

Article

Production Feature Analysis of Global Onshore Carbonate Oil Reservoirs Based on XGBoost Classifier

Guilin Qi ^{1,2,3} and Baolei Liu ^{1,2,3,*}

¹ School of Petroleum Engineering, Yangtze University, Wuhan 430100, China; 2021710246@yangtzeu.edu.cn

² Key Laboratory of Exploration Technologies for Oil and Gas Resources, Yangtze University, Ministry of Education, Wuhan 430100, China

³ Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Yangtze University, Wuhan 430100, China

* Correspondence: baoleiliu@yangtzeu.edu.cn

Abstract: Carbonate reservoirs account for 60% of global reserves for oil, making them one of the most important types of sedimentary rock reservoirs for petroleum production. This study aimed to identify key production features that significantly impact oil production rates, enhancing reservoir management and optimizing production strategies. A comprehensive dataset is built from reserves and production history data of 377 onshore carbonate oilfields globally, encompassing features such as production, recovery rate, and recovery degree of the whole lifecycle of an oilfield. XGBoost classifier is trained by K-fold cross-validation and its hyperparameters are optimized by Optuna optimization framework. The results show that XGBoost has the best performance evaluated with metrics including accuracy, precision, recall, and F1 score comparing with decision tree, random forest, and support vector machine. Key production features are identified by analyzing the classification feature importance of XGBoost classifier, including build-up stage cumulative production, plateau stage cumulative production, plateau stage recovery rate, plateau stage recovery degrees, and peak production. In conclusion, oilfield reserve size, build-up stage cumulative production, plateau stage cumulative production, and peak production increase, while plateau stage recovery rate decreases, and the plateau stage recovery degree of small-sized oilfields is slightly greater than that of moderate and large oilfields. The research methodology of this study can serve as a reference for studying production features of other types of oil and gas reservoirs. By applying the methodology to low-permeability oilfields, this paper concludes the key production features that are as follows: low-permeability oilfields generally have lower peak recovery rate, lower plateau stage recovery rate, lower decline stage recovery degree, and lower decline stage recovery rate, along with a wide but generally lower range of decline stage cumulative production compared to conventional oilfields.



Citation: Qi, G.; Liu, B. Production Feature Analysis of Global Onshore Carbonate Oil Reservoirs Based on XGBoost Classifier. *Processes* **2024**, *12*, 1137. <https://doi.org/10.3390/pr12061137>

Academic Editors: Qingbang Meng, Bin Liang and Zhan Meng

Received: 13 May 2024

Revised: 26 May 2024

Accepted: 30 May 2024

Published: 31 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Carbonate reservoirs are widely distributed with stable sediment thickness, making them one of the most significant sedimentary rock reservoir types for oil and gas [1]. According to statistics, carbonate reservoirs account for 60% and 40% of global reserves for oil and gas, respectively [2]. Their contribution to the world's oil and gas production exceeds 60% [3,4]. In terms of global production distribution, the Middle East region possesses the richest oil and gas resources, contributing to approximately two-thirds of the global production, mainly from carbonate reservoirs (70% of oil and 90% of natural gas) [5,6]. In North America, half of the total oil production comes from carbonate reservoirs [7,8]. In China, carbonate reservoirs account for 27% and 26.9% of oil and gas resources, respectively [9]. Moreover, understanding the production features of carbonate reservoirs is of paramount importance for the sustainable, efficient, safe, and rapid development of the oil and gas industry. However, the study of carbonate oilfield production features

faces challenges such as large data volume, diverse information types, and processing difficulties. Furthermore, traditional statistical analysis methods make it difficult to quickly extract useful information, while data mining methods can efficiently and accurately extract production features. Therefore, applying data mining methods to explore the implicit value in production features of onshore carbonate reservoirs can help deepen the understanding of the production features of carbonate reservoirs.

Over the past few decades, scholars have conducted extensive research on the features of carbonate reservoirs. In 1985, Perry Owen Roehl pointed out in his work "Carbonate Petroleum Reservoirs" that the formation of carbonate oilfields has undergone complex diagenesis, exhibiting characteristics of wide distribution with stable deposition thickness and good reservoir rocks [2–4]. In 1992, Dominguez presented a comprehensive treatment of the geometry, porosity, permeability evolution, and production characteristics of carbonate reservoirs [10]. In 1997, Alsharhan and Nairn researched the geological structures of the Middle East and the history of oil exploration in the previous decades in their collaborative work "Sedimentary Basins and Petroleum Geology of the Middle East", analyzing in detail the geological characteristics of carbonate oilfields in the Middle East [11]. In 2000, Lv and Jin summarized discussions on the distribution patterns of carbonate oil and gas fields both domestically and internationally [12]. In 2001, Zhang analyzed the characteristics of Ordovician carbonate fractured reservoirs in the Tarim Oilfield through dynamic reservoir calculations [13]. In 2005, Fan conducted a comprehensive study of carbonate reservoirs worldwide, identifying important factors for the formation of world-class oil and gas fields, providing a reference direction for the study of carbonate reservoirs in China [14]. In 2008, Luo et al. studied carbonate reservoirs in oil and gas areas such as Tarim, Sichuan, Ordos, Bohai Bay, southern regions, and Qiangtang, outlining the basic characteristics of carbonate reservoirs [15]. In 2008, Camacho-Velazquez et al. performed the decline-curve analysis in fractured carbonate reservoirs [16]. In 2010, Lv et al. researched water flooding in bottom-water carbonate reservoirs through laboratory experiments, providing a basis for predicting recoverable reserves and water cuts [17]. In 2011, Chang et al. summarized the patterns of water cuts in large-scale fractured carbonate oil reservoirs based on the relationship between the oil–water interface and fractures [18]. In 2014, Zhang studied the distribution and formation of large carbonate oil and gas fields, discovering distribution patterns of these fields [19]. In 2021, Cheng analyzed the production characteristics of typical fractured reservoirs using the example of the Tarim Oilfield in the Tarim Basin [20]. In 2022, Xiong and He explored the distribution of carbonate reservoirs by analyzing features such as regional distribution, lithology, trapping type, reservoir burial depth, and reserve scale of 94 global major oil and gas fields in carbonate formations [21]. In the aforementioned studies, researchers have extensively explored the characteristics of carbonate reservoirs, covering aspects such as geological structures, physical properties, and production patterns.

According to Ji et al., production features vary at different lifecycle stages of the oilfield [22]. The lifecycle stages of hydrocarbon fields are commonly divided into the build-up stage, plateau stage, and decline stage (see Figure 1) [23–25]. During the build-up stage, the continuous drilling of development wells and the construction of infield infrastructure take place. Engineers conduct adjustments and optimizations to maximize oil production, ensuring economic efficiency. In the plateau stage, the hydrocarbon field reaches its production peak and maintains a steady production fluctuating around this apex. Hirsch defined the plateau stage based on the production change rate, with the fluctuation range during this stage typically within 3–4% [26]. As time progresses, the hydrocarbon field enters a decline stage, and its production begins to decrease [27].

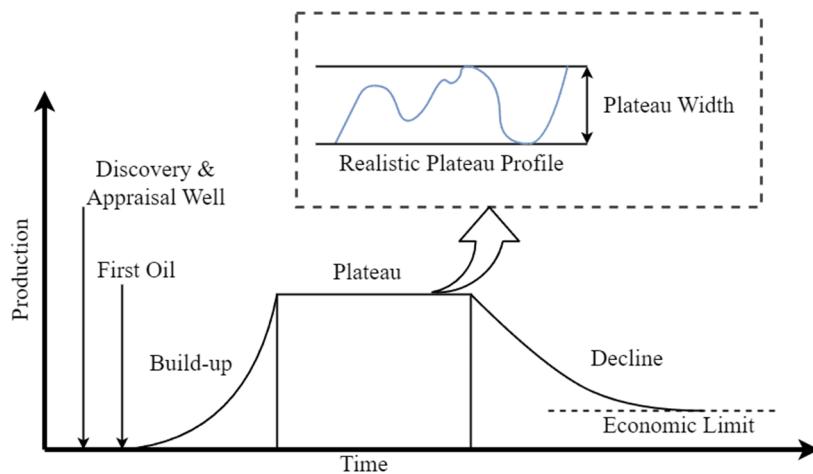


Figure 1. Lifecycle of an oilfield adapted from Höök et al. [24].

However, the current research still has shortcomings: (1) previous studies focused on wells, oilfields, or blocks, lacking a comprehensive understanding of onshore carbonate reservoirs on a global scale and (2) systematic research on the production features throughout the lifecycle of carbonate oilfields is insufficient.

Data mining is the technique of extracting hidden information, patterns, and knowledge from large-scale datasets [28]. Since the beginning of the digital era, the oil and gas industry has extensively used data mining to analyze and predict reservoir properties, geological structures, and oil and gas distribution, providing important support and guidance for decision making in exploration and development, reservoir evaluation, and production control [29–31]. Data mining tasks cover various areas such as clustering, regression, and classification (Figure 2) [32]. Clustering groups data into clusters with similar features based on certain criteria without using predefined labels [33]. Regression is used to predict numerical target variables by identifying relationships between dependent and independent variables [34]. Classification deals with categorizing data into predefined labels by training models to distinguish between different categories [35].

As a supervised learning method, classification seeks to place dataset samples into predefined groups. Common methods include random forests, logistic regression, decision trees, neural networks, and support vector machines. In 2005, Yang et al. used well-logging data to create an oil and gas layer recognition model with SVM [36]. In 2007, Liu et al. categorized oil–water layers in the Qijia-Gulong Sag in Daqing using SVM and Bayes techniques [37]. In 2008, Zhang Yinde et al. employed SVM to detect fluids in low-resistance oil layers [38]. In 2019, Ghiasi et al. evaluated the conditions for sand production in reservoirs using decision trees, random forests, and extra trees [39]. In 2022, Kumar et al. automatically detected and described homogeneous reservoirs with an accuracy rate of more than 79% with artificial neural networks and evolutionary algorithms. In 2023, Somi et al. amalgamated hidden Markov model-based clustering with XGBoost to identify sleeve incidents [40]. In 2024, Khaled et al. reliably predicted bottom-hole circulating temperature under constant conditions with XGBoost [41].

While previous studies have extensively examined various aspects of carbonate reservoirs, including their geological structures, physical properties, and production patterns, the comprehensive global perspective and systematic analysis of onshore carbonate oilfields throughout their lifecycle are still insufficient. By addressing the shortcomings of prior research, particularly the lack of a comprehensive understanding on a global scale and insufficient systematic research on production features across the lifecycle of oilfields, this paper will focus on the production features of global onshore carbonate oilfields and delve into the typical production features throughout their lifecycle to facilitate the rational development and sustainable utilization of onshore carbonate oilfields. Based on the production and reserve data, this paper will extract production features of onshore

carbonate oilfields throughout the lifecycle of oil production, including the build-up stage, plateau stage, decline stage, initial year, and peak year. Next, the XGBoost classifier will be employed to ascertain the key production feature, followed by a comprehensive analysis of the important production feature.

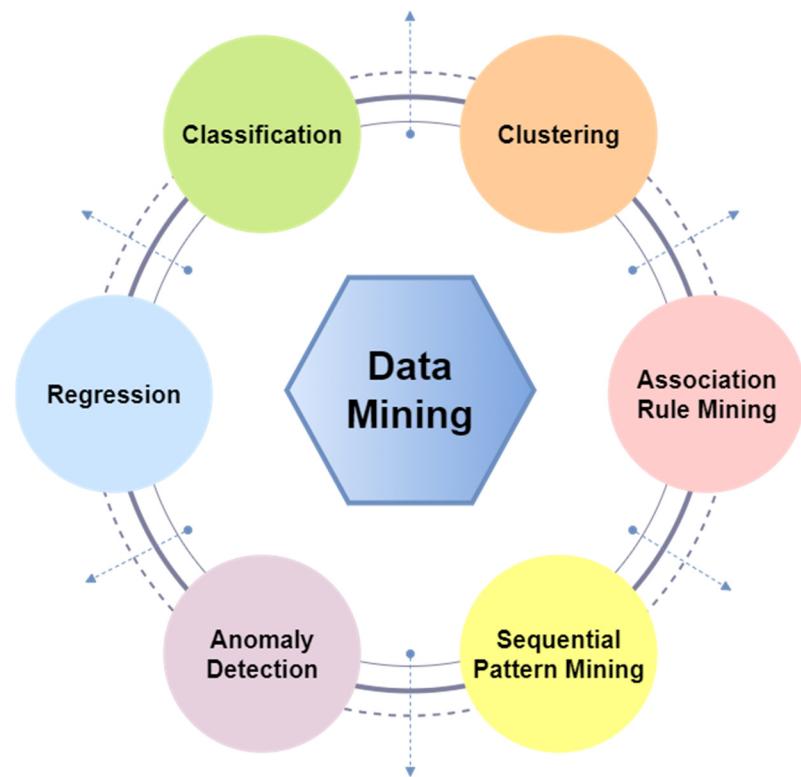


Figure 2. Types of data mining algorithms.

2. Materials and Methods

2.1. Methods

2.1.1. XGBoost

XGBoost (eXtreme Gradient Boosting), presented by Chen and Carlos [42], is a powerful and popular machine learning algorithm known for its efficiency and performance in classification problems. It is based on the gradient boosting framework, where weak learners (typically decision trees) are sequentially added and trained to correct the errors made by previous models. The objective function (L) in XGBoost is a key component that guides the training process by quantifying the loss term (l) between predicted values (\hat{y}_i) and actual values (y_i) while also incorporating regularization terms (Ω) to prevent overfitting. For a given dataset with n examples, the objective function at iteration t is shown in Equation (1).

$$L^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \quad (1)$$

where $f_t(x_i)$ represents the prediction for sample i by the t -th weak learners, x_i represents the features of the i -th examples.

The regularization term in XGBoost controls the number of trees (T) and weights of the leaves in the trees (ω), where γ and λ are hyperparameters, as shown in Equation (2). By adjusting γ and λ , the model can control the complexity of the ensemble and prevent overfitting, ultimately improving generalization performance [43].

$$\Omega(f_t) = \gamma T_t + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \quad (2)$$

However, Equation (1) includes functions as parameters and cannot be optimized using traditional optimization methods in Euclidean space. Thus, the second-order Taylor expansion is applied to Equation (1), where g_i and h_i are the first- and second-order gradient statistics of the loss function.

$$L^{(t)} \approx \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right) + \Omega(f_t), \quad (3)$$

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right), \quad (4)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l\left(y_i, \hat{y}_i^{(t-1)}\right), \quad (5)$$

For a fixed structure, the optimal weight (ω_j^*) and the corresponding optimal value can be calculated in Equations (6) and (7).

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad (6)$$

$$L^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (7)$$

It is necessary to assess every potential tree structure in order to determine the best one. Generally, enumerating every possible tree structure is impractical. A greedy approach is implemented by XGBoost, which begins with a single leaf and iteratively adds branches to the tree.

2.1.2. Evaluation Metrics

Classification models can be evaluated using several metrics, such as confusion matrix accuracy, precision, recall, and F1 score [44].

The confusion matrix is used to evaluate the performance of a classification model by comparing the actual and predicted classes [45]. The matrix consists of four primary elements: True Positive (TP) refers to the cases where the model correctly predicts the positive class; True Negative (TN) refers to the cases where the model correctly predicts the negative class; False Positive (FP) refers to the cases where the model predicts the positive class, but it is actually negative; and False Negative (FN) refers to the cases where the model predicts the negative class, but it is actually positive. A two-dimensional confusion matrix is shown in Table 1.

Table 1. An example of two-dimensional confusion matrix.

	Predicted Negative	Predicted Positive
Actual Negative	True Positive	False Positive
Actual Positive	False Positive	True Negative

Using a confusion matrix, various performance metrics like accuracy, precision, recall, and F1 score can be calculated, which help to evaluate how well the classification model is performing and where it might be making errors.

Accuracy is a metric that quantifies the ratio of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (8)$$

Precision is a metric that quantifies the ratio of correctly identified positive samples to the total number of samples projected as positive by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (9)$$

Recall quantifies the ratio of correctly identified positive samples to the total number of positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (10)$$

The F1 score is a metric that quantifies the balance between precision and recall. It is a useful measure for evaluating the accuracy and recall of a model.

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

2.2. Data Preparation

2.2.1. Data Collection and Preprocessing

This paper acquires the production and reserve data for developing global onshore carbonate oilfields from 1965 to 2023 by merging data from IHS Markit and Wood Mackenzie databases. Then, the production features of the lifecycle of oil reservoir, including build-up stages, plateau stages, decreasing stages, initial year, and peak year, are identified and extracted. Details are shown in Table 2.

Table 2. Production features of oilfield production.

Production Stage	Feature	Description	Unit
Build-up Stage	buildup_duration	duration of the build-up stage	year
	buildup_cumulative_production	total production in the build-up stage	megatons
	buildup_recovery_rate	average recover rate in the build-up stage	percent
	buildup_recovery_degree	overall recovery in the build-up stage	percent
Plateau Stage	plateau_duration	duration of the plateau stage	year
	plateau_cumulative_production	total production in the plateau stage	megatons
	plateau_recovery_rate	average recover rate in the plateau stage	percent
	plateau_recovery_degree	overall recovery in the plateau stage	percent
Decline Stage	decline_duration	duration of the decline stage	year
	decline_cumulative_production	total production in the decline stage	megatons
	decline_recovery_rate	average recover rate in the decline stage	percent
	decline_recovery_degree	overall recovery in the decline stage	percent
Initial Year	initial_production	production in the initial year	megatons
	initial_recovery_rate	recover rate in the initial year	percent
Peak Year	peak_production	production in the peak year	megatons
	peak_recovery_rate	recover rate in the peak year	percent

This study aims to provide support for the development of carbonate oilfields by exploring their production features. To ensure the accuracy and credibility of data mining, the oilfield data are preprocessed according to the following criteria:

1. Exclude oilfields with cumulative total production of less than 0.1 million tons.
2. Exclude oilfields with production years of less than or equal to 10 years.
3. Exclude oilfields with an incomplete lifecycle.

After preprocessing, this study constructed a production feature dataset of onshore carbonate oilfields with 377 instances and 16 features.

2.2.2. Data Labeling

Due to the different geological environments such as geological formations, lithology, and depositional environments, development technologies and equipment, costs, and production efficiency may vary significantly. Therefore, to explore the production features of onshore carbonate oilfields, this paper will divide them based on their recoverable reserves. Jin pointed out that oilfields can be classified into small, moderate, large, super-large, and giant fields based on recoverable reserves (Table 3) [46].

Table 3. Classification criteria of onshore oilfields [46].

Class	Geological Reserves	Recoverable Reserves
Giant oilfields	>1500	>450
Super-large oilfields	500~1500	150~450
Large oilfields	100~500	30~150
Moderate oilfields	10~100	3~30
Small oilfields	<10	<3

In this paper, to balance dataset, large, super-large, and giant oilfields are collectively referred to as large fields. Based on Jin's classification standard, the dataset comprises 377 onshore carbonate oilfields, including 290 small fields, 67 moderate fields, and 20 large fields, as detailed in Table 4.

Table 4. Statistics of onshore carbonate oilfields in the dataset.

Class	Number
Small	290
Moderate	67
Large	20

There is a significant imbalance in the oilfield distribution of classifications. This imbalance poses challenges when constructing classification models, particularly in dealing with the minority class samples. To address this issue and ensure the accuracy of the data mining model in classifying these minority samples accurately, it is necessary to assign greater weights to these samples, as shown in Equation (12).

$$w_i = \frac{N}{mn_i}, \quad (12)$$

where w_i is the weight assigned to class i , N is the total number of samples in the dataset, m is the number of classes, and n_i is the number of samples in class i .

3. Results and Discussion

3.1. Construction of XGBoost Classifier

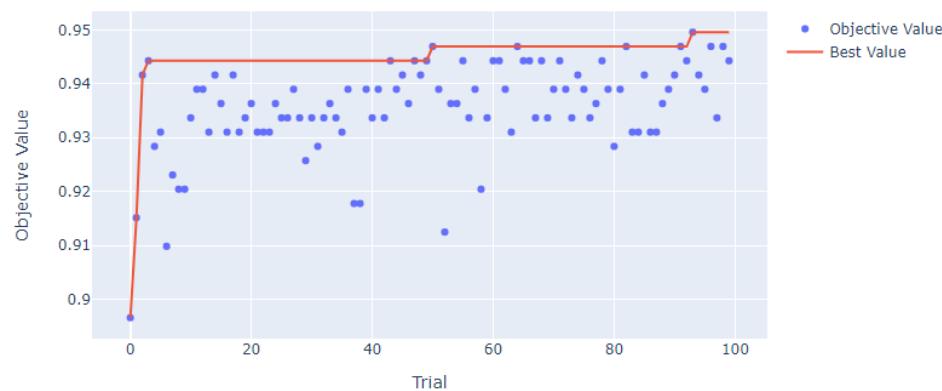
XGBoost is used for classifying onshore carbonate oilfields to determine their key production features. To search for the optimal hyperparameters for the XGBoost model, the Optuna optimization framework is employed, and hyperparameters space are listed in Table 5. Classification accuracy is used as the objective function. The cost complexity parameter (CCP_alpha) of the model is set to 0.01 to prevent overfitting, and class weights are set to balance the dataset.

Table 5. Hyperparameters space of XGBoost.

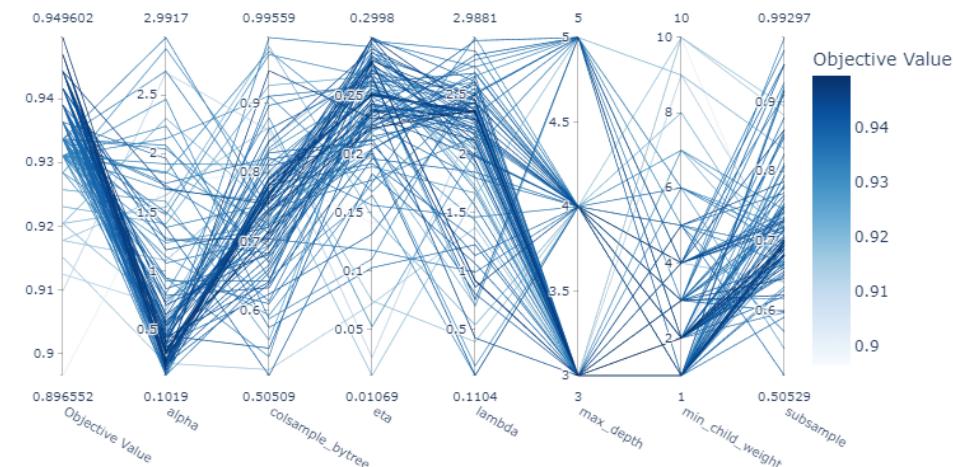
Hyperparameter	Description	Type	Search Space
eta	Learning rate	Float	[0.01, 0.3]
max_depth	Maximum depth of a tree	Int	[3, 5]
subsample	Subsample ratio of the training instance	Float	[0.5, 1.0]
colsample_bytree	Subsample ratio of columns when constructing each tree	Float	[0.5, 1.0]
alpha	L1 regularization term (alpha)	Float	[0.1, 3.0]
lambda	L2 regularization term (lambda)	Float	[0.1, 3.0]
min_child_weight	Minimum sum of instance weight needed in a child	Int	[1, 10]

After 100 trials of Optuna optimization, the gradient boosting reached its best objective value (Figure 3) and best hyperparameters (Figure 4), with the eta of 0.17, the max_depth of 4, the subsample of 0.54, the colsample_bytree of 0.55, the alpha of 0.71, the lambda of 1.35, and the min_child_weight of 4. According to the importance of hyperparameters (Figure 5), the min_child_weight has a significant impact on the accuracy of the gradient boosting.

Optimization History Plot

**Figure 3.** Optimization history plot of XGBoost classifier.

Parallel Coordinate Plot

**Figure 4.** Parallel coordinate plot of XGBoost classifier.

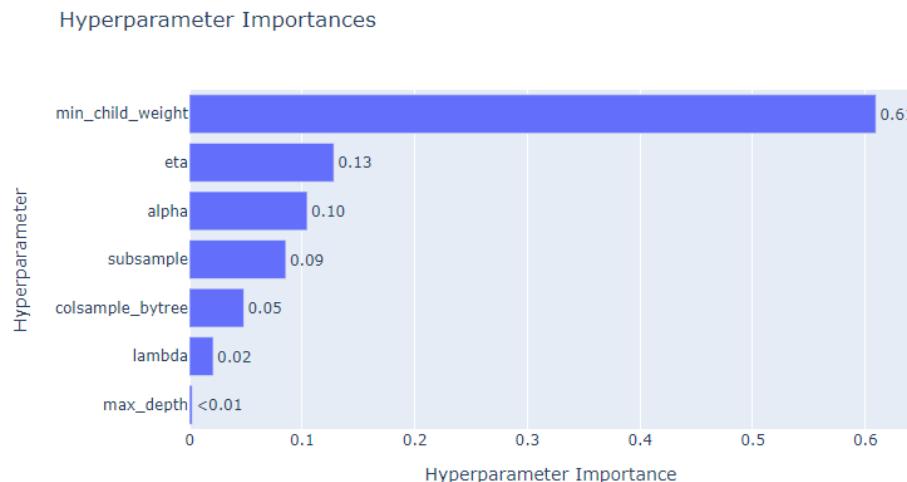


Figure 5. Hyperparameter importance plot of XGBoost classifier.

Based on the best hyperparameters, an XGBoost model is used to classify the onshore carbonate oilfield dataset, and its performance is evaluated with a five-fold cross-validation. In the analysis of the confusion matrix (Figure 6), the XGBoost model performs well overall, with particularly high precision and recall for small onshore oilfields. It demonstrates high accuracy and generalization capability in this classification task.

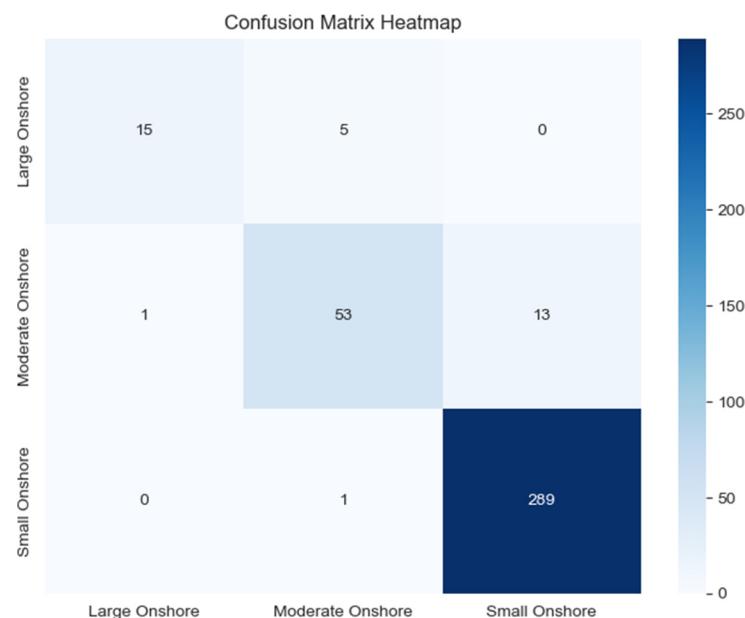


Figure 6. Confusion matrix of XGBoost classifier.

3.2. Evaluation of XGBoost Classifier

To evaluate the XGBoost classifier, decision tree (DT), random forest (RF), and support vector machine (SVM) are introduced to classify the onshore carbonate oilfields dataset as a comparison. In addition, Optuna optimization framework is utilized to optimize the hyperparameters of these models. The classification results, including XGBoost, DT, RF, and SVM, are presented in Table 6.

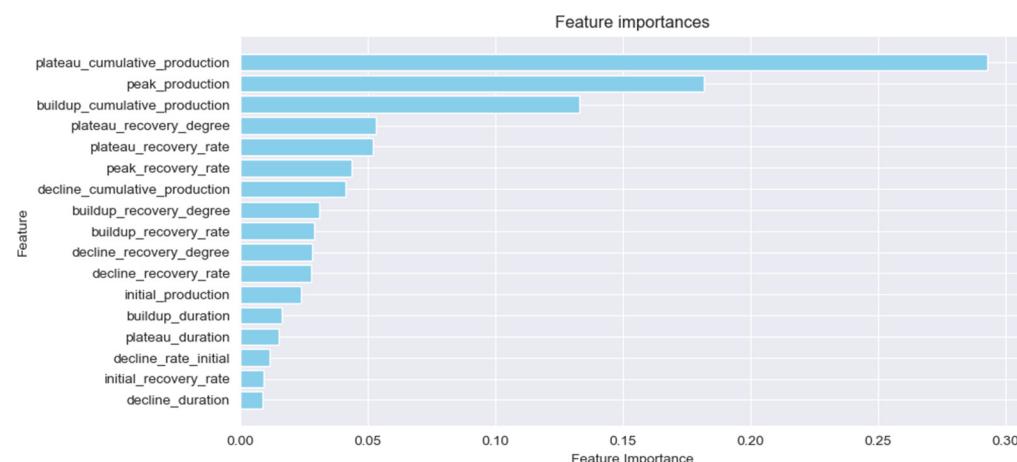
Table 6. Classification reports of DT, RF, SVM, and XGBoost in onshore carbonate oilfield dataset.

Model	Class	Precision	Recall	F1 Score	Accuracy
DT	Large	0.54	0.70	0.61	0.87
	Moderate	0.61	0.69	0.65	
	Small	0.97	0.92	0.94	
RF	Large	0.83	0.75	0.79	0.93
	Moderate	0.78	0.87	0.82	
	Small	0.98	0.96	0.97	
SVM	Large	0.87	0.65	0.74	0.92
	Moderate	0.82	0.69	0.75	
	Small	0.94	0.99	0.96	
XGBoost	Large	0.78	0.70	0.74	0.95
	Moderate	0.88	0.84	0.85	
	Small	0.98	0.99	0.98	

Based on Table 6, XGBoost demonstrates high precision and recall and balanced performance and overall accuracy compared to other models like decision tree (DT), random forest (RF), and support vector machine (SVM) in terms of precision, recall, F1 score, and accuracy across different categories of onshore oilfields.

3.3. Selection of Key Production Features

As shown in Figure 7, the feature importance of the XGBoost model such as build-up cumulative production, plateau cumulative production, plateau recovery rate, plateau recovery degree, and peak production are greater than 0.05, indicating substantial differences in these features among onshore carbonate oilfields with different reserves. Therefore, these five features are chosen as the key production features for onshore oilfield production analysis. The pair plot provides a visual analysis of five key features for small oilfields (red points), moderate oilfields (blue points), and large oilfields (green points) compared to conventional oilfields (blue points), which are shown in Figure 8.

**Figure 7.** XGBoost classifier feature importance of onshore oilfields.

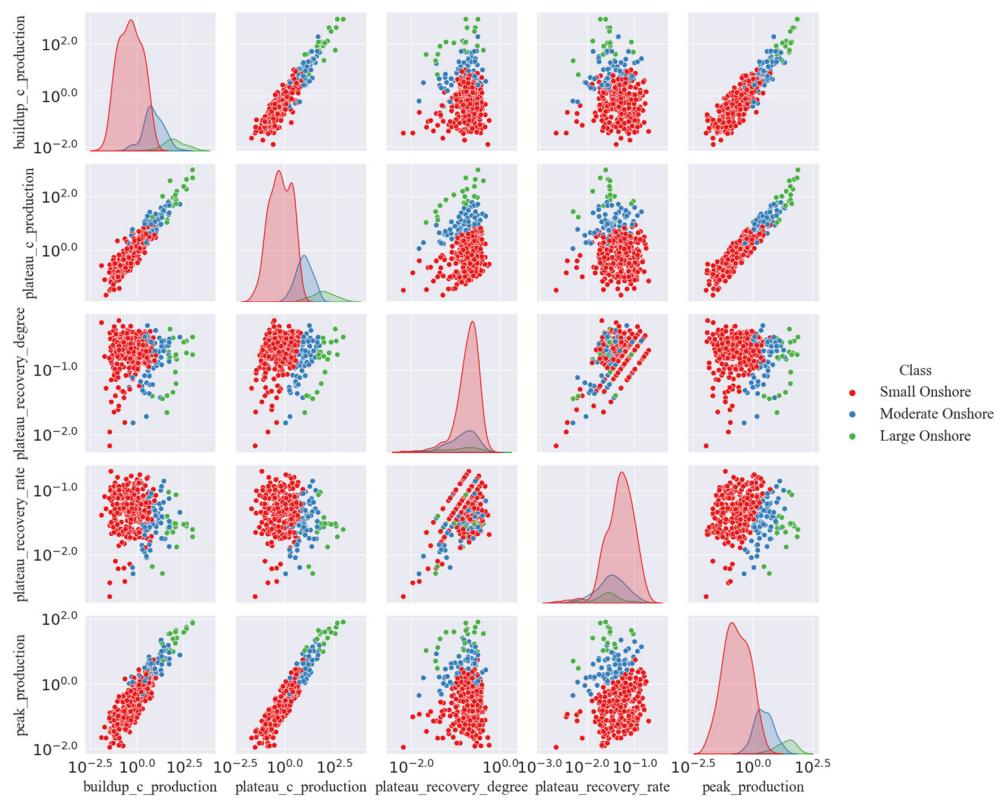


Figure 8. Pair plot of key production features.

3.4. Analysis of Key Production Features

Box plots and violin plots are utilized to provide a visual representation of the distribution of various parameters in onshore oilfields, highlighting differences in key production features between onshore carbonate oilfields. A box plot is a graphical method to display data distribution and identify outliers with statistical measures such as minimum value, first quartile, median, third quartile, and maximum value. On the other hand, a violin plot combines a box plot with a kernel density estimation plot, offering a more detailed and intuitive view of the data's density distribution.

In the analysis of build-up stage cumulative production, plateau stage cumulative production, and peak production, data span several orders of magnitude. Therefore, logarithmic scales are employed in both box plots and violin plots to appropriately represent this wide range of values. Conversely, for features such as plateau stage recovery rate and plateau stage recovery degree, data distribution tends to be more uniform. Therefore, the linear scales are suitable for visualizing these evenly distributed data.

3.4.1. Build-Up Stage Cumulative Production

Box plot and violin plot of build-up stage cumulative production are shown in Figure 9, with y-axis presented on a logarithmic scale.

Small onshore carbonate oilfields exhibit relatively lower values in the cumulative production during the build-up stage. The mean production is 2.26 megatons (Mt), with a spread that is not too wide. However, there is notable variability, suggesting diverse production capacities among these fields. The lower quartiles (0.44 Mt) reflect a range of lower production values, while the upper quartiles (2.72 Mt) indicate fields with higher production levels.

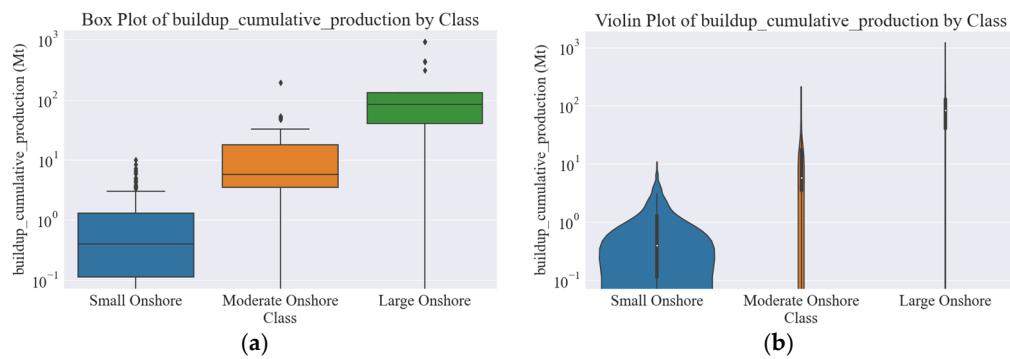


Figure 9. Box plot and violin plot of build-up stage cumulative production on onshore oilfields: (a) box plot and (b) violin plot.

Moderate onshore carbonate oilfields demonstrate a substantial increase in cumulative production during the build-up stage compared to the small onshore fields. The mean production (26.42 Mt) is significantly higher, with a wider spread indicating diverse production capacities. The lower quartiles (6.65 Mt) are notably higher, reflecting a baseline of relatively elevated production levels, while the upper quartiles (27.08 Mt) indicate fields with substantially higher production capacities.

Large onshore carbonate oilfields showcase a considerable increase in cumulative production during the build-up stage, with significantly higher mean production (218.15 Mt) compared to both small and moderate onshore fields. The spread of data is wide, indicating a broad range of production capacities among these fields. The lower quartiles (34.42 Mt) still exhibit substantial production levels, while the upper quartiles (318.60 Mt) show fields with exceptionally high production capacities.

Overall, the cumulative production of oilfields in the build-up stage increases significantly with the increase in field reserve size.

3.4.2. Plateau Stage Cumulative Production

Box plot and violin plot of plateau stage cumulative production are shown in Figure 10, with y-axis presented on a logarithmic scale.

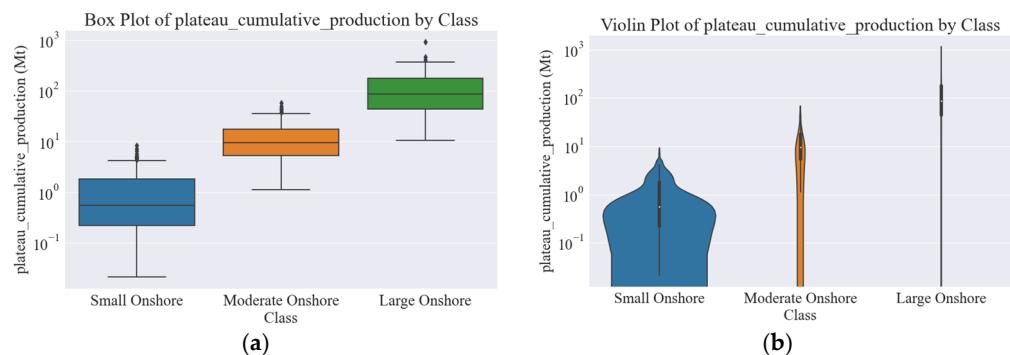


Figure 10. Box plot and violin plot of plateau stage cumulative production on onshore oilfields: (a) box plot and (b) violin plot.

Small carbonate oilfields exhibit lower average cumulative production (1.21 Mt) during the plateau stage, smaller standard deviation, and narrower data distribution range. The distribution of cumulative production during the plateau stage in this class of fields is relatively concentrated, indicating a relatively stable production level.

Moderate carbonate oilfields show a significant increase in cumulative production during the plateau stage, with an average (14.03 Mt) much higher than that of small fields. Additionally, they have a larger standard deviation, indicating greater variability in data

distribution. The data range is wide, with maximum values (57.69 Mt) reaching high levels, reflecting the potential and variability in production for moderate fields.

Large carbonate oilfields demonstrate a significant increase in both the quantity and range of cumulative production during the plateau stage. Their average (173.47 Mt) and standard deviation are much higher than those of the other two types of fields, indicating significant production differences and more dispersed data distribution. Their maximum value (926.06 Mt) far exceeds that of moderate and small carbonate oilfields, highlighting the immense potential production of large fields.

Overall, there is a gradual increase in cumulative production during the plateau stage with the increase in scale of onshore carbonate oilfields.

3.4.3. Plateau Stage Recovery Rate

Box plot and violin plot of plateau stage recovery rate are shown in Figure 11, with y-axis presented on a linear scale.

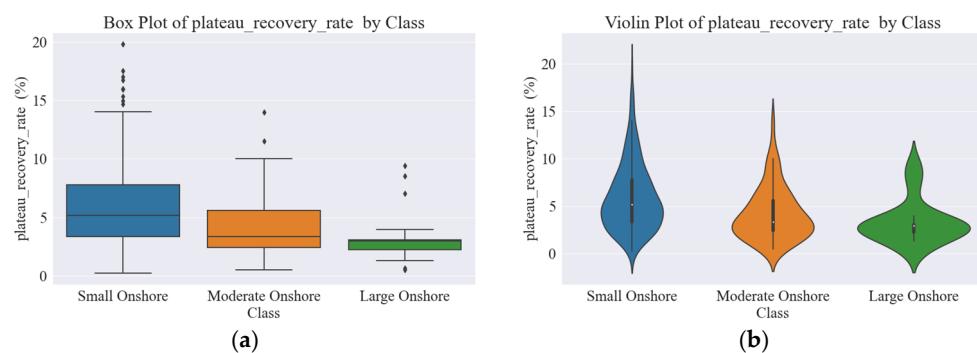


Figure 11. Box plot and violin plot of plateau stage recovery rate on onshore oilfields: (a) box plot and (b) violin plot.

Small carbonate oilfields exhibit a relatively high average recovery rate (5.89%) during the plateau stage, with a smaller standard deviation and a narrower range of data distribution. The data distribution of this field is relatively concentrated, indicating a relatively stable level of production rate.

Moderate carbonate oilfields show a relatively low average recovery rate (4.07%) during the plateau stage, with a small standard deviation, indicating less variability in data distribution.

Large carbonate oilfields demonstrate a significant decrease in both the quantity and range of recovery rate during the plateau stage. Their average (3.16%) and standard deviation are much lower than those of the other two types of fields, indicating a relatively slower recovery rate and a more concentrated data distribution.

Overall, the recovery rate decreases with the increase in field size, with the largest fields exhibiting the slowest production rates.

3.4.4. Plateau Stage Recovery Degree

Box plot and violin plot of plateau stage recovery degree are shown in Figure 12, with y-axis presented on a linear scale.

Small onshore oil fields demonstrate a relatively high average recovery degree (21.29%) with moderate variability. The distribution of recovery degrees in these fields tends to be consistent, with most fields showing recovery degrees within a narrow range around the average.

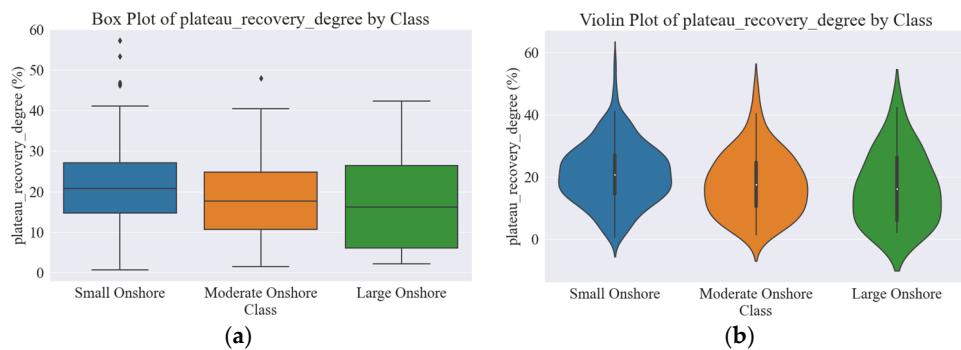


Figure 12. Box plot and violin plot of plateau stage recovery degree on onshore oilfields: (a) box plot and (b) violin plot.

Moderate onshore oil fields exhibit a slightly lower average recovery degree (17.52%) compared to small fields, with a similar level of variability. The recovery degrees in these fields are more spread out, indicating a wider range of performance among different fields.

Large onshore oil fields show the lowest average recovery degree (16.81%) among the three types, with the highest variability. The recovery degrees in these fields vary widely, indicating significant differences in performance across different fields.

Overall, as the size of the carbonate oil fields increases from small to large, the plateau stage recovery degree exhibits a trend of increasing variability and decreasing mean recovery degree. Small onshore oil fields have the highest mean recovery degree and the least variability; moderate onshore oil fields have a lower mean recovery degree with moderate variability; large onshore oil fields display the lowest mean recovery degree with the highest variability.

3.4.5. Peak Production

Box plot and violin plot of peak production are shown in Figure 13, with y-axis presented on a logarithmic scale.

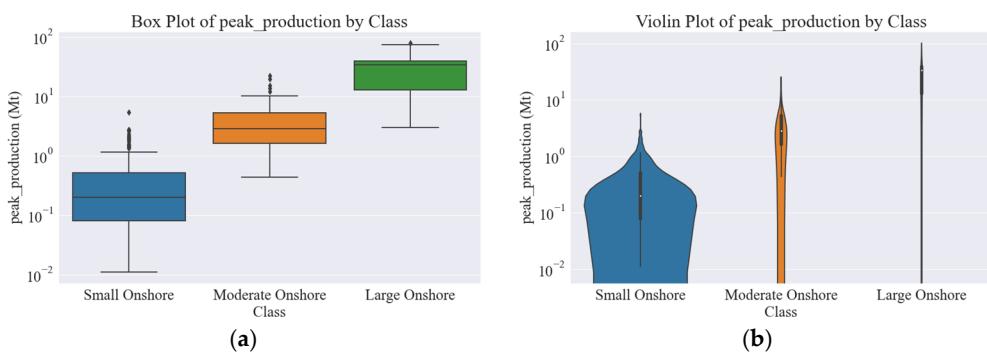


Figure 13. Box plot and violin plot of peak production on onshore oilfields: (a) box plot and (b) violin plot.

The peak production of small carbonate oilfields shows a relatively low average value (0.43 Mt), a smaller standard deviation, and a narrower range of data distribution. The data distribution in this category of fields is relatively concentrated, indicating a relatively stable level of production.

Moderate carbonate oilfields exhibit a noticeable increase in peak production, with their average value (4.31 Mt) much higher than that of small fields. Additionally, they have a larger standard deviation, indicating greater variability in data distribution. The data range is broader, with a higher maximum value (22.01 Mt), reflecting the potential and variability in the production of moderate fields.

Large carbonate oilfields demonstrate a significant increase in both the quantity and range of peak production. Their average value (30.94 Mt) and standard deviation are

much higher than those of the other two types of fields, indicating greater differences in production and a more dispersed data distribution. The maximum value (78.86 Mt) highlights the substantial potential production of large fields.

Overall, as the scale of carbonate oilfields increases, there is a gradual increase in peak production, accompanied by increased variability in data distribution.

3.5. Appliance in Low-Permeability Onshore Carbonate Oilfields

The research methodology described above is applicable and can serve as a reference for studying the production characteristics of various types of oil and gas reservoirs, including low-permeability onshore carbonate oilfields.

This paper classifies the dataset based on permeability to analyze the production features of these fields. Oilfields with permeability of less than 10 millidarcys (mD) are defined as low-permeability oilfields, while oilfields with permeability of more than 10 mD are defined as conventional reservoirs. According to this standard, the dataset comprises 149 onshore carbonate oilfields, with 22 categorized as low-permeability and 127 as conventional fields, as detailed in Table 7.

Table 7. Statistics of onshore carbonate oilfields in the dataset labeled by permeability.

Class	Number
Low-permeability reservoirs	22
Conventional reservoirs	127

To determine the feature importance of the dataset labeled by permeability, an Optuna-optimized XGBoost classifier is employed. As illustrated in Figure 14, the five most important features are peak recovery rate, decline cumulative production, plateau recovery rate, decline recovery degree, and decline recovery rate. A pair plot of these features is presented in Figure 15.

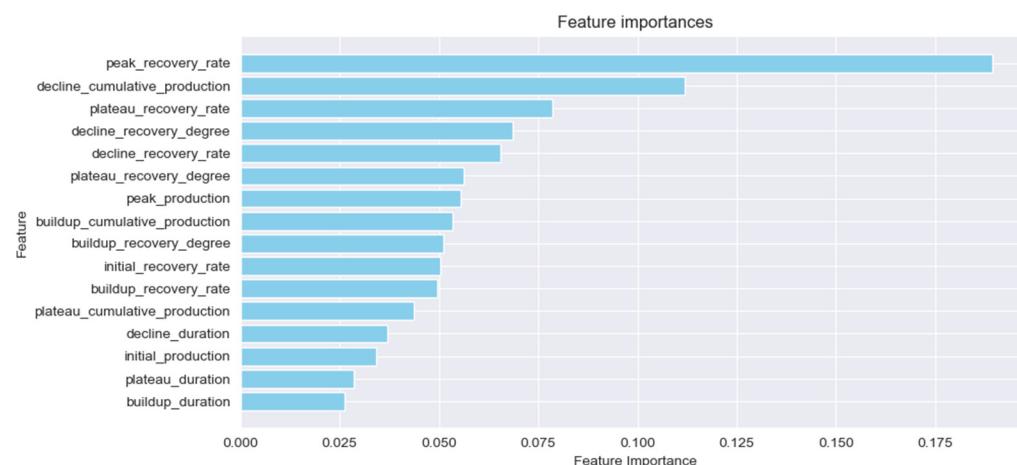


Figure 14. XGBoost classifier feature importance of the dataset labeled by permeability.

The analysis from the pair plot indicates that low-permeability oilfields generally have lower peak recovery rate, lower plateau recovery rate, lower decline recovery degree, and lower decline recovery rate, along with a wide but generally lower range of decline cumulative production compared to conventional oilfields.

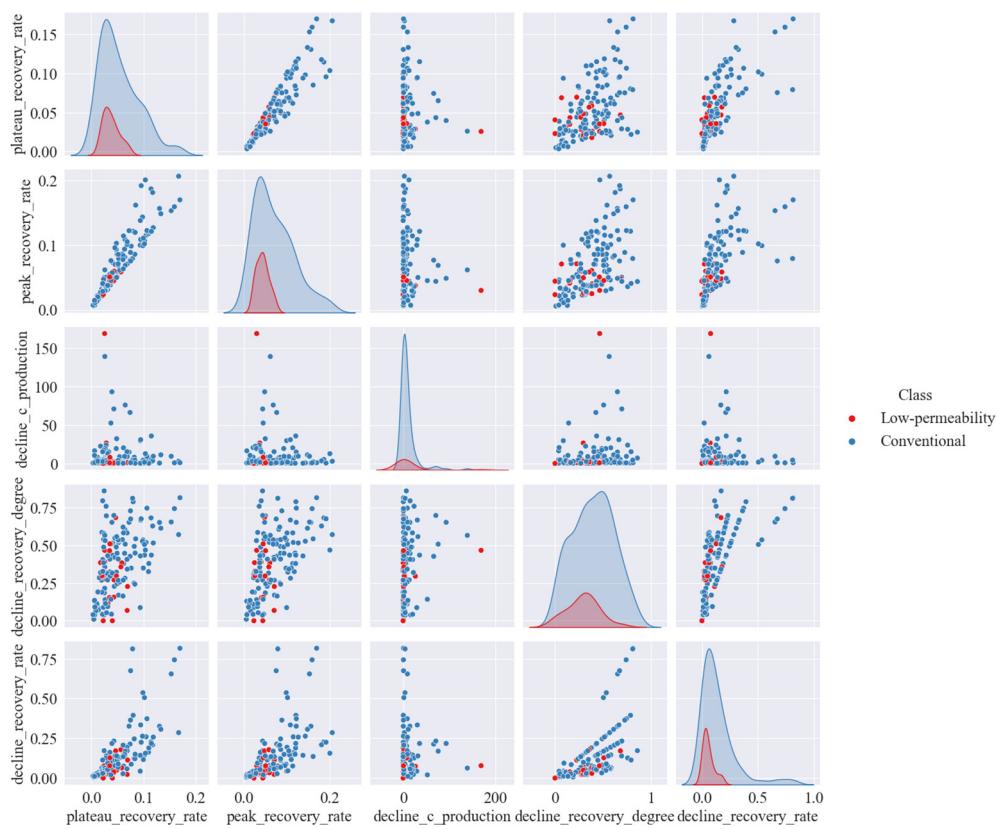


Figure 15. Pair plot of key production features of the dataset labeled by permeability.

4. Conclusions

This study delved into the onshore carbonate oilfield production features. Through the construction and evaluation of the XGBoost classifier using a comprehensive dataset from 377 onshore carbonate oilfields worldwide, key production features were selected and analyzed, which can support reservoir management and optimize production strategies. The main research findings of this study are as follows:

1. Onshore carbonate reservoirs with different reserve sizes show significant differences in build-up stage cumulative production, plateau stage cumulative production, plateau stage recovery rate, plateau stage recovery degree, and peak production. The production features of onshore carbonate reservoirs follow this pattern: oilfield reserve size, build-up stage cumulative production, plateau stage cumulative production, and peak production increase, while plateau stage recovery rate decreases, and the plateau stage recovery degree of small-sized oilfields is slightly greater than that of moderate and large oilfields.
2. The research methodology of this study can serve as a reference for studying the production features of other types of oil and gas reservoirs. The appliance of this methodology in low-permeability onshore carbonate oilfields shows that low-permeability oilfields generally have lower peak recovery rate, lower plateau stage recovery rate, lower decline stage recovery degree, and lower decline stage recovery rate, along with a wide but generally lower range of decline stage cumulative production compared to conventional oilfields.
3. The analysis regarding the differences in production features provides valuable insights for reservoir management. For example, as build-up stage cumulative production, plateau stage cumulative production, and peak production tend to increase with the size of the oilfield, this suggests that larger reserves have the potential for higher overall production. This insight can influence decisions related to resource allocation, investment prioritization, and long-term planning.

4. Future research in carbonate production could focus on incorporating advanced machine learning techniques like deep learning and ensemble methods to enhance classification accuracy, expanding dataset such as reservoir properties, number of wells, injection efficiency, aquifer, and tectonic activity, and conducting temporal analysis for better production analysis.

Author Contributions: Methodology, G.Q.; software, G.Q.; formal analysis, G.Q.; data curation, B.L.; writing—original draft, G.Q.; writing—review and editing, G.Q.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (No. 52174019), Open Fund of Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education (NO PI2021-06), and Educational Commission of Hubei Province of China (D20201302).

Data Availability Statement: The data presented in this study is available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahr, W.M. *Geology of Carbonate Reservoirs: The Identification, Description and Characterization of Hydrocarbon Reservoirs in Carbonate Rocks*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
2. Roehl, P.O.; Choquette, P.W. Perspectives on world-class carbonate Petroleum reservoirs. *AAPG Bull.* **1985**, *69*, 148.
3. Roehl, P.O.; Choquette, P.W. (Eds.) Introduction. In *Carbonate Petroleum Reservoirs*; Casebooks in Earth Sciences; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 1985; pp. 1–15.
4. Barwise, J.H.; McPherson, J.G.; Studlick, J.R. (Eds.) *Carbonate Petroleum Reservoirs*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
5. Moore, C.H.; Wade, W.J. *Carbonate Reservoirs: Porosity and Diagenesis in a Sequence Stratigraphic Framework*; Newnes: London, UK, 2013.
6. Matsumoto, K.I.; Voudouris, V.; Stasinopoulos, D.; Rigby, R.; Di Maio, C. Exploring crude oil production and export capacity of the OPEC Middle East countries. *Energy Policy* **2012**, *48*, 820–828. [[CrossRef](#)]
7. Wilson, J.L. Limestone and dolomite reservoirs. *Pet. Geol.* **1980**, *2*, 5969468.
8. Wilson, J.L. A review of carbonate reservoirs. In Proceedings of the Facts and Principles of World Petroleum Occurrence, Calgary, AB, Canada, 26–28 June 1978.
9. Li, Y.; Kang, Z.; Xue, Z.; Zheng, S. Theories and practices of carbonate reservoirs development in China. *Pet. Explor. Dev.* **2018**, *45*, 712–722. [[CrossRef](#)]
10. Dominguez, G.C. *Carbonate Reservoir Characterization: A Geologic-Engineering Analysis, Part I*; Elsevier: Amsterdam, The Netherlands, 1992.
11. Nairn, A.E.M.; Alsharhan, A.S. *Sedimentary Basins and Petroleum Geology of the Middle East*; Elsevier: Amsterdam, The Netherlands, 1997.
12. Lü, X.-X. Distribution patterns of oil-gas fields in the carbonate rock. *Acta Pet. Sin.* **2000**, *21*, 8.
13. Zhang, X.-M. The characteristics of Lower Ordovician fissure-vug carbonate oil and gas pools in Tahe oil field, Xinjiang. *Pet. Explor. Dev.* **2021**, *28*, 17–22.
14. Fan, J.-S. Characteristics of carbonate reservoirs for oil and gas fields in the world and essential controlling factors for their formation. *Dixue Qianyuan (Earth Sci. Front.)* **2005**, *12*, 23–30.
15. Luo, P.; Zhang, J.; Liu, W. Characteristics of marine carbonate hydrocarbon reservoirs in China. *Earth Sci. Front.* **2008**, *15*, 36–50.
16. Camacho Velázquez, R.; Fuentes-Cruz, G.; Vásquez-Cruz, M. Decline-curve analysis of fractured reservoirs with fractal geometry. *SPE Reserv. Eval. Eng.* **2008**, *11*, 606–619. [[CrossRef](#)]
17. Lü, A.; Yao, J.; Guo, Z. Study on typical relative permeability relationship and water displacement curves in fracture-cave reservoir of Tahe Ordovician. *Pet. Geol. Recovery Effic.* **2010**, *17*, 101–104.
18. Chang, B.; Wei, X.; Gao, S. Variation rule of water cut in large-scale fractured-cavernous reservoir. *Pet. Geol. Recovery Effic.* **2011**, *18*, 80–82.
19. Zhang, N.; He, D.; Sun, Y.; Li, H. Distribution patterns and controlling factors of giant carbonate rock oil and gas fields worldwide. *China Pet. Explor.* **2014**, *19*, 54.
20. Cheng, H. *Research on Indication Significance of Production Dynamic Curve of Fracture Vuggy Carbonate Reservoir*; Chengdu University of Technology: Chengdu, China, 2020.
21. Xiong, J.B.; He, D.F. Distribution characteristics and controlling factors of global giant carbonate stratigraphic-lithologic oil and gas fields. *Lithol. Reserv.* **2022**, *34*, 187–200.

22. Ji, B.; Xu, T.; Gao, X.; Yu, H.; Liu, H. Production evolution patterns and development stage division of waterflooding oilfields. *Pet. Explor. Dev.* **2023**, *50*, 433–441. [[CrossRef](#)]
23. Höök, M.; Söderbergh, B.; Jakobsson, K.; Aleklett, K. The evolution of giant oil field production behavior. *Nat. Resour. Res.* **2009**, *18*, 39–56. [[CrossRef](#)]
24. Mikael, H.; Hirsch, R.; Aleklett, K. Giant oil field decline rates and their influence on world oil production. *Energy Policy* **2009**, *37*, 2262–2272.
25. Aleklett, K.; Höök, M.; Jakobsson, K.; Lardelli, M.; Snowden, S.; Söderbergh, B. The peak of the oil age—analyzing the world oil production reference scenario in world energy outlook 2008. *Energy Policy* **2010**, *38*, 1398–1414. [[CrossRef](#)]
26. Hirsch, R.L. Mitigation of maximum world oil production: Shortage scenarios. *Energy Policy* **2008**, *36*, 881–889. [[CrossRef](#)]
27. Arps, J.J. Analysis of decline curves. *Trans. AIME* **1945**, *160*, 228–247. [[CrossRef](#)]
28. Guerrero, J.I.; Monedero, I.; Biscarri, F.; Biscarri, J.; Millán, R.; León, C. Detection of non-technical losses: The project MIDAS. In *Advances in Secure Computing, Internet Services, and Applications*; IGI Global: Hershey, PA, USA, 2014; pp. 140–164.
29. Kuang, L.; Liu, H.; Ren, Y.; Luo, K.; Shi, M.; Su, J.; Li, X. Application and development trend of artificial intelligence in petroleum exploration and development. *Pet. Explor. Dev.* **2021**, *48*, 1–14. [[CrossRef](#)]
30. Alkinani, H.H.; Al-Hameedi, A.T.; Dunn-Norman, S.; Flori, R.E.; Alsaba, M.T.; Amer, A.S. Applications of artificial neural networks in the petroleum industry: A review. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 18–21 March 2019.
31. Otchere, D.A.; Ganat, T.O.A.; Gholami, R.; Ridha, S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *J. Pet. Sci. Eng.* **2021**, *200*, 108182. [[CrossRef](#)]
32. Guo, S.; Jiang, W.; Meng, Q.; Ma, Y.; Wang, Q.; Xu, J.; Yan, H. Exploration on the application of data mining technology in complex oil & gas reservoirs evaluation. *Mud Logging Eng.* **2019**, *30*, 1–7.
33. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2012.
34. Draper, N.R.; Smith, H. *Applied Regression Analysis*; Wiley: Hoboken, NJ, USA, 1998.
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009.
36. Yang, B.; Kuang, L.-C.; Sun, Z.-C. On support vector machines method to identify oil & gas zone with logging and mudlog information. *Well Logging Technol.* **2005**, *29*, 511.
37. Liu, D.; Ran, Q.; Wang, B. Application of support vector machine in logging analysis of Qijia sag in Daqing Oilfield. *Geophys. Prospect. Pet.* **2007**, *46*, 156.
38. Zhang, Y.D.; Tong, K.; Zheng, J.; Wang, D. Application of support vector machine method for identifying fluid in low-resistivity oil layers. *Geophys. Prospect. Pet.* **2008**, *47*, 306–310.
39. Ghiasi, M.; Mohammad, G.; Lakhani, V.H.; Mohammadi, A.H. Distinct methodologies to assess the conditions of petroleum reservoirs with respect to onset of sand production. *Pet. Coal* **2019**, *61*, 339–350.
40. Somi, S.; Jubair, S.; Cooper, D.; Wang, P. XGSleeve: Detecting sleeve incidents in well completion by using XGBoost classifier. *Front. Artif. Intell.* **2023**, *13*, 1243584. [[CrossRef](#)]
41. Khaled, M.S.; Wang, N.; Ashok, P.; van Oort, E.; Wisian, K. Real-Time Prediction of Bottom-hole Circulating Temperature in Geothermal Wells Using Machine Learning Models. *Geoenergy Sci. Eng.* **2024**, *238*, 212891. [[CrossRef](#)]
42. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
43. Mitchell, R.; Frank, E. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci.* **2017**, *3*, e127. [[CrossRef](#)]
44. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
45. Landgrebe, T.C.W.; Duin, R.P.W. Efficient multiclass ROC approximation by decomposition via confusion matrix perturbation analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 810–822. [[CrossRef](#)] [[PubMed](#)]
46. Jin, Z. Distribution and structures of large and medium oil-gas fields in China. *Xinjiang Pet. Geol.* **2008**, *29*, 385.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.