

Research Article

Casing Damage Prediction Model Based on the Data-Driven Method

Chaodong Tan,¹ Wei Yan ,¹ Qing Tang,² Hua Wu,¹ Hongguang Bu,² Said Juma Kambi,¹ and Jiankang Liu¹

¹State Key Laboratory of Petroleum Resource and Prospecting, China University of Petroleum, Beijing, China

²PetroChina Dagang Oilfield Production Technology Research Institute, Tianjin, China

Correspondence should be addressed to Wei Yan; yanwei289@126.com

Received 18 March 2020; Revised 24 June 2020; Accepted 6 August 2020; Published 26 August 2020

Academic Editor: Nhon Nguyen-Thanh

Copyright © 2020 Chaodong Tan 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.

Casing damage caused by sand production in unconsolidated sandstone reservoirs often results in oil wells unable to produce normally. However, due to the complex mechanism of sheath damage caused by sand production, there is no more mature technology for predicting the risk of casing damage in advance. Data-driven method can better integrate various factors and use a large amount of historical data to solve complex classification prediction problems. In this paper, XGBoost and LightGBM algorithms are used to establish casing damage prediction models, and 13 model application experiments are carried out to optimize the set of casing damage factors. These two algorithms are used to calculate the feature importance of each factor and determine the final set of factors. The evaluation results of five key metrics show that both prediction models show good performance, and the prediction accuracy is 0.99 for the XGBoost model and 0.94 for the LightGBM model. Applying the established prediction model can determine reasonable range of the maximum daily liquid production of a single layer (Qlmax) to reduce the probability of casing damage. In addition, at certain Qlmax, increasing the perforation density can significantly reduce the probability of casing damage. Therefore, increasing the perforation density can achieve high production without causing casing damage.

1. Introduction

Casing damage is a common problem in sand production reservoirs. Preidentifying the risk of casing damage and taking effective preventive measures can reduce the probability of casing damage in oil wells. Some scholars analyze wellbore integrity by establishing the mechanism model. Yin and Gao [1] established a mechanical model of casing in directional well under in situ stresses and inside hydrostatic pressure. Lin et al. [2] established two theoretical models for casing strength degradation due to wear and corrosion to analyze the effects of relevant parameters on residual strength of defective casing. De Simone et al. [3] developed an analytical solution to assess the wellbore stresses and integrity during the drilling, construction, and production phases, which considered plane strain conditions, continuous homogeneous isotropic media, and linear elastic

materials. Yin et al. [4] established a finite element model to analyze nonlinear mechanical behavior of casing crossing slip formation, and the analysis results can provide a reference for casing design. Deng et al. [5] studied the collapse failure mechanism of cemented casing under nonuniform load and obtained the stress-strain laws of cemented casing by the electrical method. Han et al. [6] proposed that the incongruous displacement in mudstone is the main mechanism of casing deformation and developed a fluid-solid coupling seepage mechanical model to capture the flow and deformation in interlayers. Mohammed et al. [7] summarized the existing tools used for the assessment of well integrity issues and their respective limitations.

Some scholars use the data-driven method to solve the problems in oilfield production. Noshi et al. [8] used nine unsupervised algorithms, including support vector machine, Bootstrap, random forest, and some other algorithms to

identify the characteristics of casing failure during drilling and fracturing. Noshi et al. [9] applied artificial neural networks and boosted ensemble trees to build the prediction model of casing failure probability, which was used to evaluate 26 different features compiled from drilling, fracturing, and geologic data. Song and Zhou [10] selected 10 parameters that have the greatest impact on casing damage, including sand layer, casing, and perforation information and established the algorithm model of casing damage risk assessment by using GBDT, and the prediction accuracy of the model could reach 86.3%. Tang et al. [11] selected 19 casing damage influencing factors and established the casing damage risk prediction model by using XGBoost and LightGBM algorithms. Tan et al. [12] established a data-driven casing damage prediction model and proposed that the main control factors of casing damage were perforation density and production pressure difference.

At present, it is difficult to predict casing damage by the mechanism method. Machine learning method has been used to solve many problems in the process of oil and gas field development. The factor set of the casing damage prediction model established by Qing T and Chaodong T includes the production pressure difference, but these data are difficult to obtain completely and accurately. Therefore, it is necessary to optimize the set of casing damage influencing factors to enhance the practicability of the casing damage prediction model.

2. Influencing Factors of Casing Damage

This paper analyzes data integrity and accuracy of influencing factors of casing damage in the existing database of Gangxi oilfield and then analyzes the characteristics of each factor. Based on the original set of influencing factors of casing damage [11, 12], some parameters were deleted and some other comprehensive parameters were added to form a new casing damage dataset with 23 influencing factors, as shown in Table 1.

Production pressure difference of a small proportion of oil wells can be calculated by using the existing data in the database; this makes the calculation results uncertain, so the production pressure difference is omitted in the set of casing damage influencing factors. Since the outer diameter of casing in most wells of Gangxi oilfield is 139.7 mm, the mechanism of casing damage is different when this parameter is different. It is generally difficult to obtain a better understanding with fewer samples, so this article only focuses on the well with the casing outer diameter of 139.7 mm.

Because some of the geological parameters are difficult to obtain and the physical parameters of adjacent wells are generally close, the horizontal coordinates and vertical coordinates of the wellhead are used to approximately characterize these parameters. During oil well production, fluid carries formation sand into the wellbore. As the amount of sand produced increases, the formation around the casing in the production section is hollowed out, causing the casing to lose its lateral restraint. The overlying stratigraphic pressure on the formation acts on the casing axially, which causes the

load on the casing shaft to increase dramatically. When the critical bending load of the casing in this section is exceeded, the casing will bend, leading to the casing damage [13].

Therefore, in this paper, the maximum daily liquid production of a single layer is used as one of the casing damage influencing factors. In addition, to illustrate the casing by prolonged production of fluids, an additional arbitrage factor, i.e., cumulative liquid production of a single layer, was added. Due to differences in physical properties and perforation parameters between wells, two comprehensive parameters were added, namely, Q_{\max}/KH and Q_{\max}/HN . Q_{\max}/KH contains the physical properties and output information of a single layer, which expresses the supply and procurement balance relationship to some extent. Q_{\max}/HN represents the maximum flow rate per hole. When the flow rate per hole exceeds the critical sand carrying flow rate, the formation sand will be carried into the wellbore. Q_{cum} represents the cumulative liquid production of the single layer, reflecting the damage to the casing caused by long-term fluid production. In Table 1, Q_{\max} stands for the maximum daily liquid production of a single layer. K , H , and N represent the permeability, the perforation thickness, and the perforation density of the sand layer.

3. Casing Damage Probability Prediction

The workflow of casing damage probability prediction is shown in Figure 1. Firstly, the required data in Table 1 are extracted from the existing database to form a sample dataset of casing damage. Secondly, a casing damage probability prediction model is established based on the XGBoost and LightGBM algorithms based on the above dataset. Thirdly, the established models are used to carry out multiple data experiments, compare the prediction results of these experiments, and determine the influencing factors to be eliminated. Fourthly, the feature importance of the remaining impact factors is calculated and analyzed. Fifthly, the less important impact factor is deleted to optimize the input parameters of the model. Finally, the optimized model is used to predict the casing damage probability of the oil well.

3.1. Data Preparation and Processing. The data used in this paper include the geological parameters, engineering parameters, production data, and casing damage information of 244 production layers in 133 wells in Gangxi oilfield. Among them, the casings of 68 production layers in 64 wells were damaged. The casing damage rate of the production layers was 27.9%. The required data in Table 1 are extracted from the existing database of Gangxi oilfield to form a sample set of casing damage. In this sample set, casing steel grade and oil reservoir group are text type, and different numbers are used to represent different types. Numbers 1, 2, 3, and 4 are used to indicate the casings with steel grades of J55, K55, N80, and P110, respectively. Numbers 1 and 2 are used to indicate Ming II and Ming III reservoir groups, respectively. In addition, missing data were completely added, and incorrect data were revised. The data overview of these 244 production layers is shown in Table 2.

TABLE 1: Influencing factors of casing damage in production wells.

1	2	3	4	5	6
Wellhead abscissa	Wellhead ordinate	Perforation top	Perforation bottom	Perforation thickness	Perforation density
7	8	9	10	11	12
Perforator phasing	Casing wall thickness	Casing steel grade	Sand layer top	Sand layer bottom	Sand layer thickness
13	14	15	16	17	18
Reservoir group	Permeability	Porosity	Maximum daily liquid production of the single layer	Maximum water cut	Maximum fluid production intensity
19	20	21	22	23	
Qlcum	Qlmax/KH	Qlmax/HN	Casing service time	Sand layer production time	

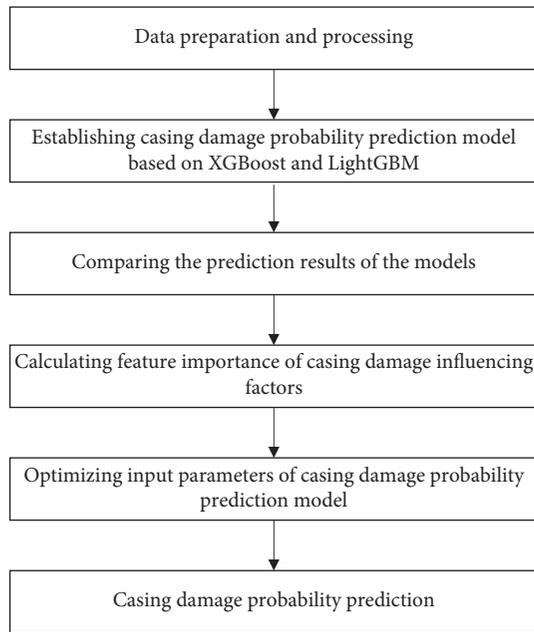


FIGURE 1: Workflow of casing damage probability prediction.

3.2. *Casing Damage Probability Prediction Model.* Casing damage prediction is a classification problem; it could be predicted whether the casing is going to be damaged or not. XGBoost and LightGBM are the two most popular algorithms that are often used to solve classification problems [11, 12]. The two algorithms have got various advantages over others. Therefore, in this paper, the two mentioned algorithms are used to establish the casing damage prediction model. In the prediction model, 23 influencing factors in Table 2 are taken as input parameters, and casing damage probability is used as the output parameter, as shown in Figure 2. In the sample data, number 1 means that the casing has been damaged, and number 0 means that the casing has not been damaged. The split function in Python is used to extract 80% of the sample data as the training dataset and 20% as the testing dataset.

3.3. *Prediction Model Application Tests.* Taking different influencing factor combinations as the input of the model, applying the established prediction model to carry out data tests, comparing the prediction results of these tests, and determining which influencing factors to eliminate, thirteen data tests were carried out, and the parameters eliminated from each test and prediction results are shown in Table 3.

Comparing the prediction results of the training dataset and the testing dataset of the 13 tests in Table 3, it is found that the prediction effects of test7 and test8 are both the best. In the training set, the XGBoost model has higher prediction accuracy than the LightGBM model. In the testing dataset, the prediction accuracy of the XGBoost model and the LightGBM model is the same, and the number of casing damaged wells predicted by the XGBoost model is closer to reality.

Perforation bottom and sand layer bottom were eliminated in these two experiments, so these two parameters are excluded from the sample data. Test7 deleted max fluid production intensity, and test8 deleted Qlcum; test13 excluded these two parameters at the same time. The prediction result of test13 is not as good as test7 and test8. In the following, we will discuss which of these two parameters to exclude.

3.4. *Feature Importance Analysis.* Figures 3 and 4 show the calculation results of feature importance using the XGBoost and the LightGBM algorithms, respectively. The parameters represented by the symbols of the ordinate are shown in Table 2. It can be seen that, in the results of the two algorithms, max fluid production intensity is less important than Qlcum. Therefore, in the final prediction sample, max fluid production intensity is eliminated. Therefore, the input parameters of the prediction model use the 20 factors used in test8.

3.5. *Prediction Model Evaluation.* The 20 casing damage influencing factors identified above are used as input parameters, and casing damage probability is used as the output parameter. XGBoost and LightGBM models are used to predict the casing damage probability of 49 production layers (testing dataset). The prediction results are shown in Figure 5.

TABLE 2: Data overview of 244 production layer samples.

Symbol	Influencing factors	Unit	Minimum	Maximum	Average
f0	Wellhead abscissa	M	—	—	—
f1	Wellhead ordinate	M	—	—	—
f2	Perforation top	M	827	1408.7	1067.1
f3	Perforation bottom	M	833.8	1412.4	1071.1
f4	Perforation thickness	M	1.0	14.0	3.9
f5	Perforation density	hole/m	10	32	—
f6	Perforator phasing	degree	90	135	—
f7	Casing wall thickness	mm	6.2	9.17	—
f8	Casing steel grade		J55, K55, N80, P110		
f9	Sand layer top	M	826.9	1408.7	1066.6
f10	Sand layer bottom	M	833.8	1412.4	1072.6
f11	Sand layer thickness	M	1	19.2	5.9
f12	Reservoir group		Ming II, Ming III		
f13	Permeability	mD	162.9	5431.1	1415.7
f14	Porosity	%	20.2	72.1	34.4
f15	Qlmax	m ³ /d	1.2	359.7	27.9
f16	Max water cut	%	28.7	100	90.9
f17	Max fluid production intensity	m ³ /m	0.31	67.36	7.82
f18	Qlcum	m ³	19.6	2,123,576	59,579
f19	Qlmax/KH	1000m ³ /(d·mD·m)	0.14	125.08	7.01
f20	Qlmax/HN	m ³ /d/hole	0.02	5.18	0.55
f21	Casing service time	Annual	0.48	50.34	15.16
f22	Sand layer production time	Annual	0.18	41.79	10.7

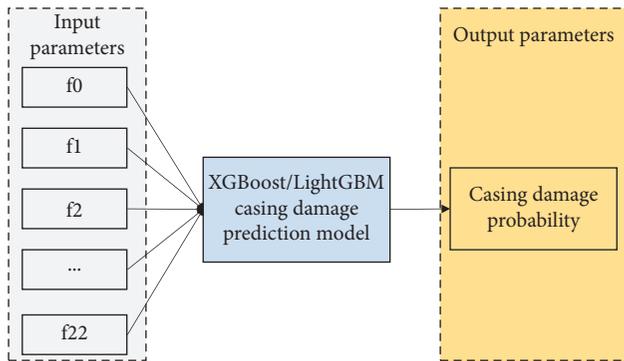


FIGURE 2: Casing damage probability prediction model.

In the testing dataset, there are actually 12 production layers with casing damage and 37 production layers with undamaged casing. Among the 12 casing damaged samples, XGBoost model and LightGBM model correctly predicted 11 and 10, respectively. In Figure 5, the 1st sample is a casing damaged sample, but it is incorrectly predicted by both models as undamaged. The 48th sample is also a casing damaged sample, which is correctly predicted by the XGBoost model but incorrectly predicted by the LightGBM model. Among the 37 undamaged samples, XGBoost model and LightGBM model correctly predicted 35 and 36, respectively. The 26th sample is an undamaged sample but is incorrectly predicted to be casing damaged by both models. The 37th sample is also an undamaged sample, which is correctly predicted by LightGBM and incorrectly predicted by the XGBoost model.

Five key metrics are used to evaluate the performance of the casing damage prediction model, including accuracy, precision, recall, F1-score, and AUC (area under ROC curve). The first four parameters can be expressed as

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{precision} = \frac{TP}{TP + FP},$$

$$\text{recall} = \frac{TP}{TP + FN},$$

$$\text{F1 - score} = \frac{2 \cdot \text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}.$$
(1)

If the casing in a production layer was actually damaged and was predicted to be damaged by the model, then it is true positive (TP). If the casing in a production layer was actually undamaged but was predicted to be damaged, then it is false positive (FP). If the casing in a production layer was actually undamaged and was predicted to be undamaged, then it is true negative (TN). Finally, if the casing in a production layer was actually damaged but was predicted to be undamaged, then it is false negative (FN).

The confusion matrix of casing damage prediction is shown in Figure 6. For the training dataset, the prediction results of the XGBoost model are all correct, while the LightGBM model predicts the 2 casing damaged layers as undamaged. For the testing dataset, the XGBoost model predicts one more casing damaged sample correctly than LightGBM but one less casing undamaged sample correctly than LightGBM.

TABLE 3: Model application test prediction results.

Project	Algorithm	Training set		Test set		Number of factors	Parameters removed
		Damaged	Undamaged	Damaged	Undamaged		
Actual	—	56	139	12	37	—	—
test1	XGBoost	56	139	11	35	23	—
	LightGBM	53	139	10	36		
test2	XGBoost	54	138	9	34	21	Casing service time Sand layer production time
	LightGBM	32	138	7	36		
test3	XGBoost	56	139	11	35	21	Perforation bottom Sand layer bottom
	LightGBM	53	139	10	36		
test4	XGBoost	56	139	10	35	20	Perforation bottom Sand layer bottom
	LightGBM	53	139	10	36		
test5	XGBoost	56	139	10	35	20	Perforator phasing Perforation bottom Sand layer bottom
	LightGBM	53	139	10	36		
test6	XGBoost	56	139	11	34	20	Qlmax Perforation bottom Sand layer bottom
	LightGBM	49	139	10	36		
test7	XGBoost	56	139	11	35	20	Perforation bottom Sand layer bottom Max fluid production intensity
	LightGBM	54	139	10	36		
test8	XGBoost	56	139	11	35	20	Perforation bottom Sand layer bottom Qlcum
	LightGBM	54	139	10	36		
test9	XGBoost	55	139	10	35	20	Perforation bottom Sand layer bottom Qlmax/KH
	LightGBM	54	139	10	36		
test10	XGBoost	56	139	11	35	20	Perforation bottom Sand layer bottom Qlmax/HN
	LightGBM	54	139	10	35		
test11	XGBoost	54	138	10	33	20	Perforation bottom Sand layer bottom Casing service time
	LightGBM	47	137	9	35		
test12	XGBoost	56	139	10	35	20	Perforation bottom Sand layer bottom Sand layer production time
	LightGBM	54	139	10	36		
test13	XGBoost	55	139	10	35	19	Perforation bottom Sand layer bottom Max fluid production intensity Qlcum
	LightGBM	54	139	10	35		

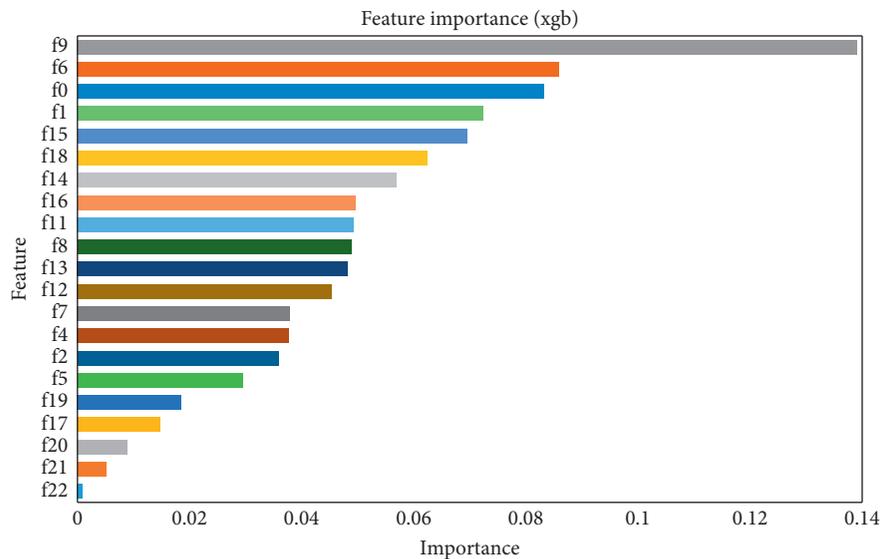


FIGURE 3: Feature importance based on the XGBoost algorithm.

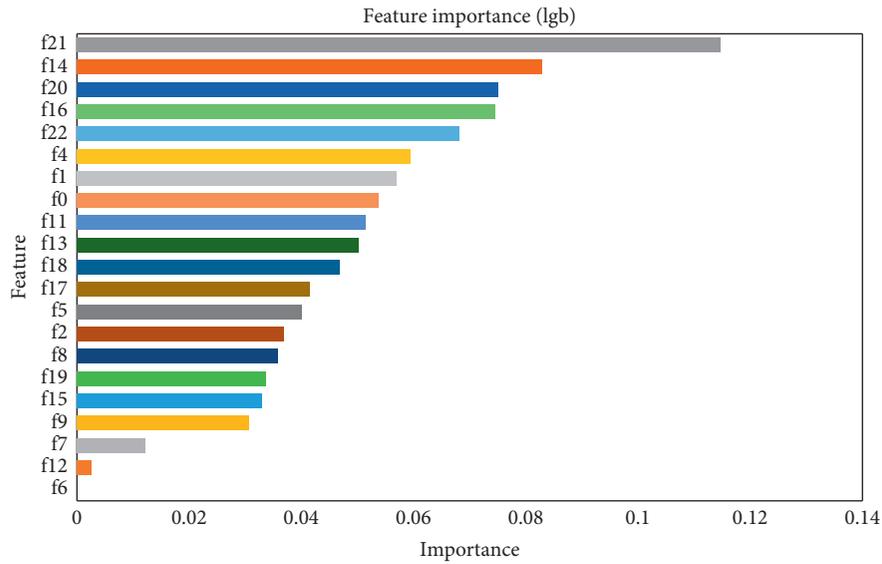


FIGURE 4: Feature importance based on the LightGBM algorithm.

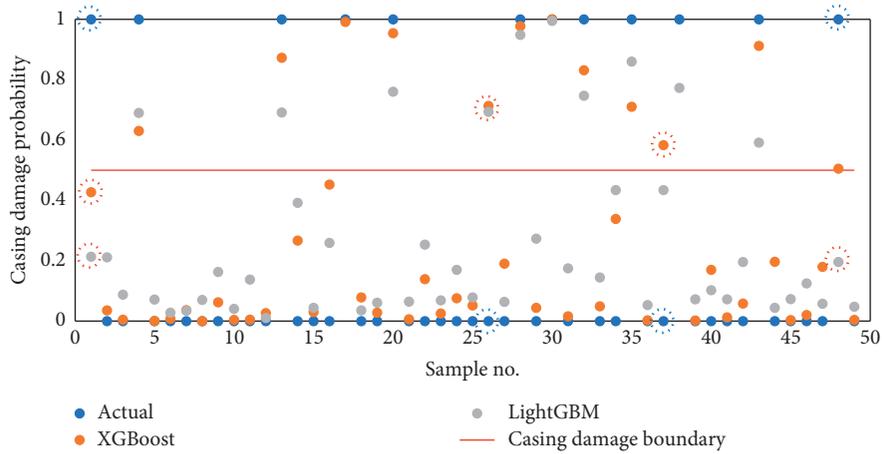


FIGURE 5: Casing damage probability prediction by XGBoost and LightGBM models.

		Actual for training dataset		Actual for testing dataset	
		Casing damaged	Casing undamaged	Casing damaged	Casing undamaged
XGBoost model predicted	Casing damaged	56	0	11	2
	Casing undamaged	0	139	1	35
LightGBM model predicted	Casing damaged	54	0	10	3
	Casing undamaged	2	139	2	36

FIGURE 6: Confusion matrix of casing damage prediction.

TABLE 4: Four key metrics for prediction model evaluation.

Sample	Model	Accuracy	Precision	Recall	F1-score
Training dataset	XGBoost	1	1	1	1
	LightGBM	0.99	1	0.96	0.98
Testing dataset	XGBoost	0.94	0.85	0.92	0.88
	LightGBM	0.94	0.91	0.83	0.87

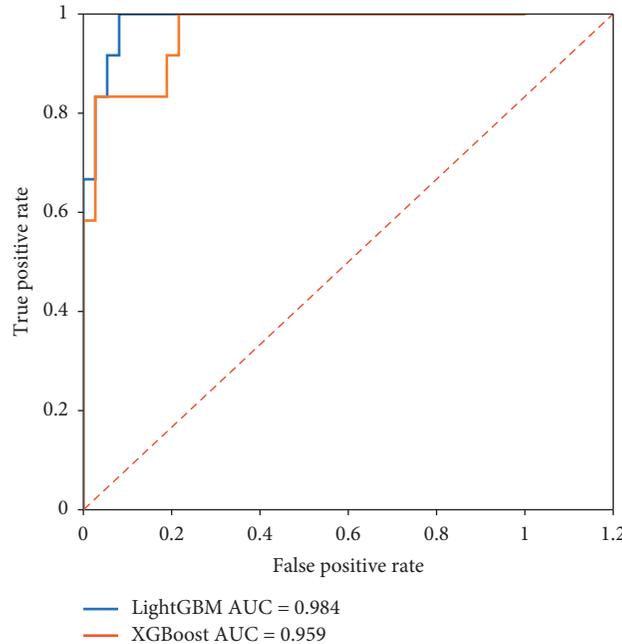


FIGURE 7: Area under the ROC curve of the testing dataset.

As can be seen from Table 4, accuracy, recall, and F1-score of the XGBoost model are higher than LightGBM, but precision is lower than LightGBM. The AUC curve is shown in Figure 7, the AUC of both models is very high, and the AUC of LightGBM is slightly higher.

In general, the prediction effects of both models are very good. In actual application, the prediction results of the two models can be combined for comprehensive evaluation. If the prediction results of the two models are both casing damaged, it reflects to a certain extent that the probability of casing damage in the target production layer is higher, and technicians should pay great attention to the well. If only one model predicts that the result is casing damaged, the technician needs to further analyze and evaluate based on the actual data on the oilfield.

4. Prediction Model Application

The casing damage probability prediction model established above is used to predict the casing damage probability of the current production layers. For the production layers with higher risk of casing damage, the controllable parameters are optimized to reduce the risk.

It can be seen from Figure 8 that, in the four production layers, when Qlmax increases, the casing damage probability of layer X27-18/1272.8~1277.5 m increases, but the predicted result is still below the casing damage limit. For the other three wells, when Qlmax increases, the probability of casing damage may exceed 0.5, which means that the casings of these layers may be damaged.

For production layers x8-9-2/962.3~965.5 m, x10-8-3/1052.5~1056.5 m, and x2-5-5/1054.1~1058.9 m, critical Qlmax of casing damage is 45, 58, and 63, respectively. It can be found that due to differences in physical properties, perforation parameters, and other parameters in the production layers between wells, critical casing damage Qlmax is also different. In actual production, Qlmax is controlled below the critical value to reduce the probability of casing damage.

It can be seen from Figure 9 that when Qlmax is lower than 80 m³/d, as the perforation density increases, the probability of casing damage is greatly reduced. Actual Qlmax of production layer X8-9-2/962.3~965.5 m is 48 m³/d, and its casing damage probability is higher than 0.5. The perforation density can be increased to reduce the probability of casing damage while maintaining a high production.

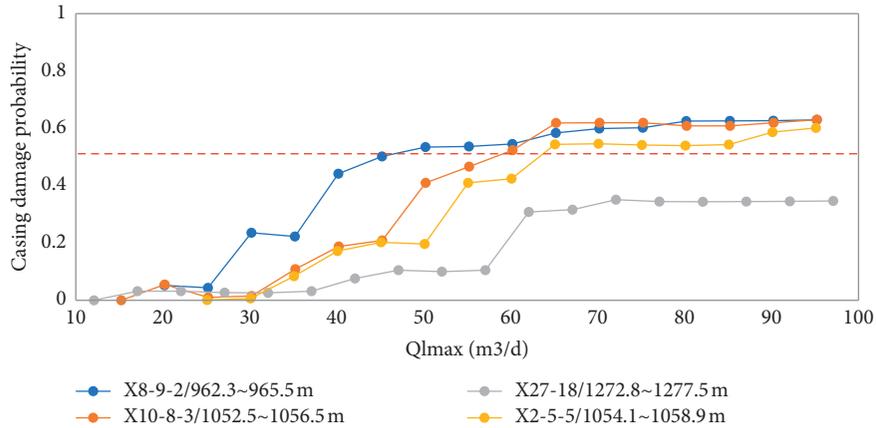


FIGURE 8: Casing damage probability of 4 production layers vs. Q_{lmax} .

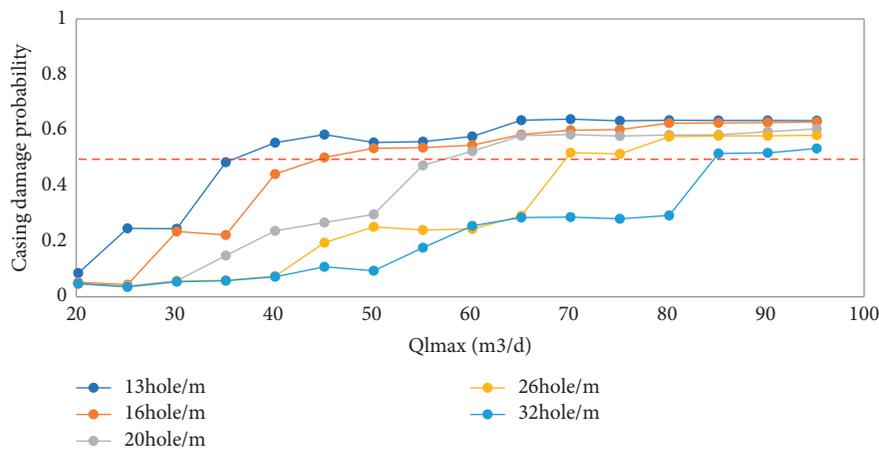


FIGURE 9: Casing damage probability of a production layer (X8-9-2/962.3~965.5 m) vs. different perforation densities.

If the perforation density is increased to 32, Q_{lmax} can be increased to $80 \text{ m}^3/\text{d}$.

5. Conclusions

- (1) XGBoost and LightGBM are used to establish the casing damage prediction model. The split function in Python is used to extract 80% of the sample data as the training dataset and 20% as the testing dataset. The evaluation results show that both XGBoost model and LightGBM model can achieve good prediction results, and the prediction accuracy is 0.99 and 0.94, respectively.
- (2) Accuracy, recall, and F1-score of the XGBoost model are higher than LightGBM, but precision is lower than LightGBM. The AUC of both models is very high, and the AUC of LightGBM is slightly higher. In general, the prediction effects of both models are very good.
- (3) The casing damage probability prediction model established above is used to predict the casing damage probability of the current production layers. In the four production layers, when Q_{lmax} increases,

the casing damage probability of layer X27-18/1272.8~1277.5 m increases. In actual production, Q_{lmax} is controlled below the critical value to reduce the probability of casing damage.

- (4) When Q_{lmax} is lower than $80 \text{ m}^3/\text{d}$, as the perforation density increases, the probability of casing damage is greatly reduced. The perforation density can be increased to reduce the probability of casing damage while maintaining a high production. If the perforation density is increased to 32, Q_{lmax} can be increased to $80 \text{ m}^3/\text{d}$.
- (5) Due to the differences in physical properties, perforation parameters, and other parameters in the production layers and wells, critical casing damage Q_{lmax} is also different. In practice, controlling Q_{lmax} below the critical value theoretically can reduce the probability of casing damage.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 51974327).

References

- [1] F. Yin and D. Gao, "Mechanical analysis and design of casing in directional well under in-situ stresses," *Journal of Natural Gas Science and Engineering*, vol. 20, pp. 285–291, 2014.
- [2] T. Lin, Q. Zhang, Z. Lian, X. Chang, K. Zhu, and Y. Liu, "Evaluation of casing integrity defects considering wear and corrosion - application to casing design," *Journal of Natural Gas Science and Engineering*, vol. 29, pp. 440–452, 2016.
- [3] M. De Simone, F. L. G. Pereira, and D. M. Roehl, "Analytical methodology for wellbore integrity assessment considering casing-cement-formation interaction," *International Journal of Rock Mechanics and Mining Sciences*, vol. 94, pp. 112–122, 2017.
- [4] F. Yin, Y. Deng, Y. He et al., "Mechanical behavior of casing crossing slip formation in waterflooding oilfields," *Journal of Petroleum Science and Engineering*, vol. 167, pp. 796–802, 2018.
- [5] K. Deng, W. Liu, T. Xia et al., "Experimental study the collapse failure mechanism of cemented casing under non-uniform load," *Engineering Failure Analysis*, vol. 73, pp. 1–10, 2017.
- [6] L. Han, F. Yin, S. Yang et al., "Coupled seepage-mechanical modeling to evaluate formation deformation and casing failure in waterflooding oilfields," *Journal of Petroleum Science and Engineering*, vol. 180, pp. 124–129, 2019.
- [7] A. I. Mohammed, B. Oyeneyin, B. Atchison et al., "Casing structural integrity and failure modes in a range of well types - a review," *Journal of Natural Gas Science and Engineering*, vol. 68, 2019.
- [8] C. Noshi, S. Noynaert, and J. Schubert, "Failure predictive analytics using data mining: how to predict unforeseen casing failures?" in *Proceedings of the Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, November 2018.
- [9] C. Noshi, S. Noynaert, and J. Schubert, "Data mining approaches for casing failure prediction and prevention," in *Proceedings of the International Petroleum Technology Conference*, Saudi Arabia, UAE, January 2019.
- [10] M. Song and X. Zhou, "A casing damage prediction method based on principal component analysis and gradient boosting decision tree algorithm//," in *Proceedings of the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, March 2019.
- [11] Q. Tang, H. Wu, G. Teng et al., "Prediction of casing damage in unconsolidated sandstone reservoirs using machine learning algorithms," in *Proceedings of the International Conference on Computation, Communication and Engineering*, pp. 185–188, Fujian, China, November 2019.
- [12] C. Tan, H. Wu, J. Liu et al., "A novel data mining approach in preventing casing damage of oil production wells," in *Proceedings of the Eurasia Conference on IOT, Communication and Engineering*, pp. 421–424, Yunlin, Taiwan, October 2019.
- [13] L. Ren, F. Wang, J. Liu et al., "Analysis of effect of sand production on casing damage in dagang Gangxi oilfield," *Oil Field Equipment*, vol. 39, no. 6, pp. 28–31, 2010.