the previous and subsequent adjacent data, $x(a)$ and $x(b)$, is taken as the estimated value of $x(i)$ to repair the data:

$$x(i) = \sqrt{x(a) \cdot x(b)}. \quad (4)$$

Historical information on production and its main control factors is established. According to the main control factors of production corresponding to different heavy oil recovery methods, the heavy oil production is taken as the production of the system and the main control factors of production as the input of the system; the data table (Table 1) arranged in the chronological order is obtained.

The data on the factors influencing the heavy oil production are not consistent in dimension, and the quantity difference is large. If the dimensionless processing of these data is not conducted, the phenomena of “large numbers eat small numbers” easily occur, resulting in imprecise

TABLE 1: Heavy oil production data table.

Time	Production	Main control factors of heavy oil production				
		x_1	x_2	x_3	$\dots$	x_j
t_0	$Q(t_0)$	$x_1(t_0)$	$x_2(t_0)$	$x_3(t_0)$	$\dots$	$x_j(t_0)$
t_1	$Q(t_1)$	$x_1(t_1)$	$x_2(t_1)$	$x_3(t_1)$	$\dots$	$x_j(t_1)$
t_2	$Q(t_2)$	$x_1(t_2)$	$x_2(t_2)$	$x_3(t_2)$	$\dots$	$x_j(t_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_n	$Q(t_n)$	$x_1(t_n)$	$x_2(t_n)$	$x_3(t_n)$	$\dots$	$x_j(t_n)$

processing results. In this study, mean-value conversion was adopted for dimensionless processing:

$$x'(t_k) = \frac{x(t_k)}{\bar{x}}, \quad k = 1, 2, \dots, n. \quad (5)$$

To weaken the randomness of historical data, the dimensionless data should be accumulated and an association model is established.

Dimensionless processing is performed for the original data of the heavy oil production Q and the main control factor index $(x_1, x_2, \dots, x_j)$ in the historical data table according to equation (5) to construct a dimensionless time sequence for processing:

$$\begin{cases} \{Q(t_k)\}, \\ \{x_j(t_k)\}, \\ j = 1, 2, \dots, m, k = 1, 2, \dots, n. \end{cases} \quad (6)$$

The corresponding first-order accumulation sequence of dimensionless data in equation (6) is constructed and is denoted as follows:

$$\begin{cases} \{Q^{(1)}(t_k)\}, \\ \{x_j^{(1)}(t_k)\}, \\ j = 1, 2, \dots, m, k = 1, 2, \dots, n, \end{cases} \quad (7)$$

where $Q^{(1)}(t_k) = \sum_{i=1}^k Q(t_i)$ and $x_j^{(1)}(t_k) = \sum_{i=1}^k x_j(t_i)$.

Superscript 1 represents the first accumulation of dimensionless data.

Furthermore, the corresponding quadratic accumulation sequence was constructed using dimensionless equation (7), which is denoted as follows:

$$\begin{aligned} Q^{(2)}(t_{k+1}) &= AQ^{(2)}(t_k) + B_1 x_1^{(2)}(t_k) + B_2 x_2^{(2)}(t_k) + \dots + B_j x_j^{(2)}(t_k) \\ &\quad + \frac{1}{2}A(Q^{(2)}(t_k))^2 + \frac{1}{2}B_1(x_1^{(2)}(t_k))^2 + \frac{1}{2}B_2(x_2^{(2)}(t_k))^2 \\ &\quad + \dots + \frac{1}{2}B_j(x_j^{(2)}(t_k))^2 + \frac{1}{6}A(Q^{(2)}(t_k))^3 + \frac{1}{6}B_1(x_1^{(2)}(t_k))^3 \\ &\quad + \frac{1}{6}B_2(x_2^{(2)}(t_k))^3 + \dots + \frac{1}{6}B_j(x_j^{(2)}(t_k))^3 + \dots, \quad j = 1, 2, \dots, m. \end{aligned} \quad (10)$$

If equation (10) is used for extrapolating one more step, multistep prediction can be performed:

$$\begin{cases} \{Q^{(2)}(t_k)\}, \\ \{x_j^{(2)}(t_k)\}, \\ j = 1, 2, \dots, m, k = 1, 2, \dots, n, \end{cases} \quad (8)$$

where $Q^{(2)}(t_k) = \sum_{i=1}^k Q^{(1)}(t_i)$ and $x_j^{(2)}(t_k) = \sum_{i=1}^k x_j^{(1)}(t_i)$.

Superscript 2 represents the second accumulation of the original dimensionless data.

The heavy oil production data table (Table 1) is a dimensionless processing and secondary accumulation data table that can finally obtain heavy oil production and main control factors after dimensionless processing and secondary accumulation, which reduces the randomness of historical data.

3.1.2. Establishment of Model. According to the grey theory, the quadratic accumulation time sequence has the property of exponential change, that is, it can establish the association model of differential simulation heavy oil production:

$$\begin{cases} \frac{DQ^{(2)}(t)}{Dt} = AQ^{(2)}(t) + \sum_{j=1}^m B_j x_j^{(2)}(t), \\ Q^{(2)}(t_0) = Q(t_0), \\ x_j^{(2)}(t_0) = x_j(t_0), \\ j = 1, 2, \dots, m. \end{cases} \quad (9)$$

In equation (9),

$Q^{(2)}(t)$: the function of the change over time of heavy oil production after dimensionless processing and secondary accumulation

$x_j^{(2)}(t)$: the function of the change over time of the j th influence factor of the heavy oil production after dimensionless processing and secondary accumulation

Based on the historical data, the parameters A and $B_1, B_2, \dots, B_j$ are identified using the least squares method for equation (9) and discretized to obtain the following:

$$\begin{aligned}
Q^{(2)}(t_{k+2}) &= AQ^{(2)}(t_{k+1}) + B_1 x_1^{(2)}(t_{k+1}) \\
&\quad + B_2 x_2^{(2)}(t_{k+1}) + \cdots + B_j x_j^{(2)}(t_{k+1}) \quad (11) \\
&\quad + \cdots, \quad j = 1, 2, \dots, m.
\end{aligned}$$

When equation (11) is used for extrapolation and multistep prediction, the parameter is not updated with the latest information, resulting in inaccurate predictions. Therefore, when equation (9) is used for multistep prediction, it is a time-varying system, and association model equation (12) of heavy oil production based on a time-varying system can be obtained:

$$\begin{cases} \frac{DQ^{(2)}(t)}{Dt} = A(t)Q^{(2)}(t) + \sum_{j=1}^m B_j(t)x_j^{(2)}(t), \\ Q^{(2)}(t_0) = Q(t_0), \\ x_j^{(2)}(t_0) = x_j(t_0), \\ j = 1, 2, \dots, m. \end{cases} \quad (12)$$

In equation (12), the identification of parameters $A(t)$ and $B_1(t), B_2(t), \dots, B_j(t)$ is obtained based on repeated cycles of historical data.

3.2. LSSVM-Based Model Parameter Identification. When multistep prediction is performed for the time-varying intelligent simulation model, the latest information is used to identify the model parameters at each step. The model parameters changed with time, and the prediction precision was higher. In equation (12), the time-varying parameters $A(t)$ and $B_1(t), B_2(t), \dots, B_j(t)$ of the model are obtained based on repeated cycles of historical data. Although they can fully reflect the relationship between the prediction results and the main control factors and the variation trend of the state variable itself, the calculation deviation increases exponentially with an increase in the time step, resulting in the imprecision of the multistep simulation and prediction. The essence of this imprecision is the accumulated error of the least squares method for parameter identification of the time-varying system.

To overcome the disadvantage of using the traditional least squares method to estimate the model parameters and improve the precision of the time-varying intelligent simulation model in a multistep prediction, an LSSVM-based method is proposed to estimate the model parameters. Because of high-precision LSSVM fitting, it can avoid the rapid increase in accumulated errors caused by the identification parameters of the least square method. This method follows the principle of structural risk minimization [30], and the algorithm is fast and accurate.

Set $x^{(2)}(t_i)$ abbreviated as x_i and $Q^{(2)}(t_i)$ abbreviated as y_i , and the training sample is $\{(x_i, y_i), x_i \in R^n, y_i \in R\}_{i=1}^N$. Consider the nonlinear regression model:

$$y = f(x) + b, \quad (13)$$

where $f(x)$ is a nonlinear function. The nonlinear mapping $\varphi: R^n \rightarrow F$, where F is the high-dimensional feature space, is obtained as follows:

$$f(x) = w^T \varphi(x), \quad (14)$$

where $w^T = \{A(t), B_1(t), B_2(t), \dots, B_j(t)\}$.

By plugging in equation (13) into (14), the nonlinear regression model can be written as follows:

$$y = w^T \varphi(x) + b. \quad (15)$$

Defining an error variable:

$$e_i = y_i - w^T \varphi(x_i) - b. \quad (16)$$

Considering the optimization problems:

$$\min \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N e_i^2 \quad (17)$$

$$\text{s.t. } e_i = y_i - w^T \varphi(x_i) - b,$$

where C is the penalty factor.

To solve this problem in equation (17), the Lagrange function is constructed to transform the optimization problem into an unconstrained problem. Considering the Lagrange function below:

$$L := \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (w^T \varphi(x_i) + b - y_i + e_i), \quad (18)$$

where α_i is a Lagrange multiplier.

Considering the Karush–Kuhn–Tucker (KKT) condition below:

$$\frac{\partial L}{\partial w(t)} = w - \sum_{i=1}^N \alpha_i \varphi(x_i) = 0 \implies w = \sum_{i=1}^N \alpha_i \varphi(x_i), \quad (19)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^N \alpha_i = 0, \quad (20)$$

$$\frac{\partial L}{\partial e_i} = C e_i - \alpha_i = 0 \implies e_i = \frac{\alpha_i}{C}, \quad (21)$$

$$\frac{\partial L}{\partial \alpha_i} = w^T \varphi(x_i) + b + e_i - y_i = 0. \quad (22)$$

By plugging in equations (18) and (21) into equation (22) and eliminating w and e_i , the following system of linear

equations can be obtained by simultaneous equations (20) and (22):

$$\begin{pmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + \frac{I}{C} \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ Y \end{pmatrix}, \quad (23)$$

wherein $\vec{1} = [1, 1, \dots, 1]^T$, $Y = [y_1, y_2, \dots, y_n]^T$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$, and I are the n -order unit matrix. The elements in the matrix Ω are defined as

$$\begin{cases} \Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j), \\ i, j = 1, 2, \dots, n, \end{cases} \quad (24)$$

where $K(x_i, x_j)$ is a kernel function that satisfies Mercer's condition [31]. In this study, the radial basis function (RBF) was selected for the kernel function:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right). \quad (25)$$

By plugging in equations (19) and (24) into nonlinear regression model (15), the following can be obtained:

$$y(x) = \sum_{i=1}^N \alpha_i \varphi(x_i)^T \varphi(x) + b. \quad (26)$$

The values of Lagrange multipliers α_i and b can be solved directly using the system of linear equation (23), and the parameter values can be substituted into (26) to obtain the time-varying parameter discrete estimates in equation (12): $A(t)$ and $B_1(t), B_2(t), \dots, B_j(t)$.

3.3. Intelligent Prediction of Heavy Oil Production. The state equation can be obtained by performing discretization processing according to the association model in equation (12) for heavy oil production based on a time-varying system:

$$\begin{aligned} Q^{(2)}(t_{k+1}) &= A(t_k)Q^{(2)}(t_k) + B_1(t_k)x_1^{(2)}(t_k) + B_2(t_k)x_2^{(2)}(t_k) + \dots + B_j(t_k)x_j^{(2)}(t_k) \\ &+ \frac{1}{2}A(t_k)(Q^{(2)}(t_k))^2 + \frac{1}{2}B_1(t_k)(x_1^{(2)}(t_k))^2 + \frac{1}{2}B_2(t_k)(x_2^{(2)}(t_k))^2 \\ &+ \dots + \frac{1}{2}B_j(t_k)(x_j^{(2)}(t_k))^2 + \frac{1}{6}A(t_k)(Q^{(2)}(t_k))^3 + \frac{1}{6}B_1(t_k)(x_1^{(2)}(t_k))^3 \\ &+ \frac{1}{6}B_2(t_k)(x_2^{(2)}(t_k))^3 + \dots + \frac{1}{6}B_j(t_k)(x_j^{(2)}(t_k))^3 + \dots, \quad j = 1, 2, \dots, m. \end{aligned} \quad (27)$$

The LSSVM is used for the parameter identification of $A(t_k)$ and $B_1(t_k), B_2(t_k), \dots, B_j(t_k)$.

Finally, the quadratic reduction is performed for $Q^{(2)}(t_{k+1})$ to obtain $Q(t_{k+1})$, and the dimension is reduced to obtain the predicted value of heavy oil production. Therefore, according to equation (27), if the main control factors of production at a given predicted time (such as steam injection volume, oil well opening times, steam injection dryness, and steam injection pressure), the predicted production can be obtained.

4. Model Application and Result Analysis

An intelligent simulation prediction method for data-driven heavy oil production is applied to an oil field in China, and various examples of heavy oil production prediction have been obtained. In this study, the production prediction is researched using CSS and SAGD as examples.

4.1. Determination of Main Control Factors of Heavy Oil. The analysis data were derived from the production data of 200 CSS development units and 200 SAGD development units in a heavy oil reservoir in China. The Apriori algorithm

and Pearson correlation coefficient were used to determine the main control factors for shadow heavy oil production.

Before the Apriori algorithm was used to analyze the heavy oil production data, the data were preprocessed and converted into the corresponding logical characteristic set data. In this study, the k -means algorithm was used in Python, the K value was set to 4, and the continuous sample data were discretized into four levels for numbering. The discretization results of CSS and SAGD production and steam injection data are shown in the following figures (Figures 3–6).

According to the same method, the sample data of all possible influencing factors of steam stimulation and SAGD development are discretized into four levels for numbering, and the total labeling rules are shown in Tables 2 and 3.

Convert the CSS and SAGD production data into a logical data table (Tables 4 and 5).

Python programming was used to realize and obtain Apriori association rules, as shown in Tables 6 and 7.

The Pearson correlation coefficient between the influencing factors of heavy oil production and each influencing factor was calculated, and the thermodynamic chart of association was drawn (Figures 7 and 8).

FIGURE 3: Discretization results of production (CSS).

FIGURE 4: Discretization results of steam injection volume (CSS).

FIGURE 5: Discretization results of production (SAGD).

By comparing the relationship between the factors influencing the CSS and SAGD production of the heavy oil, the main control factors of CSS and SAGD production were obtained (Table 8).

4.2. Intelligent Simulation and Prediction of Heavy Oil Production. Three statistical criteria were used to appraise the predictive performance of the model: coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE).

R^2 is used to recognize the goodness of fit between the observed and modeled data values, which is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^m (x_{i,e} - x_{i,p})^2}{\sum_{i=1}^m (x_{i,p} - x_{i,a})^2}. \quad (28)$$

The RMSE is one of the most reported measures of disagreement and indicates the accuracy of predictions, which is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_{i,p} - x_{i,e})^2}. \quad (29)$$

MAPE is used to evaluate the overall forecast performance of the prediction models, which is defined as follows:

FIGURE 6: Discretization results of steam injection volume (SAGD).

TABLE 2: Original data conversion rules for CSS.

Influence factor	Symbol	Converted logical value
Steam injection volume	x_1	(1, 2, 3, 4)
Steam injection pressure	x_2	(5, 6, 7, 8)
Steam injection temperature	x_3	(9, 10, 11, 12)
Steam injection dryness	x_4	(13, 14, 15, 16)
Round of CSS	x_5	(17, 18, 19, 20)
Cycle of CSS	x_6	(21, 22, 23, 24)
Crude oil viscosity	x_7	(25, 26, 27, 28)
Soaking time	x_8	(29, 30, 31, 32)
Oil well opening times	x_9	(33, 34, 35, 36)
Oil saturation	x_{10}	(37, 38, 39, 40)
Permeability	x_{11}	(41, 42, 43, 44)
Porosity	x_{12}	(45, 46, 47, 48)
Reservoir temperature	x_{13}	(49, 50, 51, 52)
Reservoir pressure	x_{14}	(53, 54, 55, 56)
Reservoir thickness	x_{15}	(57, 58, 59, 60)
Production	Q	(101, 102, 103, 104)

TABLE 3: Original data conversion rules for SAGD.

Influence factor	Symbol	Converted logical value
Steam injection volume	x_1	(1, 2, 3, 4)
Steam injection pressure	x_2	(5, 6, 7, 8)
Steam injection temperature	x_3	(9, 10, 11, 12)
Steam injection dryness	x_4	(13, 14, 15, 16)
Interlayer	x_5	(17, 18, 19, 20)
Crude oil viscosity	x_6	(21, 22, 23, 24)
Oil well opening times	x_7	(25, 26, 27, 28)
Oil saturation	x_8	(29, 30, 31, 32)
Permeability	x_9	(33, 34, 35, 36)
Porosity	x_{10}	(37, 38, 39, 40)
Reservoir temperature	x_{11}	(41, 42, 43, 44)
Reservoir pressure	x_{12}	(45, 46, 47, 48)
Reservoir thickness	x_{13}	(49, 50, 51, 52)
Production	Q	(101, 102, 103, 104)

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{x_{i,e} - x_{i,p}}{x_{i,e}} \right| \times 100\%, \quad (30)$$

where $x_{i,e}$ is the actual value, $x_{i,p}$ is the predicted value, and $x_{i,a}$ is the average of the actual values.

The CSS and SAGD of the oil field were selected to develop continuous monitoring data for 24 months from 2018 to 2019, and the actual data information was obtained, as shown in Tables 9 and 10. Table 9 shows the relationship between the CSS production of heavy oil (Q) and the main control factor steam injection volume (x_1), oil well opening times (x_2), round of CSS (x_3), steam injection dryness (x_4), and steam injection pressure (x_5), and Table 10 shows the relationship between the SAGD production (Q) and the main control factor steam injection volume (x_1), oil well opening times (x_2), steam injection dryness (x_3), steam injection pressure (x_4), and steam injection temperature (x_5).

In this study, two traditional prediction methods, the BP neural network model (M1) [32] and the GM (1, N) model (M2) [33], were selected for comparison with the prediction effect of the data-driven intelligent simulation and prediction model (M3) for heavy oil production. The first 18 months of heavy oil production data were used for model parameter training, and the last six months of production data were used to evaluate the model's prediction effect. The three prediction models were used for predicting CSS and SAGD production, respectively, and the results are shown in Tables 11 and 12.

Table 11 shows that the MAPE values of the training and prediction data of M3 are 7.65, and 4.89, respectively. The RMSE of the M3 model is 10872, which is lower than the value of the other two models, reflecting the superiority of model M3; R^2 of model M3 is closer to 1, indicating that model M3 has a good prediction effect. Based on the analysis of the three evaluation indexes, model M3 had the best prediction effect on CSS production among the three prediction models.

Table 12 shows that the MAPE values of the training data and prediction data of M3 are 7.65, and 2.78, respectively. The RMSE of the M3 model is 4878, which is lower than the value of the other two models, reflecting the superiority of model M3; R^2 of model M3 is closer to 1, indicating that model M3 has a good prediction effect. Based on the analysis of the three evaluation indexes, model M3 had the best

TABLE 4: Logical data table of CSS.

Num	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	Q
1	2	1	1	2	2	2	3	4	2	2	1	1	2	4	2	3
2	2	1	1	1	2	2	3	4	2	2	1	1	2	4	2	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
200	4	1	1	2	2	2	3	4	2	2	1	1	2	4	2	3

TABLE 5: Logical data table of SAGD.

Num	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	Q
1	3	1	1	2	2	2	3	4	2	2	1	1	2	4
2	3	1	1	2	2	2	3	4	2	2	1	1	2	4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
200	1	1	1	2	2	3	3	4	2	2	2	1	2	4

TABLE 6: Production association rule results of CSS.

Influence factor	Association rule	Confidence
Steam injection volume	forzenset({"101"})-->forzenset({"1"})	0.728
Steam injection pressure	forzenset({"101"})-->forzenset({"5"})	0.512
Steam injection temperature	forzenset({"101"})-->forzenset({"10"})	0.382
Steam injection dryness	forzenset({"101"})-->forzenset({"13"})	0.583
Round of CSS	forzenset({"101"})-->forzenset({"20"})	0.628
Cycle of CSS	forzenset({"101"})-->forzenset({"22"})	0.457
Crude oil viscosity	forzenset({"101"})-->forzenset({"28"})	0.493
Soaking time	forzenset({"101"})-->forzenset({"29"})	0.378
Oil well opening times	forzenset({"101"})-->forzenset({"33"})	0.697
Oil saturation	forzenset({"101"})-->forzenset({"37"})	0.483
Permeability	forzenset({"101"})-->forzenset({"41"})	0.498
Porosity	forzenset({"101"})-->forzenset({"45"})	0.447
Reservoir temperature	forzenset({"101"})-->forzenset({"49"})	0.323
Reservoir pressure	forzenset({"101"})-->forzenset({"53"})	0.428
Reservoir thickness	forzenset({"101"})-->forzenset({"57"})	0.497

TABLE 7: Production association rule results of SAGD.

Influence factor	Association rule	Confidence
Steam injection volume	forzenset({"101"})-->forzenset({"1"})	0.748
Steam injection pressure	forzenset({"101"})-->forzenset({"5"})	0.522
Steam injection temperature	forzenset({"101"})-->forzenset({"9"})	0.587
Steam injection dryness	forzenset({"101"})-->forzenset({"13"})	0.593
Interlayer	forzenset({"101"})-->forzenset({"20"})	0.474
Crude oil viscosity	forzenset({"101"})-->forzenset({"24"})	0.498
Oil well opening times	forzenset({"101"})-->forzenset({"26"})	0.693
Oil saturation	forzenset({"101"})-->forzenset({"29"})	0.411
Permeability	forzenset({"101"})-->forzenset({"33"})	0.405
Porosity	forzenset({"101"})-->forzenset({"37"})	0.393
Reservoir temperature	forzenset({"101"})-->forzenset({"42"})	0.498
Reservoir pressure	forzenset({"101"})-->forzenset({"45"})	0.457
Reservoir thickness	forzenset({"101"})-->forzenset({"49"})	0.493

prediction effect on SAGD production among the three prediction models.

Figure 9 shows the comparison curves of production prediction results of models M1, M2, and M3 for CSS and

predicted values with the original data, and Figure 10 shows the comparison curves of production prediction results of models M1, M2, and M3 for SAGD with the original data. There are 24 sets of data for the test in total, in which the first

FIGURE 7: Thermodynamic chart of association of influence factors (CSS).

FIGURE 8: Thermodynamic chart of association of influence factors (SAGD).

18 sets of data were the model parameter training data, and the rest of 6 sets of data were used for testing the model prediction results.

As shown in Figures 9 and 10, the predictive ability of the intelligent simulation model (M3) of data-driven heavy oil production is better than that of the other two prediction

TABLE 8: Main control factors of heavy oil production.

Development mode	CSS	SAGD
Main control factor	Steam injection volume	Steam injection volume
	Oil well opening times	Oil well opening times
	Round of CSS	Steam injection dryness
	Steam injection dryness	Steam injection temperature
	Steam injection pressure	Steam injection pressure

TABLE 9: Production and influence factor data table (CSS).

$Q(t)$	x_1 (m ³)	x_2	x_3	x_4 (%)	x_5 (MPa)
194262	1847583	8352	8.29	65.8	4.87
327610	3198991	7964	8.31	67.2	5.23
275026	2654110	8838	8.35	65.3	5.26
288456	2764428	8962	8.47	62.5	5.28
255918	2454015	9108	8.52	63.2	5.35
293678	2761422	9150	8.57	50.1	5.36
203667	1887667	9254	8.69	60.8	5.42
178398	1655084	9193	8.71	61.7	5.34
167375	1548297	9010	8.73	56.5	4.88
212629	1976453	8925	8.79	78.3	6.65
224312	2114896	8655	8.80	79.1	6.31
171843	1654428	8254	8.83	78.4	6.11
159637	1682486	8109	8.84	78.7	6.12
200824	2135608	7795	8.52	78.6	6.23
219048	2336129	8422	8.57	78.5	6.55
209800	2263761	8511	8.65	78.2	6.83
225100	2392000	8716	8.75	78.1	6.92
215533	2235672	8838	8.83	81.6	6.76
213587	2005806	9420	8.92	81.2	6.69
206547	1918222	9449	8.95	81.4	6.85
185624	2246018	9474	9.01	80.9	6.53
245624	2249241	9525	9.13	80.9	6.42
223487	1735016	9537	9.19	80.6	6.58
174543	1784021	9578	9.24	80.8	6.32

TABLE 10: Production and influence factor data table (SAGD).

$Q(t)$	x_1 (m ³)	x_2	x_3 (%)	x_4 (MPa)	x_5 (°C)
64318	506112	115	57.9	4.83	275
71580	463157	139	66.2	4.13	281
77561	710444	145	72.1	3.75	284
81522	604951	136	75.2	4.95	298
78062	505637	189	52.7	5.08	264
63922	244004	148	62.7	5.98	276
67442	659807	132	60.4	5.58	158
70778	362603	158	55.8	4.08	159
76694	544993	129	54.1	5.16	169
75826	471763	178	49.7	3.58	171
75845	331719	186	56.0	4.12	140
70331	290967	192	76.7	4.05	146
79806	411983	193	77.0	6.14	138
86313	310045	198	77.4	4.61	141
86264	481391	208	77.5	4.33	145
82425	839562	206	77.3	4.16	135
74014	344731	209	78.9	4.52	134
88153	530710	211	81.6	3.75	205
91579	318079	236	80.7	4.95	196
93351	331737	223	82.7	1.24	195
79844	400203	256	82.9	3.85	195
94128	462791	242	82.8	6.03	189
95254	335283	466	84.2	3.98	194
96953	454539	278	84.2	5.22	202

TABLE 11: Production prediction results (CSS).

Original data	M1		M2		M3	
	Predicted value (t)	Error (%)	Predicted value (t)	Error (%)	Predicted value (t)	Error (%)
194262	198204	2.03	194262	0.00	194262	0.00
327610	321036	2.01	330581	0.91	315426	3.72
275026	279416	1.60	306316	11.38	318620	15.85
288456	289507	0.36	276199	4.25	278959	3.29
255918	259727	1.49	247188	3.41	239756	6.32
293678	286736	2.36	231969	21.01	231826	21.06
203667	203210	0.22	213548	4.85	207474	1.87
178398	181112	1.52	209309	17.33	213798	19.84
167375	171073	2.21	198825	18.79	193887	15.84
212629	197722	7.01	199468	6.19	203776	4.16
224312	209678	6.52	203945	9.08	220345	1.77
171843	162112	5.66	195732	13.90	194037	12.92
159637	162263	1.64	180199	12.88	165218	3.50
200824	203115	1.14	182559	9.10	181642	9.55
219048	228156	4.16	201112	8.19	219341	0.13
209800	218604	4.20	208572	0.59	224027	6.78
225100	231560	2.87	204722	9.05	207988	7.60
215533	219850	2.00	205807	4.51	207897	3.54
MAPE	2.72		8.63		7.65	
213587	206639	3.25	219031	2.55	218799	2.44
206547	197102	4.57	210478	1.90	209906	1.63
185624	230109	23.97	198837	7.12	198276	6.82
245624	230379	6.21	236368	3.77	235859	3.98
223487	180238	19.35	239546	7.19	239084	6.98
174543	186955	7.11	188129	7.78	187661	7.52
R^2	-0.32		0.77		0.79	
RMSE	26998		11164		10872	
MAPE	10.74		5.05		4.89	

TABLE 12: Production prediction results (SAGD).

Original data	M1		M2		M3	
	Predicted value (t)	Error (%)	Predicted value (t)	Error (%)	Predicted value (t)	Error (%)
64318	67922	5.60	64318	0.00	64318	0.00
71580	71973	0.55	68622	4.13	69053	3.53
77561	78971	1.82	70001	9.75	70356	9.29
81522	75817	7.00	83599	2.55	84596	3.77
78062	78370	0.39	81830	4.83	81929	4.95
63922	67270	5.24	85714	34.09	85637	33.97
67442	72317	7.23	74104	9.88	74103	9.88
70778	70610	0.24	71460	0.96	71468	0.97
76694	69529	9.34	67715	11.71	67844	11.54
75826	75224	0.79	67232	11.33	67541	10.93
75845	74076	2.33	73472	3.13	73623	2.93
70331	77015	9.50	72380	2.91	72326	2.84
79806	78614	1.49	76091	4.66	79368	0.55
86313	78173	9.43	82120	4.86	82689	4.20
86264	82683	4.15	78645	8.83	78695	8.77
82425	86226	4.61	84322	2.30	83697	1.54
74014	80542	8.82	90628	22.45	89553	20.99
88153	85575	2.92	81537	7.51	81958	7.03
MAPE	4.53		8.10		7.65	
91579	84737	7.47	89049	2.76	90457	1.23
93351	84272	9.73	91602	1.87	92325	1.10
79844	89769	12.43	82746	3.63	81241	1.75
94128	87789	6.73	96231	2.23	96534	2.56
95254	116091	21.88	97832	2.71	97958	2.84
96953	93405	3.66	102683	5.91	102584	5.81
R^2	-0.25		0.95		0.96	
RMSE	27293		5406		4878	
MAPE	16.32		3.50		2.78

FIGURE 9: Production simulation and prediction results of CSS.

FIGURE 10: Production simulation and prediction results of SAGD.

models. Model M3 has a higher prediction precision for heavy oil production and better adaptability in heavy oil reservoirs with different recovery methods.

5. Conclusions

Based on the characteristics of different heavy oil recovery methods, this study uses a hybrid data-driven method to determine the main controlling factors of heavy oil production and establishes an intelligent simulation prediction model for heavy oil production based on these main controlling factors. The anticipated effect using the data-driven intelligent simulation and prediction model to forecast heavy oil production is better than the BP neural network model and the GM(1, N) model under the same data conditions. The intelligent simulation and prediction model of data-driven heavy oil production can accurately predict the CSS of heavy oil and SAGD production, indicating that the model has good adaptability to heavy oil production prediction with different recovery methods, and the model can be used to predict heavy oil production.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by the Major Program of Sichuan Province (no. 20QYCX0030).

References

- [1] J. J. Arps, "Analysis of decline curves," *Transactions of the AIME*, vol. 160, no. 1, pp. 228–247, 1945.
- [2] M. J. Fetkovich, "Decline curve analysis using type curves," *Journal of Petroleum Technology*, vol. 32, no. 6, pp. 1065–1077, 1980.

- [3] R. G. Agarwal, D. C. Gardner, S. W. Kleinsteiber, and D. Fussell, "Analyzing well production data using combined type curve and decline curve analysis concepts," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, September 1998.
- [4] K. Li, U. Yangtze, and R. N. Horne, "A decline curve analysis model based on fluid flow mechanisms," in *Proceedings of the SPE Western Regional/AAPG Pacific Section Joint Meeting*, Society of Petroleum Engineers, Long Beach, CA, USA, May 2003.
- [5] K. Ling and J. He, "Theoretical bases of ARPS empirical decline curves," in *Proceedings of the Dhahi International Petroleum Conference and Exhibition*, Society of Petroleum Engineers, Abu Dhabi, UAE, November 2012.
- [6] K. Jongkittinarukorn, N. Last, F. Escobar, and K. Maneeintr, "A new decline curve analysis method for layered reservoirs," in *Proceedings of the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, March 2019.
- [7] A. Zerzar and Y. Bettam, "Interpretation of multiple hydraulically fractured horizontal wells in closed systems," in *Proceedings of the Canadian International Petroleum Conference*, pp. 1–14, Petroleum Society of Canada, Calgary, Alberta, June 2004.
- [8] I. Makhotin, D. Koroteev, and E. Burnaev, "Gradient boosting to boost the efficiency of hydraulic fracturing," *Journal of Petroleum Exploration & Production Technologies*, vol. 9, no. 3, pp. 1–7, 2019.
- [9] M.-A. Ahmadi, M. R. Ahmadi, S. M. Hosseini, and M. Ebadi, "Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: application of artificial intelligence," *Journal of Petroleum Science and Engineering*, vol. 123, pp. 183–200, 2014.
- [10] M. A. Al-Marhoun, S. Nizamuddin, A. A. A. Raheem, S. S. Ali, and A. A. Muhammadain, "Prediction of crude oil viscosity curve using artificial intelligence techniques," *Journal of Petroleum Science and Engineering*, vol. 86–87, pp. 111–117, 2012.
- [11] S. Bhattacharya, P. K. Ghahfarokhi, T. R. Carr, and S. Pantaleone, "Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: a case study from the Marcellus Shale, North America," *Journal of Petroleum Science and Engineering*, vol. 176, pp. 1–49, 2019.
- [12] A. Davtyan, A. Rodin, I. Muchnik, and A. Romashkin, "Oil production forecast models based on sliding window regression," *Journal of Petroleum Science and Engineering*, vol. 195, pp. 1–10, 2020.
- [13] Q. Cao, R. Banerjee, S. Gupta, J. Li, W. Zhou, and B. Jeyachandra, "Data driven production forecasting using machine learning," in *Proceedings of the SPE Argentina Exploration and Production of Unconventional Resources Symposium*, pp. 1–10, SPE-180984-MS, Buenos Aires, Argentina, June 2016.
- [14] K. Zhang, X. G. Zhao, L. M. Zhang et al., "Current status and prospect for the research and application of big data and intelligent optimization methods in oilfield development," *Journal of China University of Petroleum (Edition of Natural Science)*, vol. 44, no. 4, pp. 28–38, 2020.
- [15] X. Wu, V. Kumar, J. Ross Quinlan et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2007.
- [16] Y. Zhong, L. Zhao, Z. Liu, Y. Xu, and R. Li, "Using a support vector machine method to predict the development indices of very high water cut oilfields," *Petroleum Science*, vol. 7, no. 3, pp. 379–384, 2010.
- [17] M. A. Elhaj, F. Anifowose, and A. Abdulraheem, "Single gas flow prediction through chokes using artificial intelligence techniques," in *Proceedings of the SPE Saudi Arabia Section Annual Technical Symposium and Exhibition*, Society of Petroleum Engineers, Al-Khobar, Saudi Arabia, April 2015.
- [18] J. Peng, Y. Qiao, D. Tang, L. Ge, Q. Xia, and T. Chen, "The least squares SVM for the prediction of production in the field of oil and gas," *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 12, no. 1, pp. 60–74, 2018.
- [19] M. C. Machado de Almeida Duque, G. Souza Chaves, D. de Oliveira Monteiro, L. Velasco Medani, and V. J. Martins Ferreira Filho, "Machine learning models to automatically validate petroleum production tests," in *Proceedings of the SPE Latin American and Caribbean Petroleum Engineering Conference*, SPE-199112-MS, July 2020.
- [20] J. L. Deng, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.
- [21] U. Kumar and V. K. Jain, "Time series models (Grey-Markov, grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India," *Energy*, vol. 35, no. 4, pp. 1709–1716, 2010.
- [22] X.-W. Ren, Y.-Q. Tang, J. Li, and Q. Yang, "A prediction method using grey model for cumulative plastic deformation under cyclic loads," *Natural Hazards*, vol. 64, no. 1, pp. 441–457, 2012.
- [23] H.-T. Pao and C.-M. Tsai, "Modeling and forecasting the CO₂ emissions, energy consumption, and economic growth in Brazil," *Energy*, vol. 36, no. 5, pp. 2450–2458, 2011.
- [24] N.-m. Xie and S.-f. Liu, "Discrete grey forecasting model and its optimization," *Applied Mathematical Modelling*, vol. 33, no. 2, pp. 1173–1186, 2009.
- [25] L. Wu, S. Liu, L. Yao, S. Yan, and D. Liu, "Grey system model with the fractional order accumulation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 7, pp. 1775–1785, 2013.
- [26] B. Zeng, H. Duan, and Y. Zhou, "A new multivariable grey prediction model with structure compatibility," *Applied Mathematical Modelling*, vol. 75, pp. 385–397, 2019.
- [27] T. L. Tien, "A research on the grey prediction model GM(1, n)," *Applied Mathematics and Computation*, vol. 218, no. 9, pp. 4903–4916, 2012.
- [28] R. Jia and Q. Y. Xiong, "Two-dimensional temperature field distribution reconstruction based on least square method and radial basis function approximation," *Mathematical Problems in Engineering*, vol. 2017, Article ID 1213605, 7 pages, 2017.
- [29] B. M. Ayeni, "Making ordinary least squares linear classifiers more robust," *Machine Learning*, vol. 107, pp. 1–9, 2018.
- [30] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, Heidelberg, Germany, 1995.
- [31] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 23, no. 3, pp. 533–536, 1986.
- [33] B. Zeng, C. M. Luo, and S. F. Liu, "Development of an optimization method for the GM(1, N) model," *Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation*, vol. 55, pp. 553–362, 2016.