

Research Article

Application of BigML in the Classification Evaluation of Top Coal Caving

Meng Wang , Caiwang Tai , Qiaofeng Zhang , Zongwei Yang , Jiazheng Li ,
Kejun Shen , and Kang Wang 

College of Mining Engineering, Liaoning Technical University, Fuxin 123000, China

Correspondence should be addressed to Caiwang Tai; 471920665@stu.lntu.edu.cn

Received 14 June 2021; Revised 5 August 2021; Accepted 17 August 2021; Published 2 September 2021

Academic Editor: Jianwei Cheng

Copyright © 2021 Meng Wang 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.

Longwall top coal caving mining is one of the main methods of mining thick coal seams in China. Therefore, carrying out the classification evaluation of top coal caving is of great significance to ensure mining success and reduce the risk of mining technology. In order to realize the classification evaluation of top coal caving, this article introduces the method of using BigML to establish the classification evaluation model of top coal caving. Furthermore, using the data from the CNKI database as sample data, a classification evaluation model of top coal caving is established on BigML. After training, testing, and optimization, the model is used to evaluate the top coal caving in No. 3 coal seam of Gucheng Coal Mine, and the evaluation result is grade 1, which is consistent with the engineering practice. The final research results show that the application of BigML in the classification evaluation of top coal caving is successful; the evaluation of top coal caving through BigML is reliable; BigML provides another scientific reliability way for the classification evaluation of top coal caving.

1. Introduction

According to the *World Energy Statistics Review* published in 2020, global coal consumption decreased by 0.6% in 2019, and the proportion of coal in primary energy reached the lowest level in 16 years, but the proportion of coal in primary energy is still up to 27% [1]. Therefore, in the continuous development of other energy sources, coal is still one of the most critical energy sources [2], especially for a country such as China that is “rich in coal, poor in oil, and less in gas” and whose economy is developing rapidly, the status of coal is vital. In 2015, China’s coal production and consumption had reached 47% and 50.01% of the world’s coal production and consumption [3], respectively. By 2019, when global coal consumption is declining, China’s coal consumption will still account for 51.7% of the world’s coal consumption [1]. It shows that a large amount of coal needs to be consumed to support China’s rapid economic development. Among the proven recoverable coal reserves in China, thick coal seams account for 44% of the total recoverable coal reserves [4], and coal seams with a thickness of more than 5 m are as

much as 10 billion tons [5], so nearly half of the coal consumed comes from the mining of thick coal seams [6]. As the longwall top coal caving mining technology has the advantages of high output, high efficiency, low energy consumption, low cost, and strong adaptability [7], it is estimated that the use of longwall top coal caving mining in thick seams can double the production efficiency and reduce the production cost by 30% to 40% [8]. Therefore, since China introduced it from abroad in the 1980s, it has become the primary technology for mining thick coal seam in China, Australia, Turkey, Russia, and Vietnam through continuous development and improvement [9–16]. The top coal caving (that is, the difficulty of top coal mining under the action of in situ stress and gravity [17]) is one of the critical factors determining the success of longwall top coal caving mining, and it is also an essential reference for designing longwall top coal caving mining technology and improving the recovery rate of top coal [18]. In addition, if the top coal caving can be understood from the feasibility study to the formal design stage, the relevant technical risks will be significantly reduced [19]. Therefore, the evaluation of top coal caving has

always been a hot spot in the research of longwall top coal caving mining at home and abroad.

At present, the evaluation of top coal caving at home and abroad is mainly based on observation and empirical methods [7, 19], established mathematical evaluation models [20], and numerical simulations [21–23]. However, there are some shortcomings in these methods. For example, based on observation and experience, there are strict requirements on the experience level of the evaluator, and at the same time, they also need to bear the massive risk of mistakes due to lack of experience. The establishment of mathematical models and numerical models to evaluate top coal caving requires that the model builder has a very high level of professional knowledge. At the same time, as the simulation size of the current numerical simulation is getting larger and larger, the requirements for the computer are getting higher and higher, which makes the technical cost and time cost of the evaluation very high, and the evaluation efficiency is not high. In recent years, with the continuous development of science and technology, advanced artificial intelligence and machine learning algorithms such as ant colony clustering algorithms, expert systems, and artificial neural networks have been gradually used in coal production [24–26]. Therefore, experts and scholars are also trying to use artificial intelligence algorithms. Moreover, machine learning algorithms are used to evaluate top coal caving and have achieved certain results. For example, Mohammadi et al. [27] used fuzzy multicriteria decision-making methods to establish a classification system for evaluating the caving of the direct roof of coal seams; Yongkui et al. [28] used Bayesian theory and rough set theory to establish a Bayesian classifier model used for the evaluation and prediction of roof caving properties of coal seams, which can accurately classify; Oraee and Rostami [29] used fuzzy logic algorithms to establish a fuzzy system for quantitative analysis of roof caving in longwall top coal caving mining face and applied the model to Tabas-Palward Mine's longwall top coal caving mining face which located in Palward District, Yazd Province, and the model prediction results in application are consistent with the on-site measured results; Shi et al. [17] established a top coal caving prediction model based on vector support vector machines, and the test results showed that the model has a certain feasibility and generalization; Yu and Mao [30] used SPSS statistical software to establish a top coal caving prediction model based on an artificial neural network. The training and test results show that the model has good top coal caving prediction capabilities. However, most of the prediction models of the above experts and scholars are built by computer programming, which requires higher computer language and professional level of users, especially when users want to optimize and modify the model to make it conform to the actual situation. Therefore, it leads to the poor portability of the model and the difficulty of popularization and application.

Therefore, in order to make the model have better portability, at the same time, without programming, it is easy to use the established model to evaluate the caving ability of top coal, and even easy to modify and optimize the established model to make it more in line with their practical

application. In this article, BigML is applied to establish the classification evaluation model of top coal caving. It is hoped that through this, it is possible to easily use the established model to evaluate the top coal caving without programming, and even easily modify and optimize the established model to make it more in line with their actual application situation.

2. Introduction to BigML

BigML (<https://bigml.com>) is a cloud-based machine learning platform dedicated to enabling all people who understand computer language and do not understand computer language to build their machine learning prediction model without writing a line of code and to make it automatic, remote callable, programmable, and extensible. It can perfectly and easily solve modeling tasks such as classification, regression, time series forecasting, cluster analysis, anomaly detection, correlation analysis, and realize model visualization. Because BigML has powerful functions and advanced algorithms and it is easy to use, it can realize a one-stop service from data loading, data cleaning, model building, and model evaluation to the final model prediction. So, BigML has been widely used in agriculture [31], medicine [32], finance [33], and other scientific research fields and has successfully helped thousands of analysts, software developers, and scientists all over the world solve machine learning tasks from “end-to-end” and seamlessly transform data into operational models for remote services, or embed them locally into applications for prediction. In addition, BigML has more than 147000 global users, and more than 600 universities and research institutions have cooperated with it. The global user distribution of BigML is shown in Figure 1.

3. Influencing Factors of Top Coal Caving and Its Evaluation Grade Division

3.1. Influencing Factors of Top Coal Caving. Two main factors are affecting top coal caving: geological and mining technology factors. In the longwall top coal caving mining, the mining technology is generally determined by geological factors [34], so this article mainly studies the top coal caving under the influence of geological factors. The practical experience [30] and research [9] show that the buried depth (H), thickness of coal seam (M), thickness of gangue (MG), uniaxial compressive strength of coal (Rc), fracture development degree (DN , that is, the product of the number of through cracks N_{1m} on the surface of 1 m^2 coal and the fractal dimension D_1 of the number of cracks counted by the coal sample), and filling coefficient of direct roof (K , $K = \sum hk_p/M$) are essential geological factors affecting top coal caving [35]. Therefore, this article considers the factors above influencing factors for the classification evaluation of top coal caving.

3.2. Top Coal Caving Evaluation Grade Division. The top coal caving is generally divided into five grades, namely, “(1) excellent caving,” “(2) good caving,” “(3) fair caving,” “(4) poor caving,” and “(5) extremely poor caving” [18].

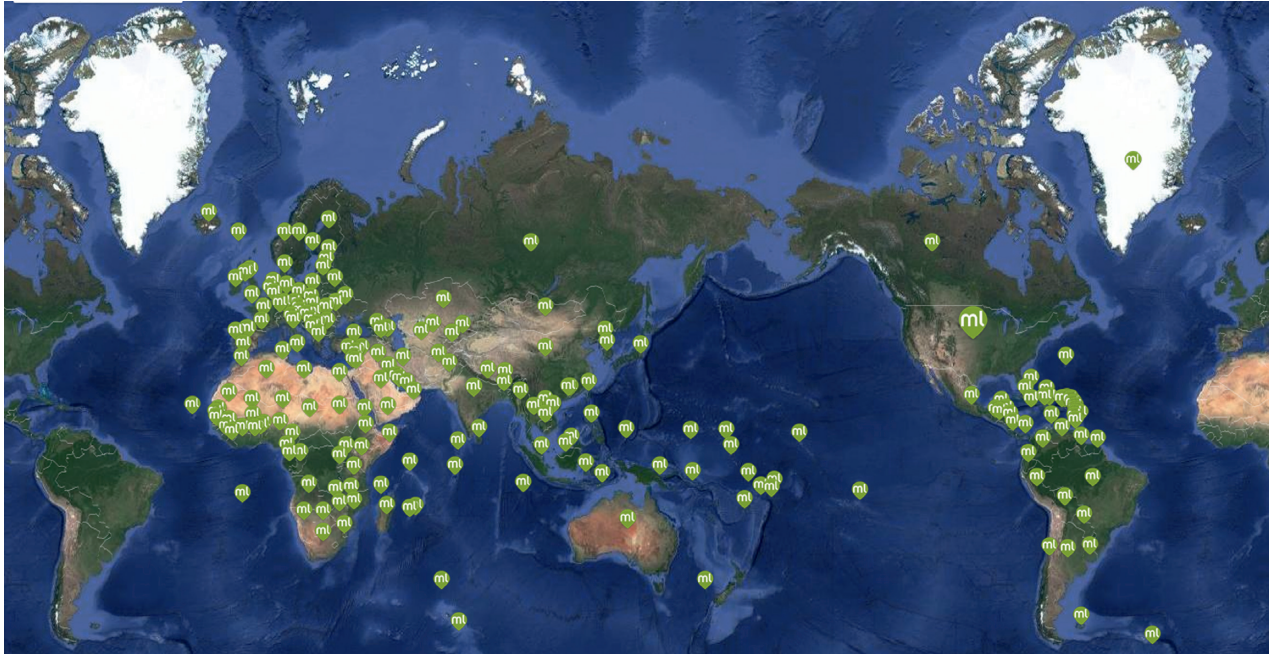


FIGURE 1: Global user distribution map of BigML.

However, the engineering practice shows little difference between grades “4” and “5.” Both are difficult to cave. Therefore, in this article, the top coal caving is only divided into four grades, which are “(1) excellent caving,” “(2) good caving,” “(3) fair caving,” and “(4) poor caving.” The specific situation of each grade is shown in Table 1.

4. Sample Data and Data Preprocessing

4.1. Source of Sample Data. There are a large number of research studies on top coal caving in the CNKI publicly published paper database. Therefore, to meet the needs of model training and testing, this article obtained 68 sets of sample data from the CNKI publicly published paper database. The specific conditions of the sample data are shown in Table 2, and the data distribution is shown in Figure 2. Through the shape and volume of the violin in the data distribution diagram (Figure 2), it can be seen that the “grade” of the obtained sample data is unbalanced. Therefore, to ensure that the established model has good stability and strong prediction ability, the sample imbalance is a problem worthy of attention, and it is necessary to reasonably select model performance evaluation indicators to evaluate the model.

4.2. Data Cleaning and Segmentation. Although all the sample data are from CNKI’s publicly published paper database, outliers are inevitable in these sample data. The quality of the data sample generally has an essential impact on establishing the model and the predictive ability of the established model. Therefore, cleaning the data sample and eliminating outliers make the established model better predictive. BigML’s data anomaly detection function is based on the most advanced isolated forest algorithm [36], which

has a powerful ability of outlier detection. Therefore, this article uses BigML’s data anomaly detection function to do outlier detection on the data samples in this article. Before BigML performs data anomaly detection, it is set to find at least three sets of relatively abnormal sample data. After testing, only the sample with the coal seam name “No. 8-1 coal seam of Baode Mine” is an outlier (in BigML, it is usually considered that the score is more than 60%, which is the actual abnormality [37]). The BigML anomaly detection result is shown in Figure 3. Therefore, “No. 8-1 coal seam of Baode Mine” is removed from the sample data set to obtain clean sample data. The sample data distribution after cleaning is shown in Figure 4.

In order to train the model and test the model’s performance, using the data segmentation function of BigML, the sample data are randomly divided according to 7 : 3, 70% of which is the training set and 30% is the test set. The sample data distribution of the training set and test set after segmentation is shown in Figure 5.

5. Selection of Model Performance Evaluation Indicators

Top coal caving classification evaluation is a problem of classification and prediction. In classification prediction model training and testing, model performance evaluation indicators play a vital role in obtaining the optimal classifier. Therefore, choosing appropriate model performance evaluation indicators is essential to identifying and obtaining the optimal classifier [38]. Most of the performance indicators of the current evaluation and classification prediction model are based on the confusion matrix similar to Table 3, and these performance indicators can be divided into three categories.

TABLE 1: Caving grade of top coal.

Grade	Top coal caving	Description
1	Excellent	The top coal can cave very well. As long as the appropriate caving support is selected, the coal can be discharged, and no extra measures need to be taken.
2	Good	The top coal can also cave well. Similarly, the coal can be discharged after selecting the appropriate caving support, but there are large blocks in the discharged coal, which is easy to cause the phenomenon of the bayonet, so corresponding measures need to be taken.
3	Fair	The top coal can cave, but not well. At the same time, the discharged coal is large and often has a bayonet phenomenon. Corresponding measures must be taken to discharge the coal.
4	Poor	The top coal is very difficult to cave, and more measures are needed to release the coal.

TABLE 2: Sample data.

Number	Coal seam name	H (m)	R_c (MPa)	MG (m)	M (m)	DN	K	Grade
1	No. 3 coal seam of Yaoqiao Mine	193	25.00	0.30	6.70	7.50	0.71	4
2	No. 13-1 coal seam of Xieqiao Mine	357	10.00	0.12	12.00	14.00	2.44	2
3	No. 3 coal seam of Dayan No. 2 Mine	435	16.60	0.00	8.80	9.23	0.52	2
4	No. 3 coal seam of Qishan Mine	300	14.00	0.00	8.00	10.10	1.00	2
5	No. 2 coal seam of Taoyer Mine	290	6.50	0.20	7.20	16.77	1.44	1
6	No. 2 coal seam of Xuecun Mine	300	16.00	0.40	14.50	16.20	0.36	3
7	No. 4-2 coal seam of Cuijiagou Mine	262	17.50	0.05	6.85	9.50	1.78	2
8	No. 5 coal seam of Xuangangjiaojiazhai	140	2.80	0.00	6.50	14.71	1.20	2
9	No. 15 coal seam of Yangquan No. 1 Mine	250	20.00	0.30	6.00	8.01	1.91	3
10	No. 7-8 coal seam of Pingdingshan No. 1 Mine	300	14.00	0.00	7.90	12.09	1.16	2
11	No. 3 coal seam of Lu'anwangzhuang Mine	200	16.00	0.00	7.00	7.43	0.56	2
12	No. 4-2 coal seam of Xiashijie Mine	177	17.50	0.00	12.00	9.10	0.17	3
13	No. 18 coal seam of Hegang Nanshan Mine	150	10.00	0.40	12.20	11.20	1.34	3
14	No. 15 coal seam of Yinying Mine	220	8.50	0.00	6.30	10.65	0.82	2
15	No. 2-4 coal seam of Gujiaojialequan	127.5	10.00	0.52	8.20	9.11	0.56	4
16	No. 3 coal seam of Xinzhuang Mine	175	10.00	0.00	7.00	13.02	0.78	2
17	No. 3 coal seam of Tang'an Mine	200	15.00	0.30	6.50	13.82	0.81	2
18	No. 5 coal seam of Xiangshan Mine	230	8.50	0.00	6.50	16.08	0.72	2
19	No. 2-4 coal seam of Dongliang Mine	300	16.00	0.30	12.50	14.80	1.14	3
20	No. 2-14 coal seam of Wulong Mine	300	16.00	0.40	14.50	16.20	0.36	3
21	No. 11 coal seam of Yaodatong Mine	300	30.00	0.40	8.60	9.85	0.25	4
22	No. 3 coal seam of Fenghuangshan Mine	140	35.00	0.10	6.50	8.01	0.62	4
23	No. 10 coal seam of Shuiyu Mine	190	6.50	0.00	7.20	16.77	1.44	1
24	No. 5 coal seam of Liujialiang Mine	140	2.80	0.00	6.00	19.40	1.20	2
25	Shenyang Puhe Mine lignite	357	10.00	0.12	12.00	14.00	2.44	2
26	No. 8-9 coal seam of Wangfeng Mine	90	14.00	0.70	8.00	10.10	0.00	4
27	No. 3 coal seam of Baodian Mine	435	16.60	0.00	8.80	9.23	0.52	2
28	No. 3 coal seam of Xinglongzhuang Mine	412.5	13.74	0.30	7.80	12.88	0.58	2
29	No. 2 coal seam of Chaohua Mine	160	5.00	0.00	8.40	16.00	0.73	2
30	No. 5 coal seam of Dalong Mine	200	6.70	0.20	6.50	17.00	0.60	2
31	No. 3 coal seam of Yong'an Mine	193	25.00	0.10	6.70	7.50	0.71	4
32	No. 15 coal seam of Yangquan No. 4 Mine	240	20.00	0.20	6.50	8.60	0.40	3
33	No. 6-17 coal seam of Gaozhuang Mine	300	14.00	0.00	8.00	10.05	1.00	2
34	No. 2-1 coal seam of Micun Mine	130	5.00	0.00	8.50	17.60	1.14	2
35	No. 2 coal seam of Xingtai Mine	360	20.00	0.40	6.20	8.30	0.41	3
36	No. 8-1 coal seam of Baode Mine	263.5	8.00	0.15	7.46	9.50	0.52	3
37	No. 4-2 coal seam of Buertai Mine	424.5	15.00	0.12	6.71	15.00	2.04	2
38	No. 3 coal seam of Wangpo Mine	380	25.00	0.00	6.97	12.00	0.39	2
39	No. 12 coal seam of Huojitu Mine	106	33.00	0.14	6.45	14.20	1.20	2
40	X coal seam of No.3 Hebi Mine	750	6.50	0.30	8.00	16.00	0.33	2
41	No. 8 coal seam of Wobei Mine	600	8.91	0.88	2.75	26.16	1.04	1
42	No. 8301-5 coal seam of X Mine	484	24.00	0.26	6.39	9.00	1.40	4
43	No. 8101-5 coal seam of X Mine	240	13.80	0.45	7.43	12.40	0.61	3
44	No. 8101-8 + 10 coal seam of X Mine	273	13.80	0.30	9.00	12.00	2.30	3
45	No. 2-1304 coal seam of Zhaozhuang Mine	440	14.03	0.19	6.36	11.98	0.44	2
46	No. 28071-2 coal seam of Anlin Mine	365	13.58	0.00	5.06	3.00	2.60	3

TABLE 2: Continued.

Number	Coal seam name	H (m)	R_c (MPa)	MG (m)	M (m)	DN	K	Grade
47	Coal seam of Zhangshuanglou Mine	600	14.60	0.30	4.50	18.00	1.25	2
48	Coal seam of Changcun Mine	350	20.00	0.30	6.70	2.00	1.23	2
49	Coal seam of Wangcun Mine	250	25.00	0.20	6.50	2.00	1.28	2
50	Coal seam of Zhangcun Mine	230	20.00	0.20	7.00	2.00	0.68	2
51	No. 16 coal seam of Lutian Coal Mine	150	9.55	0.30	8.07	30.25	0.90	2
52	No. X coal seam of Datong No. 2 Mine	325	30.00	0.35	7.75	6.00	0.25	4
53	No. X coal seam of Tiefsa Daming Mine	200	22.50	0.25	8.25	10.00	0.95	3
54	No. X coal seam of Nantun Mine	500	7.00	0.15	7.00	12.00	0.25	1
55	No. X coal seam of Xinji Mine	325	17.50	0.25	8.00	8.00	0.80	3
56	Mine big coal of No. 6 Hebi Mine	325	12.50	0.25	6.50	10.00	1.50	2
57	No. 7 coal seam of Sanhejian Mine	200	17.50	0.05	7.50	12.00	0.60	2
58	No. X coal seam of Guzhuang Mine	200	17.50	0.15	7.00	12.00	0.95	2
59	No. X coal seam of Yangchangwan Mine	350	14.50	0.25	10.00	12.50	1.50	2
60	No. X coal seam of Baicaoyu Mine	500	7.00	0.25	9.75	14.00	0.95	2
61	No. X coal seam of Yuhua Mine	500	12.50	0.15	7.75	14.00	0.80	2
62	No. X coal seam of Jiahe Mine	405	17.5	0	2.4	13.5	1.45	1
63	No. 3 coal seam of Bofang Mine	390	12	0	4.5	13	1.5	1
64	No. 13 coal seam of Jinshan Mine	235	6.4	0.5	13.53	14.5	2	1
65	No. 4 coal seam of Majialiang Mine	540	25	0	7	15	1	1
66	No. X coal seam of Wajinwan Mine	310	11	0	4.22	15	1.5	1
67	No. X coal seam of Shigejie Mine	424	16.5	0.3	6.64	13.5	1.65	1
68	No. X coal seam of Wuyang Mine	350	10	0.1	6.59	14	0.77	1

*Data come from the database of publicly published papers on CNKI.

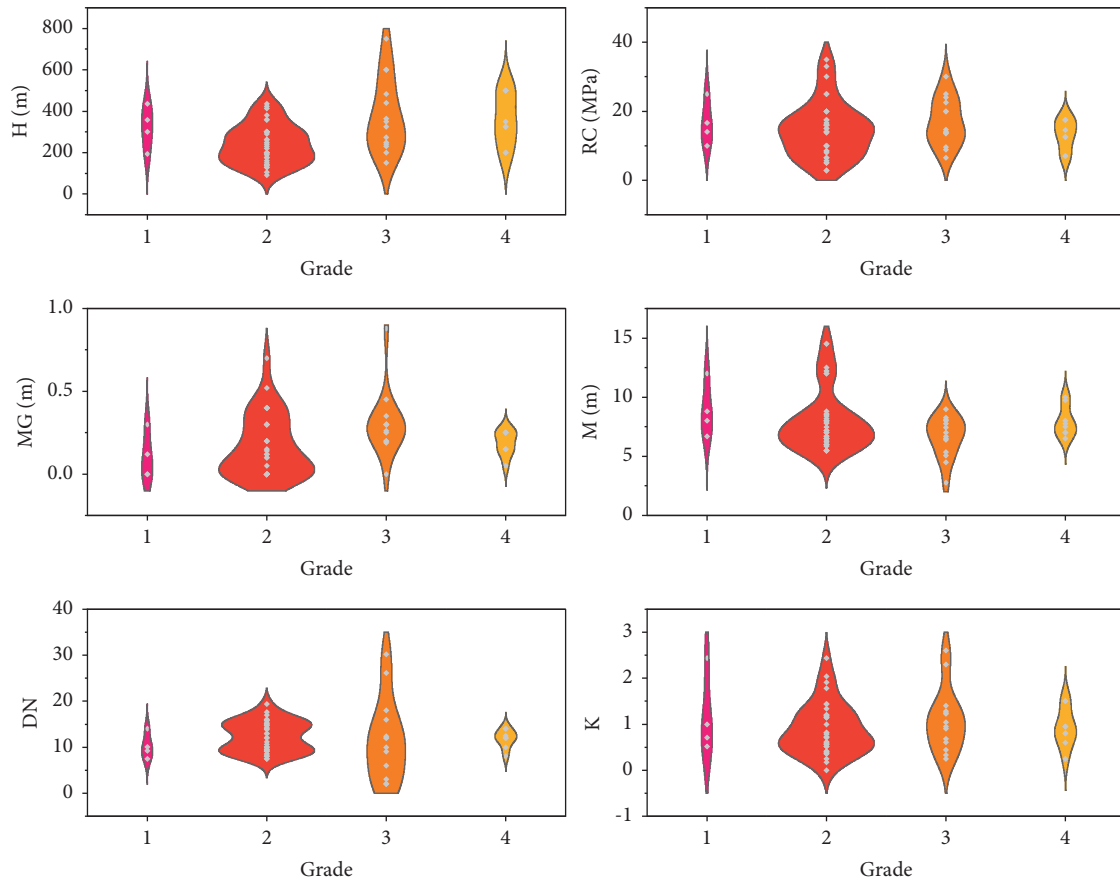


FIGURE 2: Distribution of sample data.

5.1. Paired Indicators. The paired indicators mainly include the evaluation indicators of the binary prediction model, such as accuracy and error rate (Acc&Err), precision and recall (P&R), and true positive rate and true false rate (TPR&TFR), and the evaluation indicators of the multi-classification prediction model, such as macro-accuracy and macro-error rate and macro-precision and macro-recall, which are expanded from binary classification to multiclassification.

Accuracy and error rate (Acc&Err) is used to calculate the proportion of the samples with correct prediction classification and the samples with wrong prediction classification to the total samples. The calculation formula of accuracy and error rate is shown in equations (1) and (2). The range of accuracy and error rate (Acc&Err) is both [0, 1]. Generally, the closer the accuracy is to 1, the better the performance of the model. On the contrary, the closer the error rate is to 0, the better the model's performance:

$$accuracy = 1 - \frac{FP + FN}{N_{sample}} = \frac{TP + TN}{N_{sample}}, \quad (1)$$

$$error\ rate = \frac{FP + FN}{N_{sample}}, \quad (2)$$

where FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, TP is the number of samples predicted to be true positives, TN is the number of samples predicted to be true negatives, and N_{sample} is the total number of samples.

Precision rate and recall rate, respectively, calculated the proportion of the predicted positive samples to the predicted positive samples and the actual positive samples. The calculation formula of precision and recall is shown in equations (3) and (4). The range of precision and recall is both [0, 1]. Ideally, the closer the precision and recall are to 1, the better the model's performance. However, in practice, the relationship between FP and FN is the relationship between type I error and type II error, so precision and recall (P&R) is a contradictory relationship. Therefore, it is generally necessary to find a balance between the precision and recall:

$$precision = \frac{TP}{TP + FP}, \quad (3)$$

$$recall = \frac{TP}{TP + FN}, \quad (4)$$

where TP is the number of samples predicted to be true positives, TN is the number of samples predicted to be true negatives, and FP is the number of samples predicted to be false positives.

True positive rate and true false rate (also called sensitivity and specificity, TPR&TFR), respectively, calculate the proportion of samples correctly predicted as positive cases to the total positive samples, and the proportion of samples correctly predicted as negative cases to the total negative samples. The calculation formula of true positive rate and true false rate is shown in equations (5) and (6). The range of true positive rate and true false rate is both [0, 1]. Ideally, the

closer the true positive rate is to 1 and the closer the true false rate is to 0, the better the model performance:

$$TPR = sensitivity = \frac{TP}{TP + FN}, \quad (5)$$

$$TFR = specificity = \frac{TN}{TN + FP}, \quad (6)$$

where FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, TP is the number of samples predicted to be true positives, and TN is the number of samples predicted to be true negatives.

Macro-accuracy and macro-error rate and macro-precision and macro-recall are also called averaged-accuracy and averaged-error rate and averaged-precision and averaged-recall, which extended from the problem of dichotomy for requirements of multiclassification problems. They all treat each category equally, add up the same indicators of different categories, and then calculate the average to realize the evaluation of the multiclassification prediction model. Therefore, their value range and significance are the same as evaluating the two-classification prediction models. Their calculation formula is shown in the following equations:

$$macro - accuracy = \frac{1}{n} \times \sum_{i=1}^n \frac{TP_i + TN_i}{N_{isample}}, \quad (7)$$

$$macro - error\ rate = \frac{1}{n} \times \sum_{i=1}^n \frac{FP_i + FN_i}{N_{isample}}, \quad (8)$$

$$macro - precision = \frac{1}{n} \times \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i}, \quad (9)$$

$$macro - recall = \frac{1}{n} \times \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i}, \quad (10)$$

where FP_i is the number of samples predicted to be false positive in class i , FN_i is the number of samples predicted to be false negative in class i , TP_i is the number of samples predicted to be true positive in class i , TN_i is the number of samples predicted to be true negative in class i , $N_{isample}$ is the total number of samples of class i , and n is the number of categories.

5.2. Comprehensive Indicators. The comprehensive indicators mainly include F -measure (also known as F -score, $F1$) [40], Matthews correlation coefficient (Phi coefficient) [41], Kendall's tau, and Spearman's rho, used for binary classification. Furthermore, as well as for evaluating the multi-classification prediction model, the macro- F -measure and macro-Matthews correlation coefficient (macro-Phi coefficient) of the evaluation indicators of the multiclassification prediction model are expanded from the evaluation indicators of the two-classification prediction model.

F -measure is proposed to solve the contradictory model performance measurement value of precision and recall.

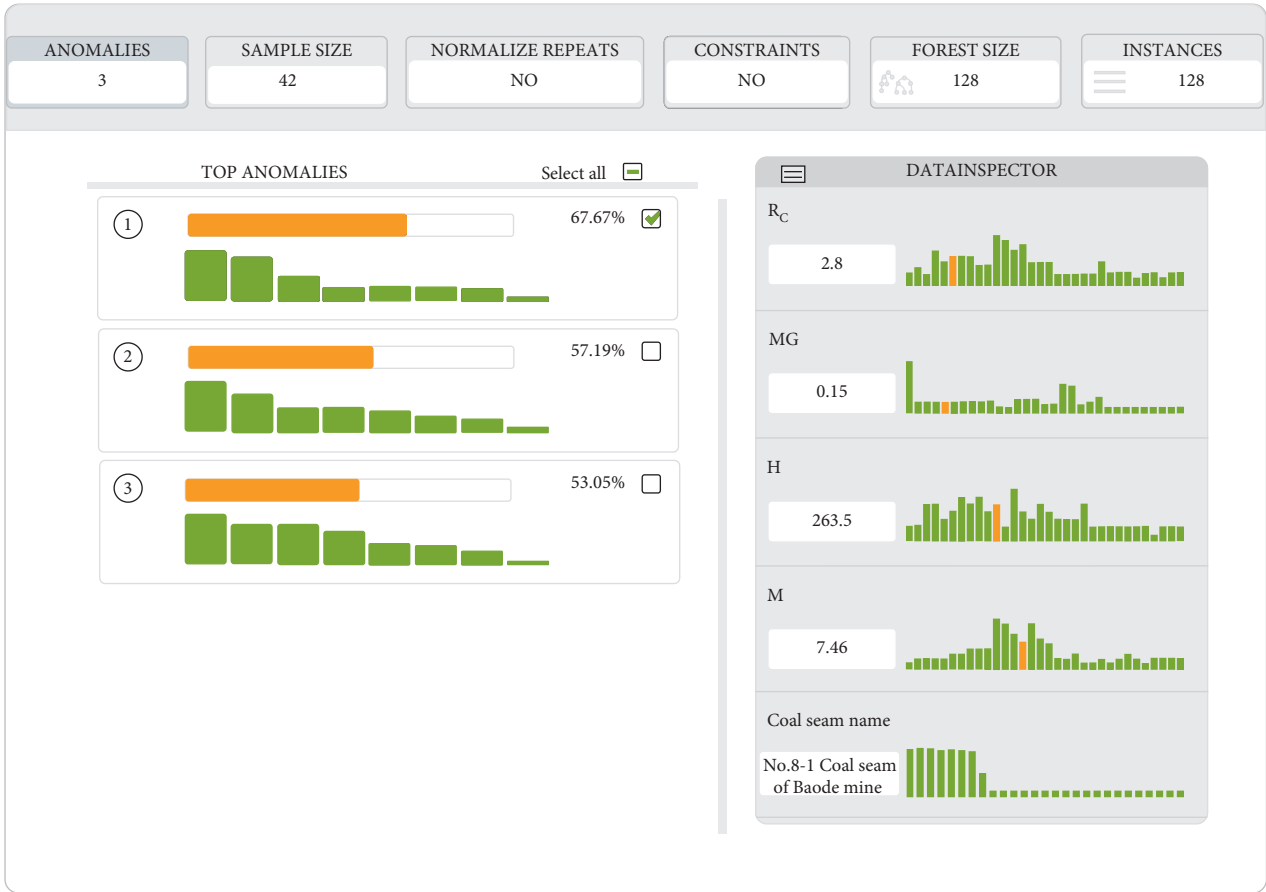


FIGURE 3: Anomaly detection results of sample data.

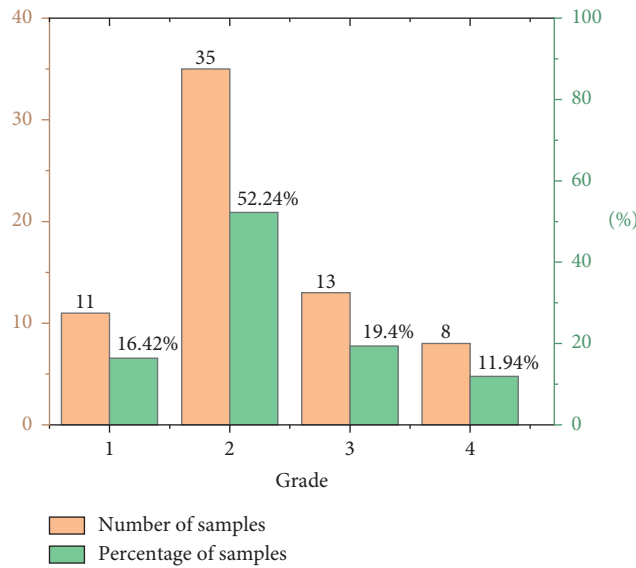


FIGURE 4: Sample data distribution after cleaning.

Therefore, F -measure is a balance point between precision and recall (that is, the harmonic average of precision and recall), which can take into account the precision and recall of the classification model at the same time. Its formula is shown

in equation (11). The value range of F -measure is $[0, 1]$. Ideally, the closer the F -measure value is to 1, the better the model performance, and vice versa, the worse the model performance:

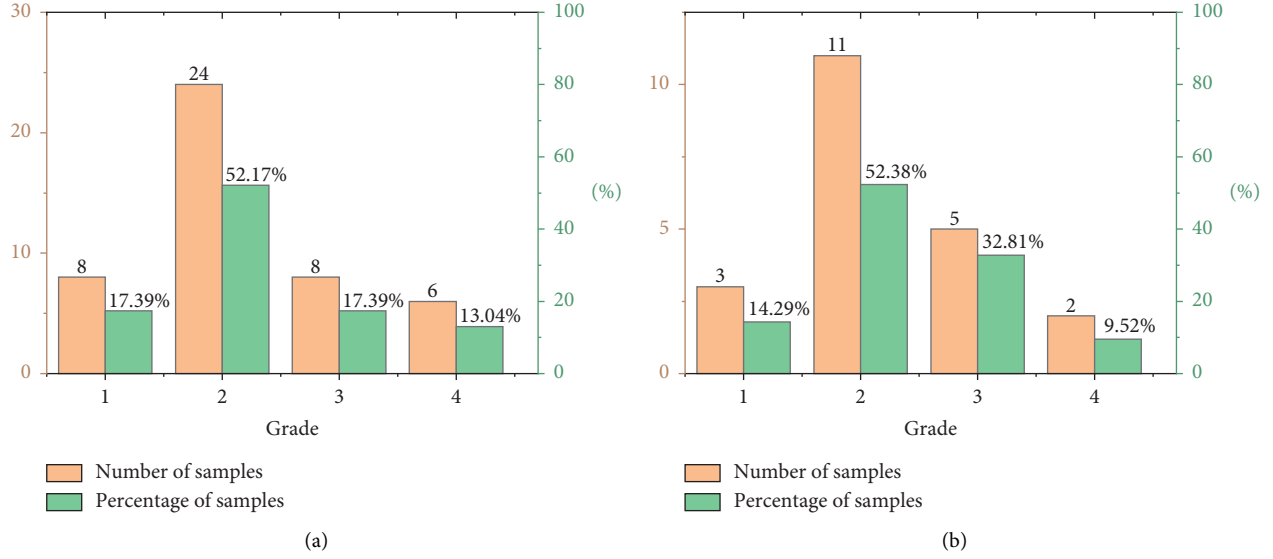


FIGURE 5: Sample data distribution of training set and test set. (a) Distribution of training sample data. (b) Distribution of test sample data.

$$\begin{cases} \frac{1}{F_1} = \frac{1}{2} \cdot \left(\frac{1}{P} + \frac{1}{R} \right), \\ F_1 = \frac{2 \times P \times R}{P + R} = \frac{2TP}{2TP + FP + FN}, \end{cases} \quad (11)$$

where P is precision, R is recall, FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, and TP is the number of samples predicted to be true positives.

The Matthews correlation coefficient (Phi coefficient), which is mainly used to measure the two classification

problems, is a relatively balanced indicator. It comprehensively considers TP , TN , FP , and FN , and it can also be used in the case of unbalanced sample data categories. The value range of Phi coefficient is $[-1, 1]$, a value of 1 indicates that the prediction is entirely consistent with the actual, a value of 0 indicates that the predicted result is not as good as the random predicted result, and -1 indicates that the predicted result is utterly inconsistent with the actual result [42]. Its calculation formula is shown in the following equation:

$$\text{Phi coefficient} = \frac{TP \times TN - TP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (12)$$

where FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, TP is the number of samples predicted to be true positives, and TN is the number of samples predicted to be true negatives.

Macro- F -measure and macro-Matthews correlation coefficient (macro-Phi coefficient) are also called averaged F -measure and averaged-Matthews correlation coefficient, respectively. It is also based on the needs of the

multiclassification problem and extended from the two classification problems. They all treat each category equally, add up the same indicators of different categories, and then calculate the average to realize the evaluation of the multiclassification prediction model. Their value range and significance are the same as the evaluation of the two-classification prediction models. Their calculation formula is shown in the following equations:

$$\text{macro-}F_1 = \frac{1}{n} \times \sum_{i=1}^n \frac{2 \times P_i \times R_i}{P_i + R_i} = \frac{1}{n} \times \sum_{i=1}^n \frac{2TP_i}{2TP_i + FP_i + FN_i}, \quad (13)$$

$$\text{macro-} \text{Phi coefficient} = \frac{1}{n} \times \sum_{i=1}^n \frac{TP_i \times TN_i - TP_i \times FN_i}{\sqrt{(TP_i + FP_i) \times (TP_i + FN_i) \times (TN_i + FP_i) \times (TN_i + FN_i)}}, \quad (14)$$

TABLE 3: Confusion matrix of two-classification problems [39].

	Actual positive class	Actual negative class
Predicted positive class	True positive (TP)	False negative (FN)
Predicted negative class	False positive (FP)	True negative (TN)

where FP_i is the number of samples predicted to be false positive in class i , FN_i is the number of samples predicted to be false negative in class i , TP_i is the number of samples predicted to be true positive in class i , TN_i is the number of samples predicted to be true negative in class i , and n is the number of categories.

5.3. Visual Indicators. Visual indicators mainly include ROC curve [43] and AUC [44], precision-recall curve (also known as PR curve) [45], gain curve [46], K-S curve and K-S statistical value [47], and lift curve [48] and lift value.

ROC curve, also known as receiver operating characteristic curve, is a comprehensive indicator that reflects TPR and FPR with the decision threshold [49]. It is a curve composed of points (TPR and FPR), the abscissa is FPR, and the ordinate is TPR. ROC curve is mainly used to compare the relative performance of different classification models. However, when the ROC curves of different classification models intersect, it is not easy to reasonably evaluate the models' relative performance.

AUC, also known as the area under the ROC curve, is often used in conjunction with the ROC curve. The value range of AUC is [0, 1]. According to experience, when the AUC value is less than 0.5, the predictive ability of the model is worse than random guessing, but if the prediction is reversed, it is better than random guessing; when the AUC value is equal to 0.5, the model has no predictive value, just as a random guess; when AUC value is more than 0.7, the model's predictive ability can be considered acceptable; when the AUC value is equal to 1, the model's predictive ability is perfect, and using this model, no matter what threshold is set, a perfect prediction can be obtained (most of the time does not exist). The specific AUC value range and its empirical evaluation of the model are shown in Table 4.

PR curve is a curve that reacts to the relationship between precision and recall. It is also used to evaluate the relative performance of different classification models and can be used with the AUC value. The PR curve is an essential supplement to the ROC curve, especially in unbalanced sample categories; the PR curve can reflect the classifier's quality more effectively than the ROC curve.

The gain curve (or cumulative gain curve) is an indicator to describe the global accuracy. It represents the relationship between the percentage of correct predictions for positive cases and the effort required to achieve them, measured by the percentage of prediction cases. The Y-axis in the gain curve is equal to recall and sensitivity (TPR), and the X-axis is the percentage of positive instances. The formulas of these indicators are shown in the following equations:

$$gain = recall = TPR = \frac{TP}{TP + FN}, \quad (15)$$

$$\% \text{ of positive instances} = \frac{TP + FP}{TP + FP + TN + FN}, \quad (16)$$

where FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, TP is the number of samples predicted to be true positives, and TN is the number of samples predicted to be true negatives.

K-S curve (Kolmogorov–Smirnov curve), also called Lorentz curve, is used to describe the quality of the classification model. The K-S curve draws two curves with TPR and FPR as the vertical axis and the threshold as the horizontal axis. Thus, it reflects the difference between TPR and FPR at the same threshold. In general, the farther the two broken lines are, the better the classification model distinguishes between positive and negative samples. The K-S statistical value measures the maximum difference between TPR and FPR within the range of all possible thresholds, that is, the upper limit of the classification model for the discrimination between positive and negative samples. The calculation formula of the K-S statistic is shown in equation (17). The value range of the K-S statistic is [0, 1]. Ideally, the closer the K-S statistic is to 1, the stronger the classification model's ability to distinguish between positive and negative samples. According to experience, when the K-S statistic is less than 0.2, the model cannot distinguish between positive and negative samples, and when the K-S statistic is more significant than 0.4, the model can distinguish between positive and negative samples [51]. The specific range of K-S statistics and its empirical evaluation of the model are shown in Table 5:

$$K - S \text{ statistic} = \max(TPR - FPR), \quad (17)$$

where FPR is the negative cumulative response, $FPR = FP / TR + FP$, and TPR is sensitivity.

Lift curve, which measures the increase in the accuracy of the model's prediction results under a certain threshold compared to the accuracy of the random prediction results without the model. In short, it is how much the prediction effect has been improved by using this model and not using this model for prediction. The larger the lift, the better the prediction effect of the model. The calculation of lift is shown in the following equation:

$$\begin{aligned} \text{lift} &= \frac{TP/TP + FP}{TP + FN/TP + TN + FP + FN} \\ &= \frac{TP \times (TP + TN + FP + FN)}{(TP + FP) \times (TP + FN)}, \end{aligned} \quad (18)$$

where FP is the number of samples predicted to be false positives, FN is the number of samples predicted to be false negatives, TP is the number of samples predicted to be true positives, and TN is the number of samples predicted to be true negatives.

The above parameters and indicators for evaluating model performance are provided in BigML. Through the

TABLE 4: AUC value range and its empirical evaluation of the model [50].

Range of AUC values	Model performance evaluation
0.50~0.60	Fail
0.60~0.70	Poor
0.70~0.80	Fair
0.80~0.90	Good
1.00	Excellent

brief introduction of the above indicators, it is not difficult to find that these indicators evaluate the performance of the classification prediction model from different evaluation angles to realize the multidirectional and multiangle analysis and measurement of the prediction model performance [52]. In addition, it can also be seen that there is no single indicator that can evaluate the performance of the classification prediction model in an all-around way, and more is the collaborative evaluation of multiple parameter indicators. It can be seen from Figure 4 that the sample data of top coal caving have severe category imbalance (i.e., grade imbalance), and the classification evaluation of top coal caving is a multiclass prediction problem. So, it is necessary to select multiple parameters and indicators from the above indicators to evaluate the prediction model's performance to ensure that the prediction model is robust.

Accuracy/error rate is the most commonly used indicator for researchers to evaluate the performance of classification prediction models, because they calculated the ratio of the number of correctly classified predictions to the total number of predictions and the number of incorrectly classified predictions accounted for the total number of predictions, and they can objectively reflect the global quality of the model. However, accuracy/error rate is not a good indicator when the sample data categories are unbalanced, because when the sample data categories are unbalanced, the prediction will favor the category that accounts for the majority of the total sample and ignore the category that accounts for the minority of the total sample, resulting in the category that accounts for the minority of the total sample does not have the classification prediction ability or the classification prediction ability is weak. In addition, because the top coal caving classification evaluation is a multiclassification prediction problem, this article will not take the accuracy/error rate and precision and recall rate as the indicators of model performance evaluation. So, only take the macro-accuracy/macro-error rate and macro-precision and the macro-recall rate as the auxiliary indicators. Moreover, taking the macro-Matthews correlation coefficient, ROC curve, PR curve, K-S curve, gain curve, lift curve K-S statistics, AUC value, and lift value are the leading indicators to evaluate the model's performance.

6. Predictive Model Establishment and Its Performance Evaluation

It is often not easy to get a more robust and stable classification prediction model, and therefore, it is also challenging to achieve overnight. Therefore, more needs to continue

exploring and optimizing the model to obtain a more robust and stable classification prediction model. In BigML, the methods of establishing classification prediction models include models (decision trees), ensemble (bagging, random decision forests, and boosted trees), deep nets, logistic regression, and other methods. However, it is not easy to know in advance the specific ways and methods to obtain. So, it is necessary to use the modeling method provided by BigML to establish one or more exploratory models and continuously evaluate and optimize the models to obtain a robust and stable classification evaluation model of top coal caving.

BigML is not only a very friendly machine learning platform, which can build models (decision tree), ensemble (bagging, random decision forests, and boosted trees), deep nets, and logical regression models with one click. Nevertheless, it also considers that noncomputer professional users may have model parameter adjustment and optimization problems, so the "automatic optimization" function is also specially set up. The user can efficiently complete the modeling task through this function by simply specifying training samples and training objectives. The model parameters will be automatically adjusted to the theoretical optimal by BigML. In order to evaluate the performance of model establishment, BigML has set up modules such as single model evaluation, multimodel evaluation, and cross-validation evaluation. These modules only need the user to specify the model to be evaluated and the test set sample used to evaluate the model, and then the model performance evaluation can be easily completed.

6.1. Establishment of the Classification Evaluation Model of Top Coal Caving. The decision tree is the most commonly used method to establish a classification prediction model, so this article uses the decision tree in BigML to establish the top coal caving exploratory classification evaluation model. In addition, when the model (decision tree) in BigML is used to establish the classification evaluation model of top coal caving exploratory, the "automatic optimization" function in BigML is used to build the model, and the prepared training set samples are taken as the model training samples, and the "Grade" in the samples is taken as the training target. Finally, BigML obtains the classification evaluation model of top coal caving based on the training set samples, as shown in Figures 6–9. In order to evaluate the performance of the top coal caving grading evaluation model established by BigML, the "single model evaluation" module of BigML is used to test the model and evaluate the model's performance. The test sample is the test set sample because the sample is small, with only 21 groups, so the sampling method is set to be replaceable. The evaluation results are shown in Figures 10–13 and Tables 6 and 7.

According to Tables 6 and 7 and Figures 10–13, it can be seen that the established classification evaluation model of top coal caving can be barely accepted when the probability threshold is 50% (the default probability threshold of the classification prediction model is 50%, which is also a commonly used threshold for establishing classification

TABLE 5: K-S statistical range and its empirical assessment of model differentiation ability.

Range of K-S statistics	Model differentiation ability
0.00~0.20	Fail
0.20~0.40	Poor
0.40~0.50	Fair
0.50~0.60	Good
0.60~0.750	Very good
0.75~1.00	Excellent

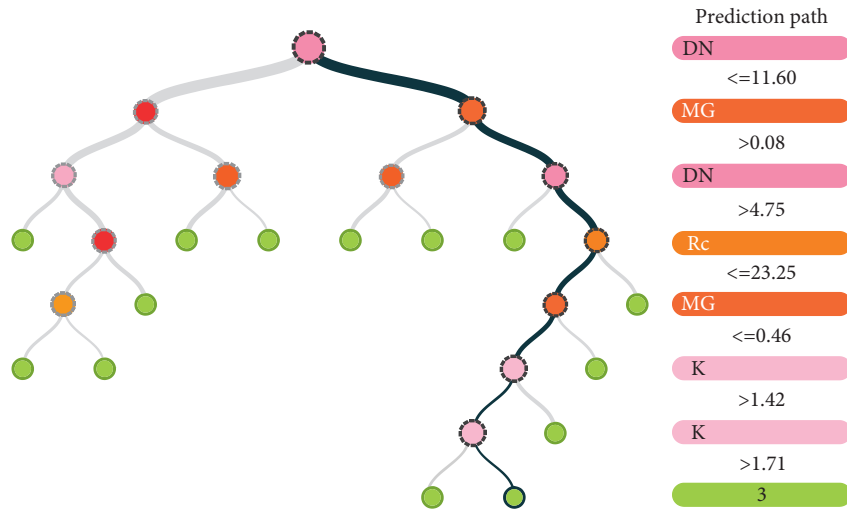


FIGURE 6: The established decision tree model.

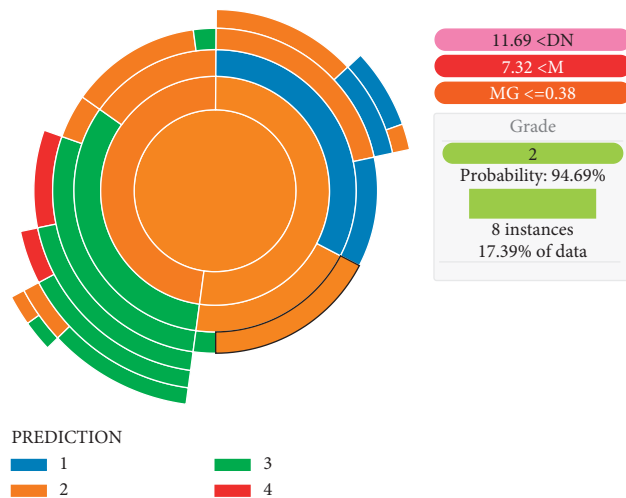


FIGURE 7: Prediction distribution of the decision tree model from the perspective of the sunburst chart.

prediction models [53]). The global prediction accuracy rate (ACCURACY) of the model is 80.95%, the average recall rate (AVG.RECALL) and precision rate (AVG. PRECISION) are 79.17% and 85.42%, respectively, and the average Matthews correlation coefficient (AVG. Phi) and average *F*-measure (AVG. *F*) of the model reach 0.7436 and 0.8087, respectively. However, from the model’s graphical performance evaluation indicators, lift curve (Figure 13), the lift value of each grade is greater than 100% (that is, the model’s prediction

ability for top coal caving of each grade is stronger than the random model). However, from the values of ROC AUC, PR AUC, and K-S in the ROC curve (Figure 10), PR curve (Figure 11), and K-S curve (Figure 12), the model has a certain prediction ability for top coal caving of grade 2 and grade 4, while it has low prediction ability for top coal caving of grade 1 and grade 2. The ROC AUC and PR AUC of grades 2 and 4 are greater than 0.7, and the K-S values are 100%, while the ROC AUC and PR AUC of grades 1 and 3

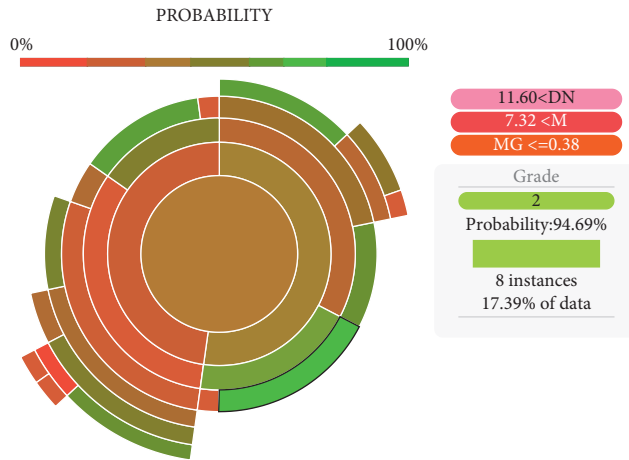


FIGURE 8: Prediction probability distribution of the decision tree model from the perspective of the sunburst chart.

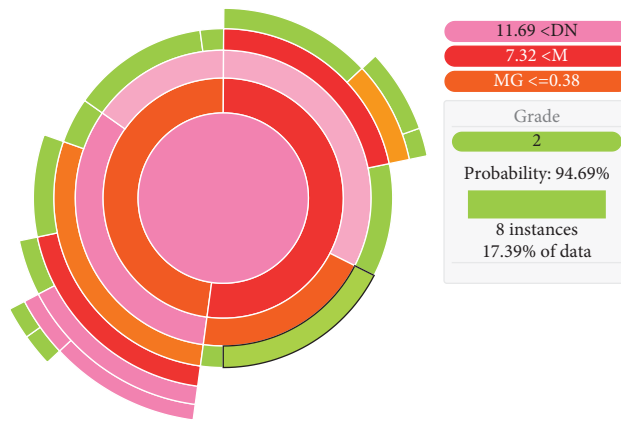


FIGURE 9: Variable distribution of the decision tree model from the perspective of the sunburst chart.

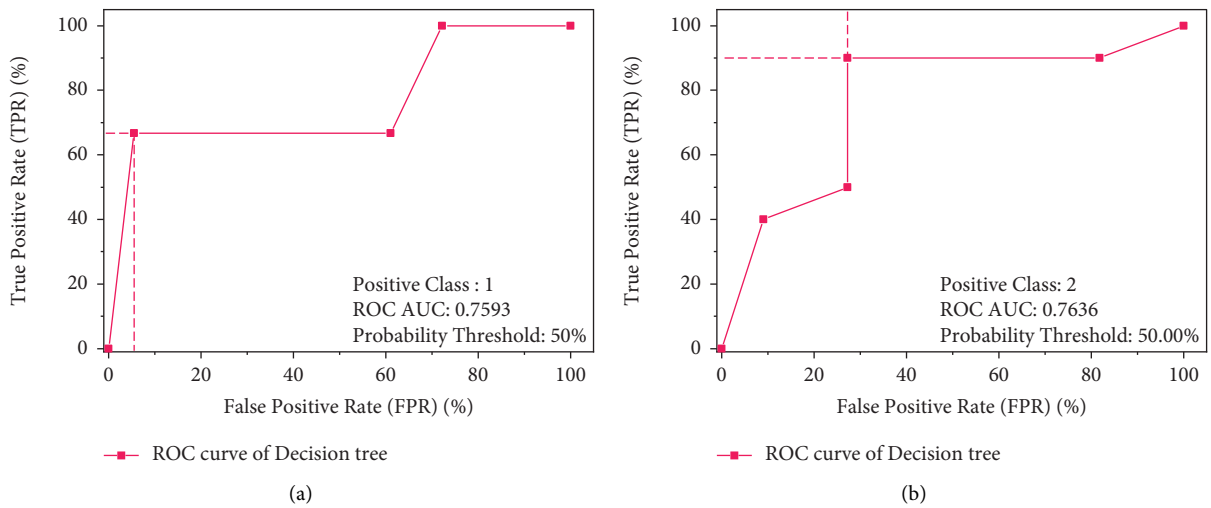


FIGURE 10: Continued.

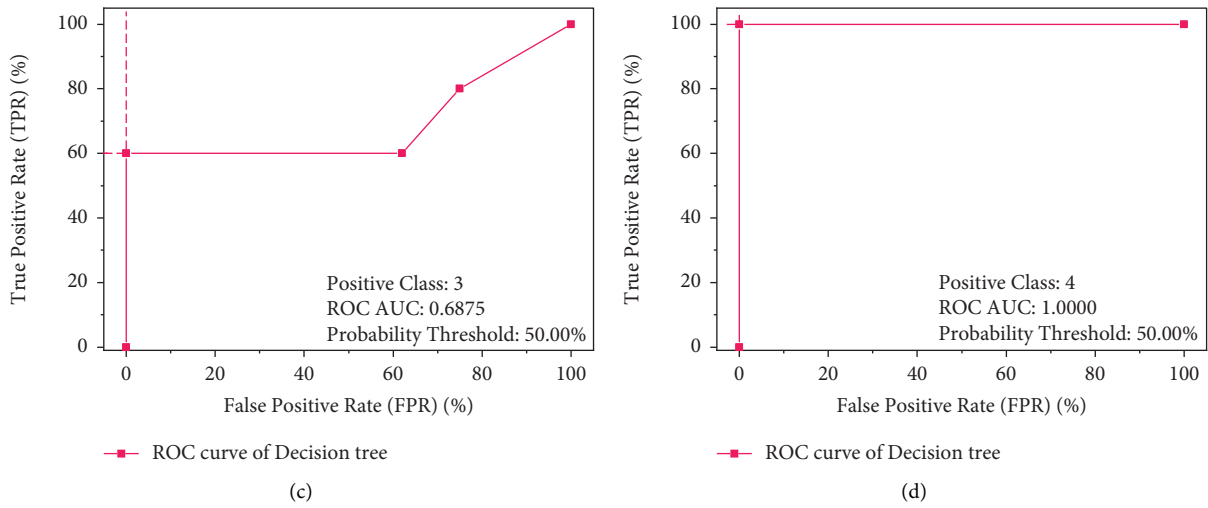


FIGURE 10: ROC curve of each classification test of the decision tree model.

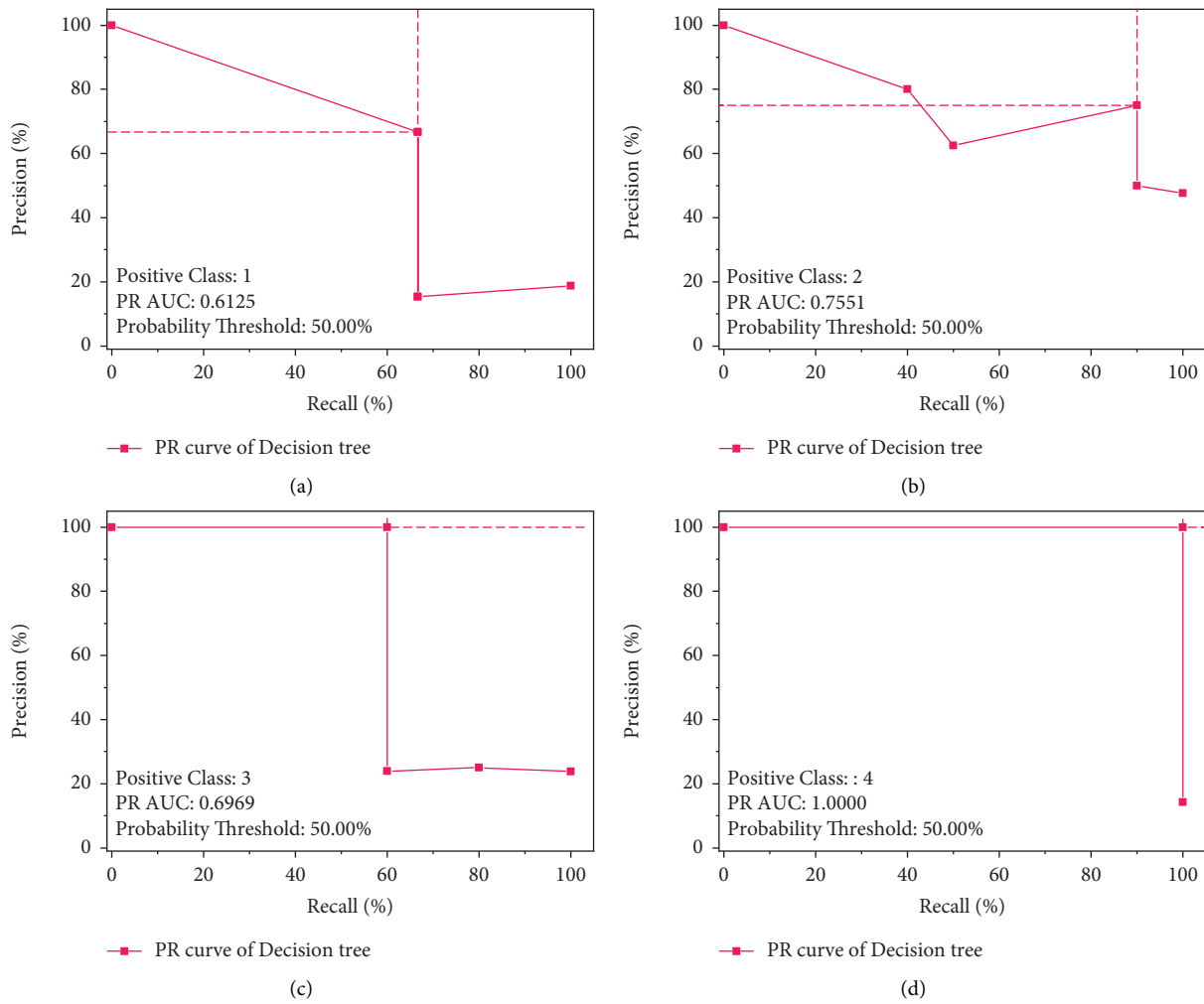


FIGURE 11: PR curve of each classification test of the decision tree model.

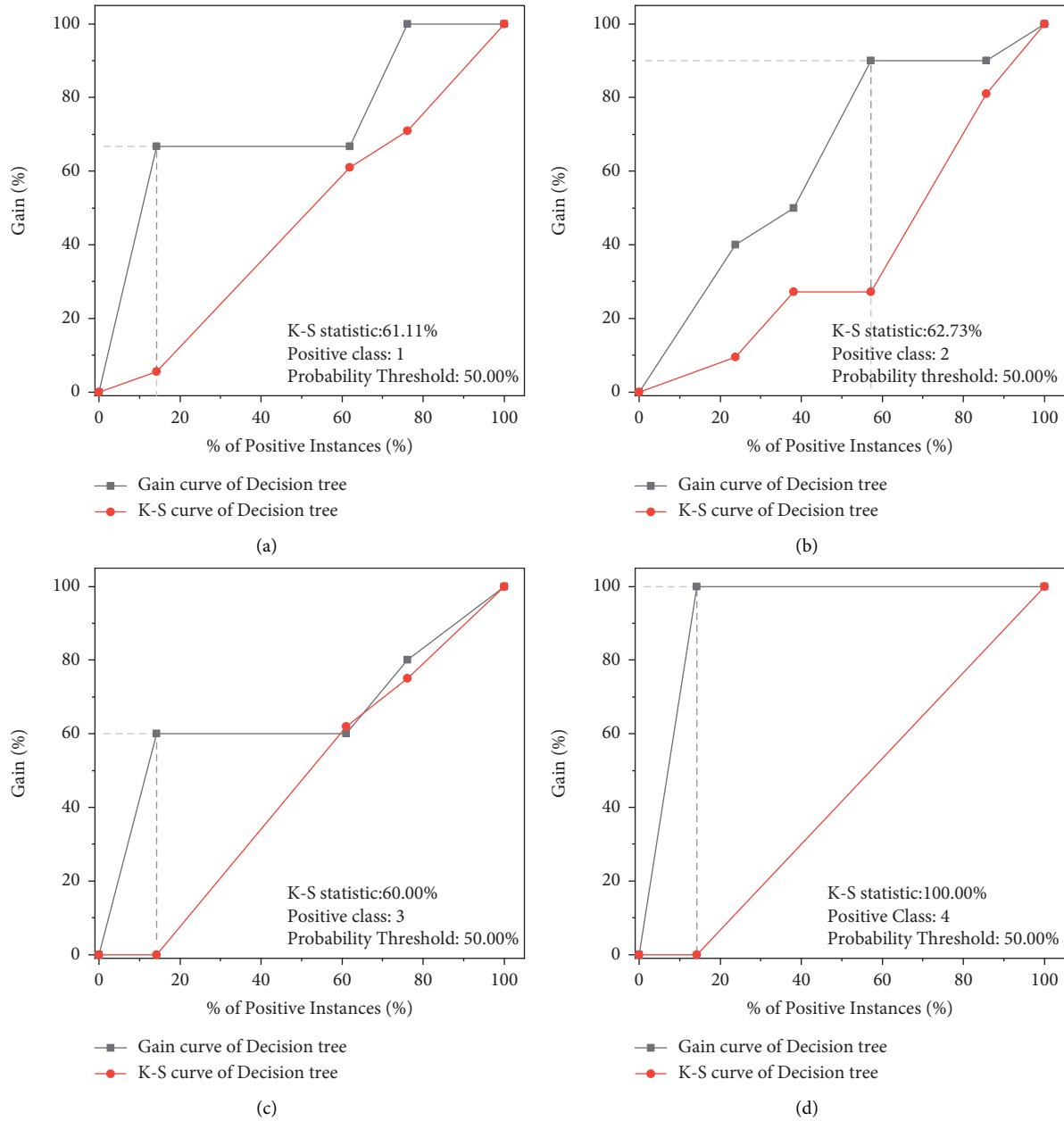


FIGURE 12: K-S and gain curve of each classification test of the decision tree model.

are less than 0.7, and the K-S values are about 60%. Optimization or a better modeling method can be considered to establish a classification evaluation model of top coal caving to achieve a perfect prediction of each grade top coal caving.

The deep network is also a method to establish a classification prediction model, so try to use the deep network in BigML to establish a better top coal classification and caving classification evaluation model. When using the deep network in BigML to establish the top coal classification and caving classification evaluation model, the “automatic optimization” function is also used for modeling, and the automatic method is selected as “Network search.” The model training sample is the training set sample, and the training target is the “Grade” in the sample. After the model

is trained, use the test set sample data to test and evaluate the model’s performance, and the test sampling method is still set to replaceable sampling. The test results are shown in Figures 14 –17 and Tables 8 and 9.

According to Tables 8 and 9, when the probability threshold is 50%, the global performance of the classification evaluation model of top coal caving established by the deep network is not much better than that established by the decision tree. The global prediction ACCURACY of the classification evaluation model of top coal caving established by the deep network is 80.95%, the same as that established by the decision tree; the AVG.RECALL and AVG.PRECISION are 82.26% and 83.04%, respectively, and the AVG. Phi and AVG. *F* are 0.7571 and 0.8084, respectively, which

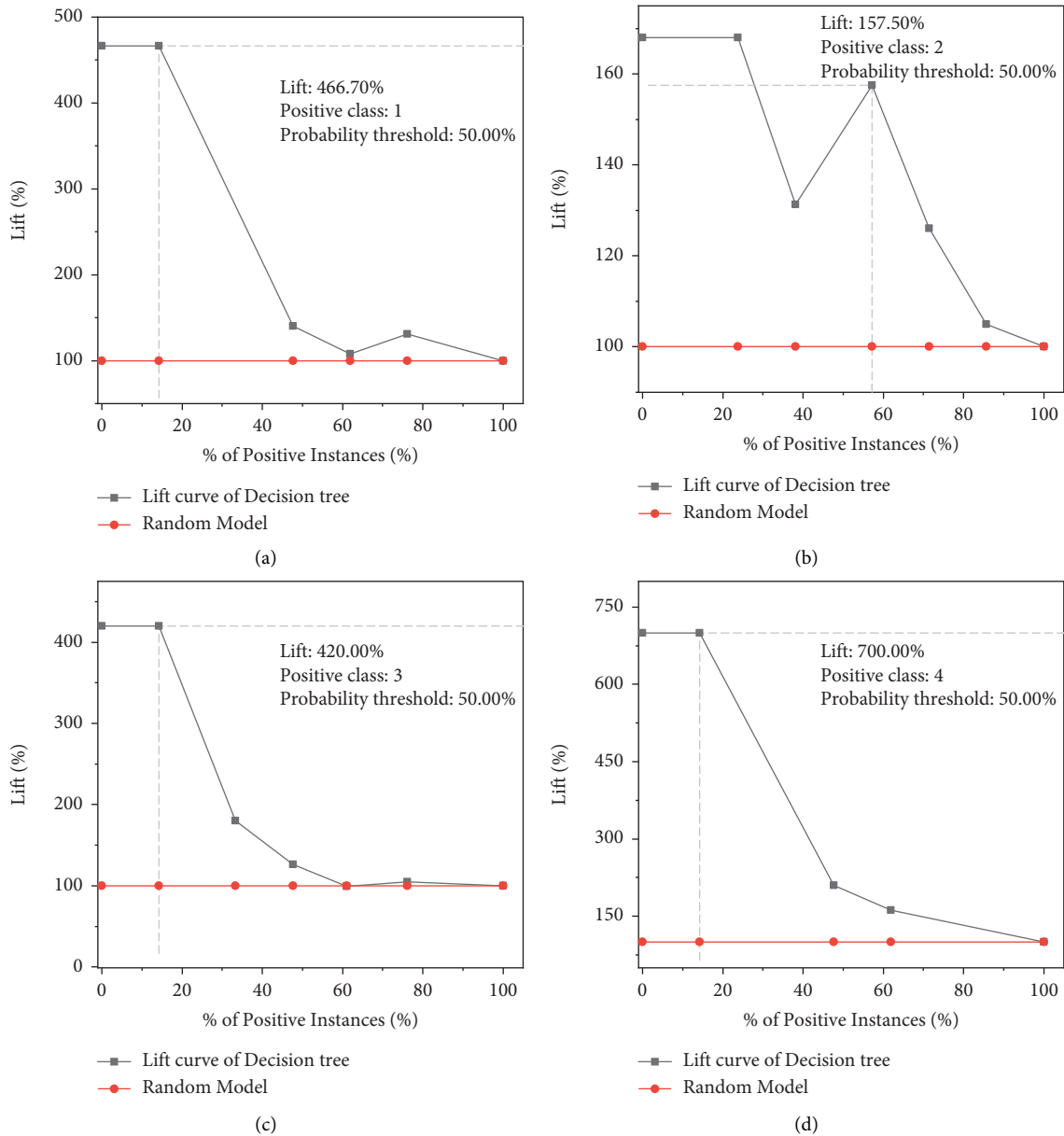


FIGURE 13: Lift curve of each classification test of the decision tree model.

are little different from the AVG.RECALL, AVG.PRECISSION, AVG. Phi, and AVG. *F* of the classification evaluation model of top coal caving established by a decision tree. The global performance parameters of these models show that the global performance of the classification evaluation model of top coal caving established by the deep network is not much better than that established by the decision tree.

However, according to Figures 14–17, among the local evaluation parameters of the model, the model established by the deep network is better than the decision tree. From the model’s graphical performance evaluation indicators, lift curve (Figure 17), the lift value of each evaluation grade is greater than 100%, which means that the prediction ability of the model established by the deep network for top coal caving of each grade is stronger than the

random model. From the ROC AUC, PR AUC, and K-S values of each grade in ROC curve (Figure 14), PR curve (Figure 15) and K-S curve (Figure 16), the ROC AUC, PR AUC, and K-S values of each grade in the classification evaluation model of top coal caving established by the deep network are greater than or equal to that established by the decision tree. In addition, from the ROC AUC, PR AUC, and K-S values of each grade of the classification evaluation model top coal caving established by the deep network, it can be seen that the model has good prediction ability for the top coal caving of each grade. The ROC AUC of each grade in model is greater than 0.7, the PR AUC is basically greater than or equal to 0.7, and the K-S value is greater than 80%. The above analysis shows that under the probability threshold of 50%, although the

TABLE 6: Confusion matrix of decision tree model test results and model evaluation related indicators.

Actual vs. predicted	1	2	3	4	Actual	Recall (%)
1	2	1	0	0	3	66.67
2	1	9	0	0	10	90.00
3	0	2	3	0	5	60.00
4	0	0	0	4	3	100.00
Predicted	3	12	3	3	21	79.17
Precision (%)	66.67	75.00	100.00	100.00	86.42	AVG.RECALL
					AVG.PRECISION	80.95
						ACCURACY

TABLE 7: Performance evaluation results of the decision tree model.

Grade	F-measure	Phi coefficient
1	0.67	0.61
2	0.82	0.63
3	0.75	0.73
4	1.00	1.00
	0.8087	0.7436
	AVG. F	AVG. Phi

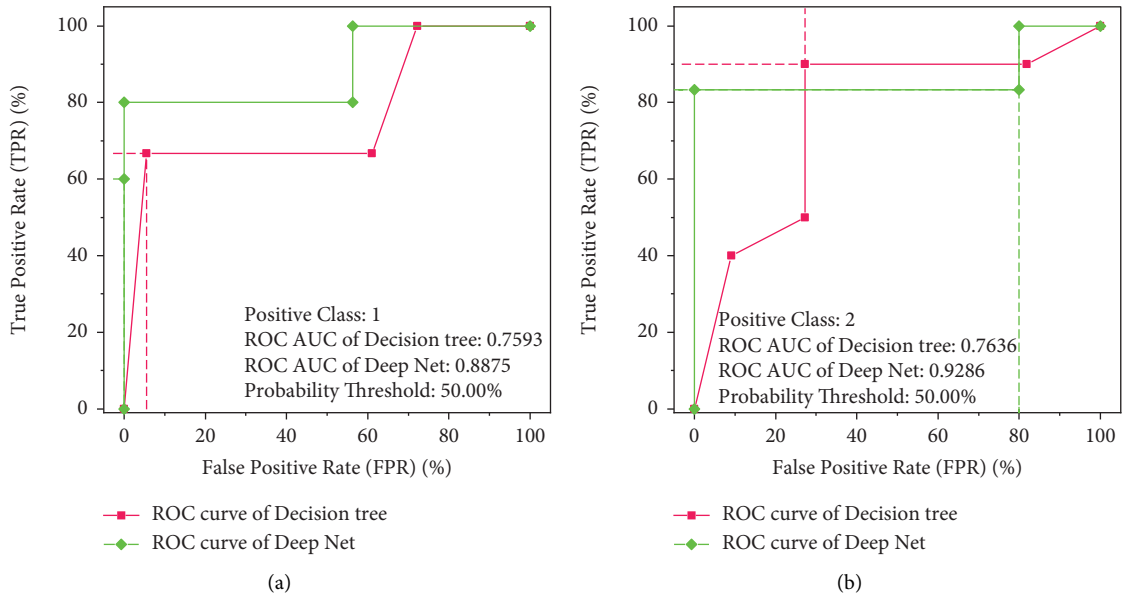


FIGURE 14: Continued.

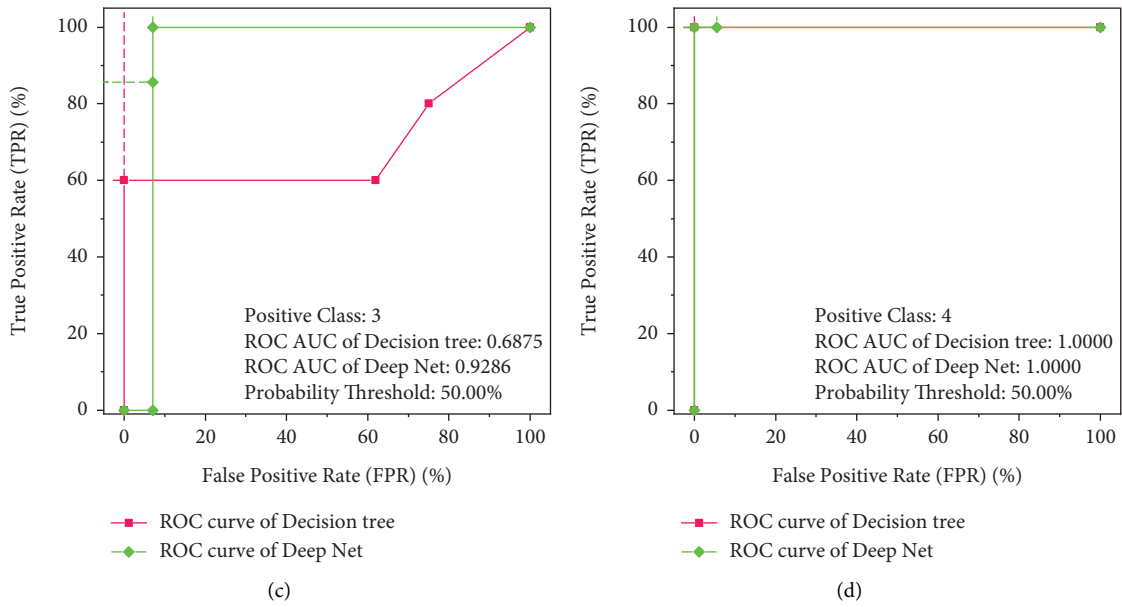


FIGURE 14: ROC curve of each classification test of the deep network model.

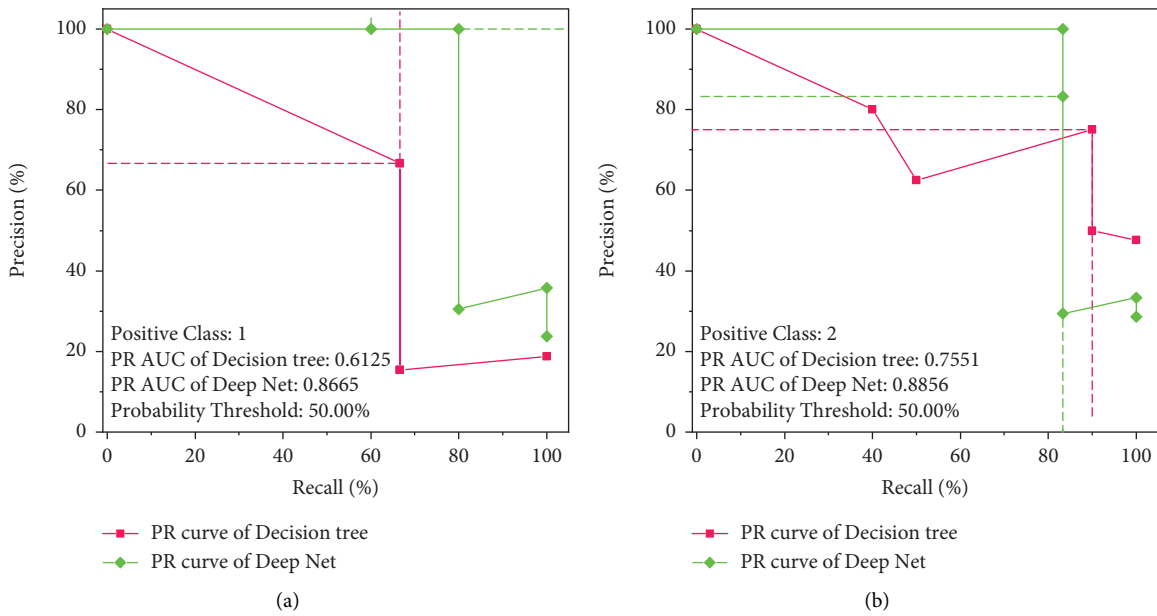


FIGURE 15: Continued.

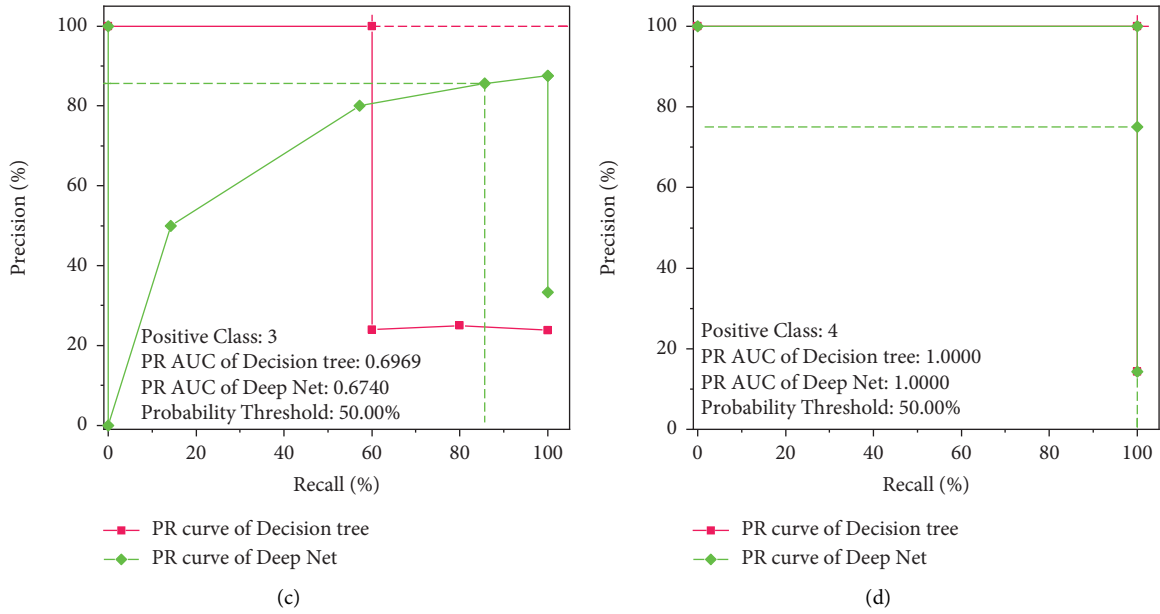


FIGURE 15: PR curve of each classification test of the deep network model.

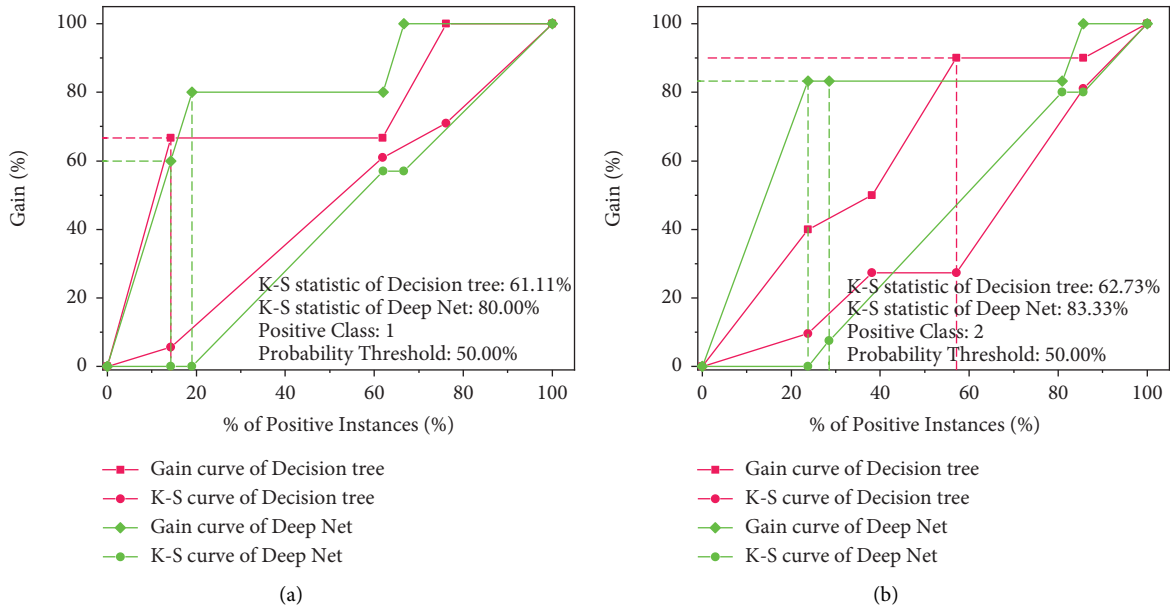


FIGURE 16: Continued.

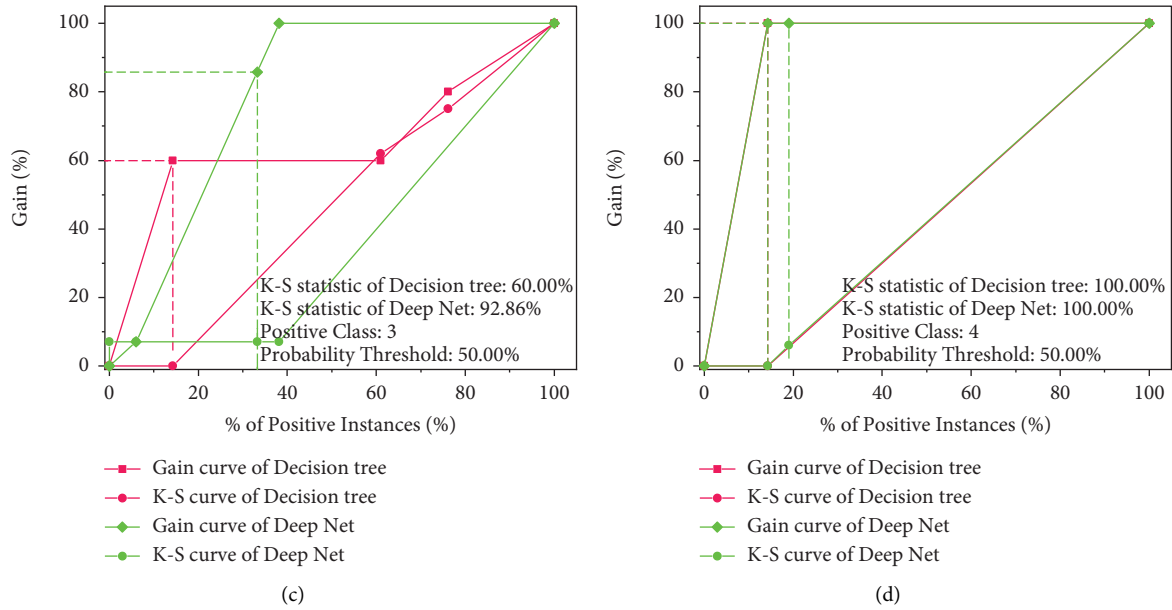


FIGURE 16: K-S curve and gain curve of each classification test of the deep network model.

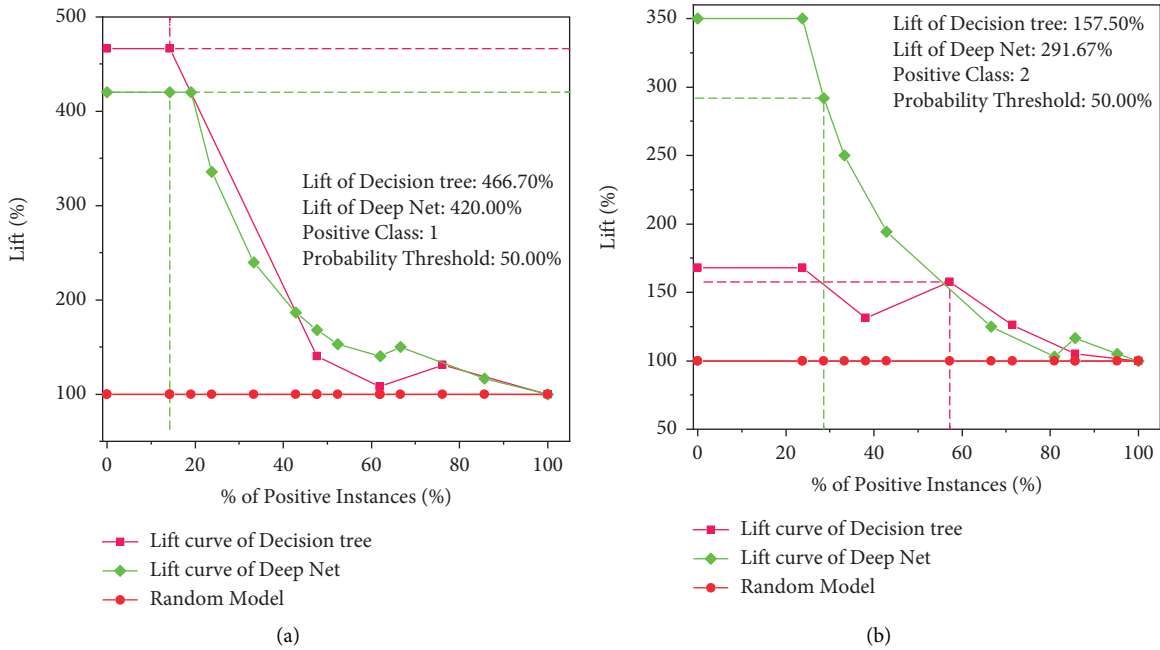


FIGURE 17: Continued.

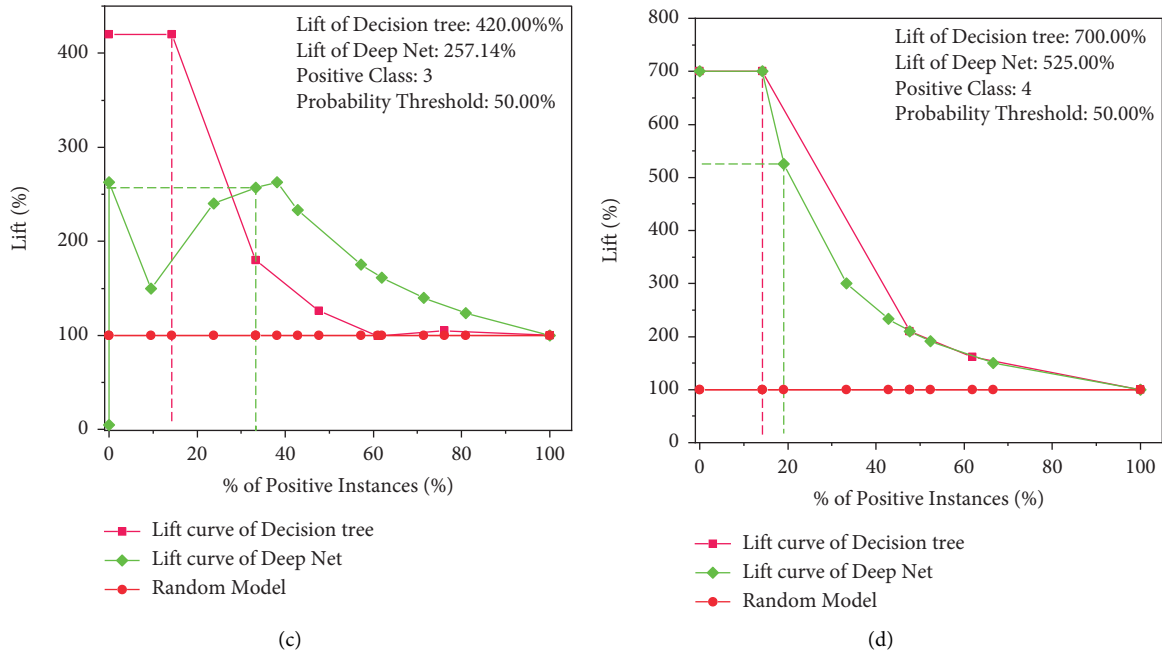


FIGURE 17: Lift curve of each classification test of the deep network model.

classification evaluation model of top coal caving established by deep network and the classification evaluation model of top coal caving established by decision tree have little difference in global performance, the classification evaluation model of top coal caving established by deep network is better than that established by decision tree model in local. In order to get a more robust classification evaluation model of top coal caving, the combination and average of the two models can be considered.

6.2. Optimization of Top Coal Caving Classification Evaluation Model. In the modeling process, it is not difficult to find that, in general, if several different models can be combined and their prediction results are averaged, the ideal prediction results can be obtained. At the same time, if the combined average model can balance the shortcomings of a single participating combined model, then the final model generally obtained is robust and stable. However, it is based on this idea to develop a fusion modeling method in BigML. The fusion modeling method combines different models and averages their predictions to balance the weaknesses of each model so that the model can produce better performance. The principle is similar to model integration, except that the fusion modeling method can combine and average a single decision tree and combine and average models such as logistic regression and deep network.

In order to optimize the model and get a more robust and stable classification evaluation model of top coal caving, the classification evaluation model of top coal caving established by decision tree and depth network is fused. According to the model performance evaluation parameters, although the global performance of the classification evaluation model of top coal caving established by decision tree and deep network is similar,

the local performance of the model established by the deep network is better than that of the model established by a decision tree. Therefore, the weight of the prediction result of the model established by the decision tree and depth network is 1 : 3. After the model fusion, the model's performance is tested and evaluated with the sample data of the test set, and the test sampling mode is set to replaceable sampling. The test results are shown in Figures 18–21 and Tables 10 and 11.

According to Tables 10 and 11 and Figures 18–21, it can be seen that under the probability threshold value of 50%, the classification evaluation model of top coal caving established by fusion is perfect and robust no matter from the global or local view, which has fully met the demand of prediction. The global prediction ACCURACY, AVG. RECALL, and AVG. PRECISION of the classification evaluation model of top coal caving established by fusion reached 90.45%, 95.45%, and 88.75%, respectively. In addition, the AVG. Phi and AVG. F of the classification evaluation model of top coal caving established by fusion reached 0.8838 and 0.9115, respectively. From the ROC AUC, PR AUC, and K-S values of each grade in the ROC curve (Figure 18), PR curve (Figure 19), and K-S curve (Figure 20) in model local performance evaluation parameters, the ROC AUC, PR AUC and K-S values of each grade of the classification evaluation model of top coal caving established by fusion are greater than or equal to that established by decision tree and depth network, respectively. At the same time, from the model's lift curve (Figure 21), the lift value of each evaluation grade is greater than 100%, which also shows that the model's prediction ability established by fusion is stronger than the random model. In addition, through the ROC AUC, PR AUC, and K-S values of each grade of the classification evaluation model of top coal caving established by fusion, it

TABLE 8: Confusion matrix of deep network model test results and model evaluation related indicators.

Actual vs. predicted	1	2	3	4	Actual	Recall (%)
1	3	1	1	0	5	60.00
2	0	5	0	1	6	83.33
3	0	1	6	0	7	86.71
4	0	0	0	3	3	100.00
Predicted	3	7	7	4	21	82.26
					83.04	AVG.RECALL
Precision (%)	100.00	71.43	86.71	75.00	AVG.PRECISION	80.95
						ACCURACY

TABLE 9: Performance evaluation results of the deep network model.

Grade	F-measure	Phi coefficient
1	0.75	0.73
2	0.77	0.67
3	0.86	0.79
4	0.86	0.84
	0.81	0.7436
	AVG. F	AVG. Phi

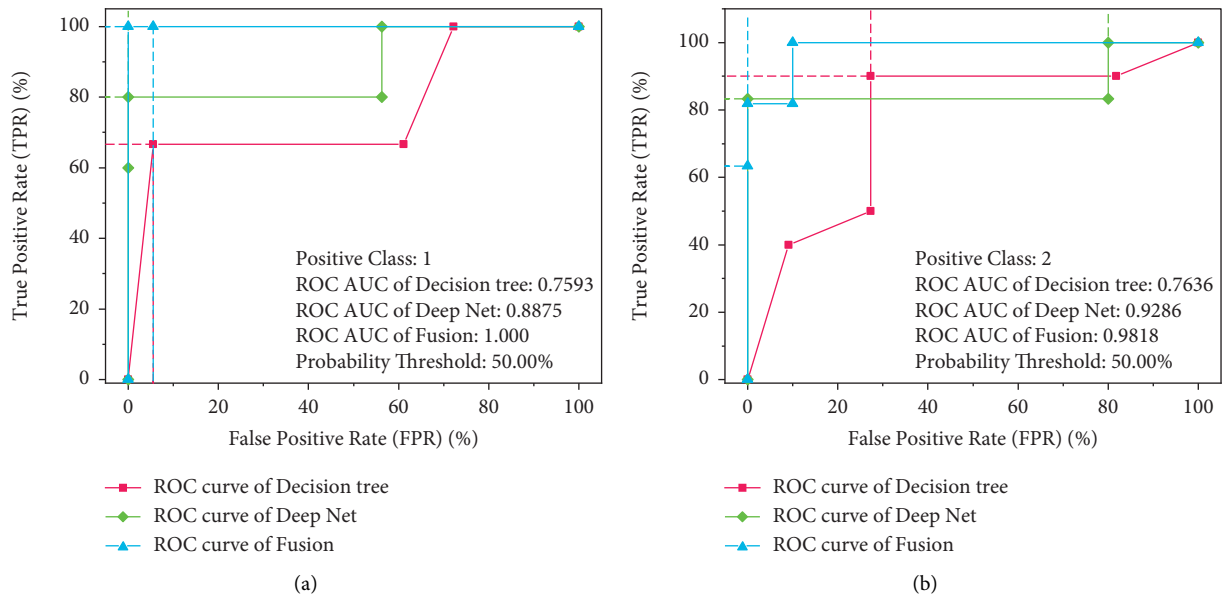


FIGURE 18: Continued.

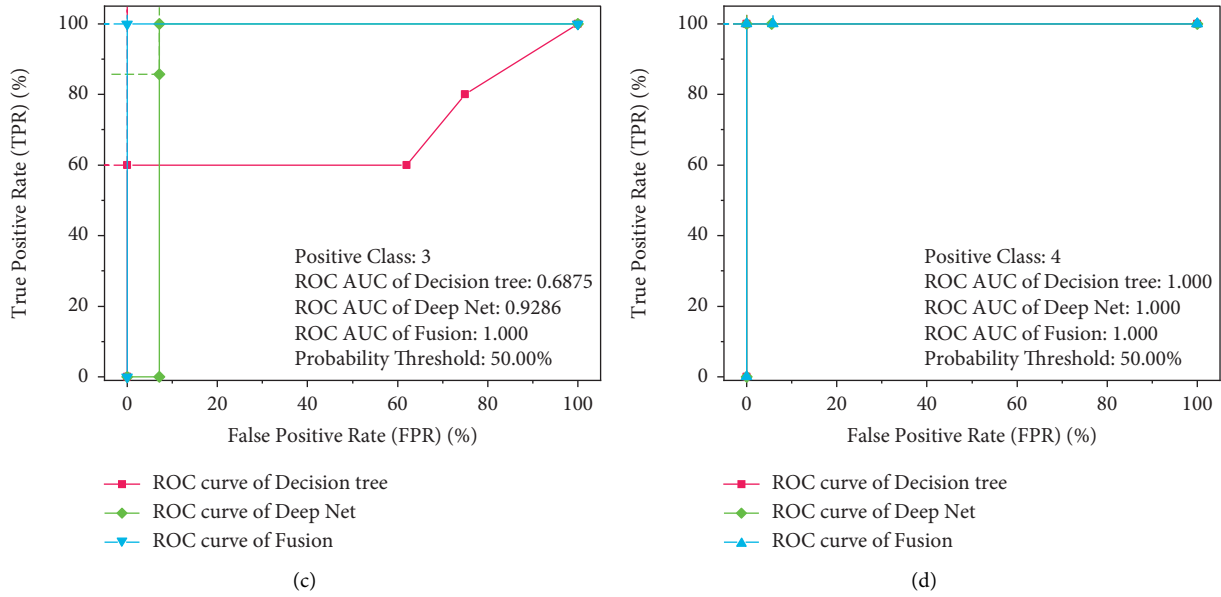


FIGURE 18: ROC curve of each classification test of the fusion model.

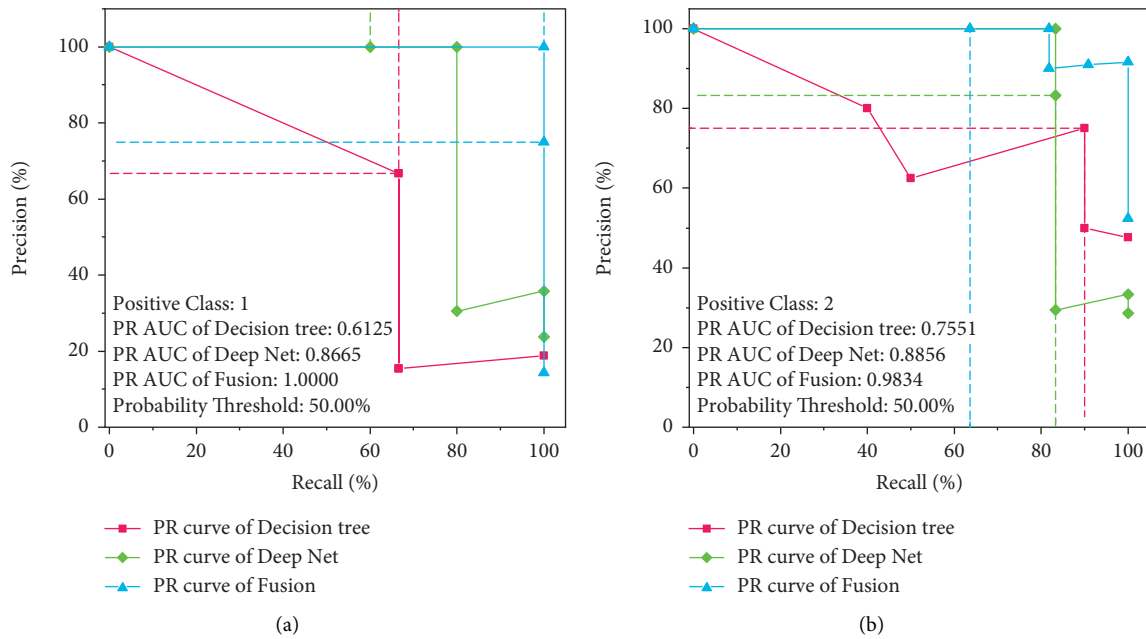


FIGURE 19: Continued.

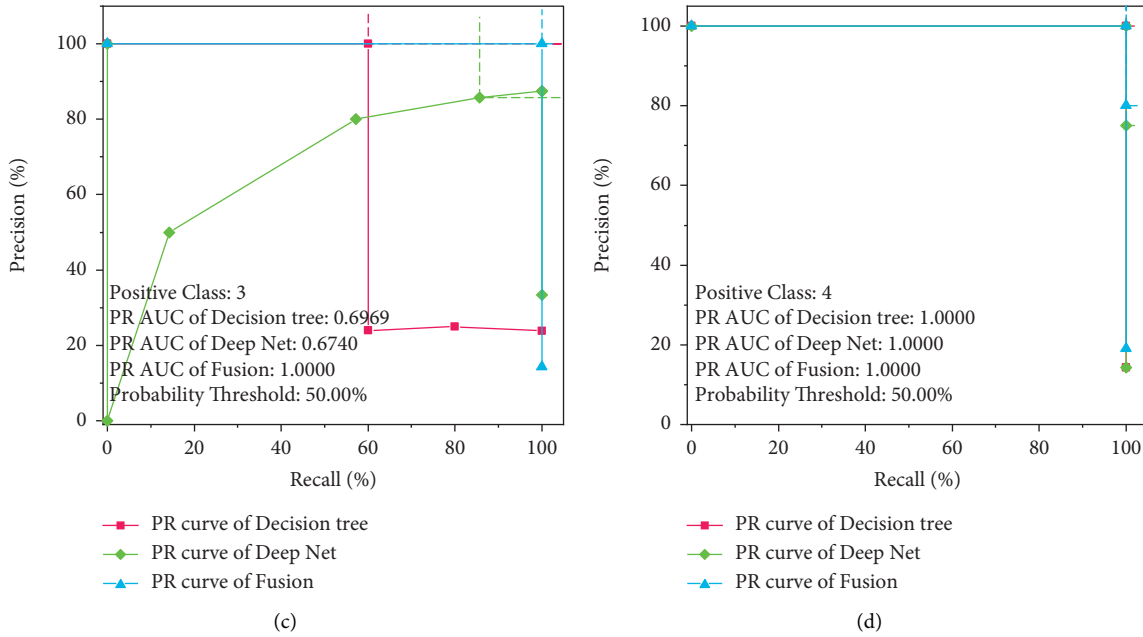


FIGURE 19: PR curve of each classification test of the fusion model.

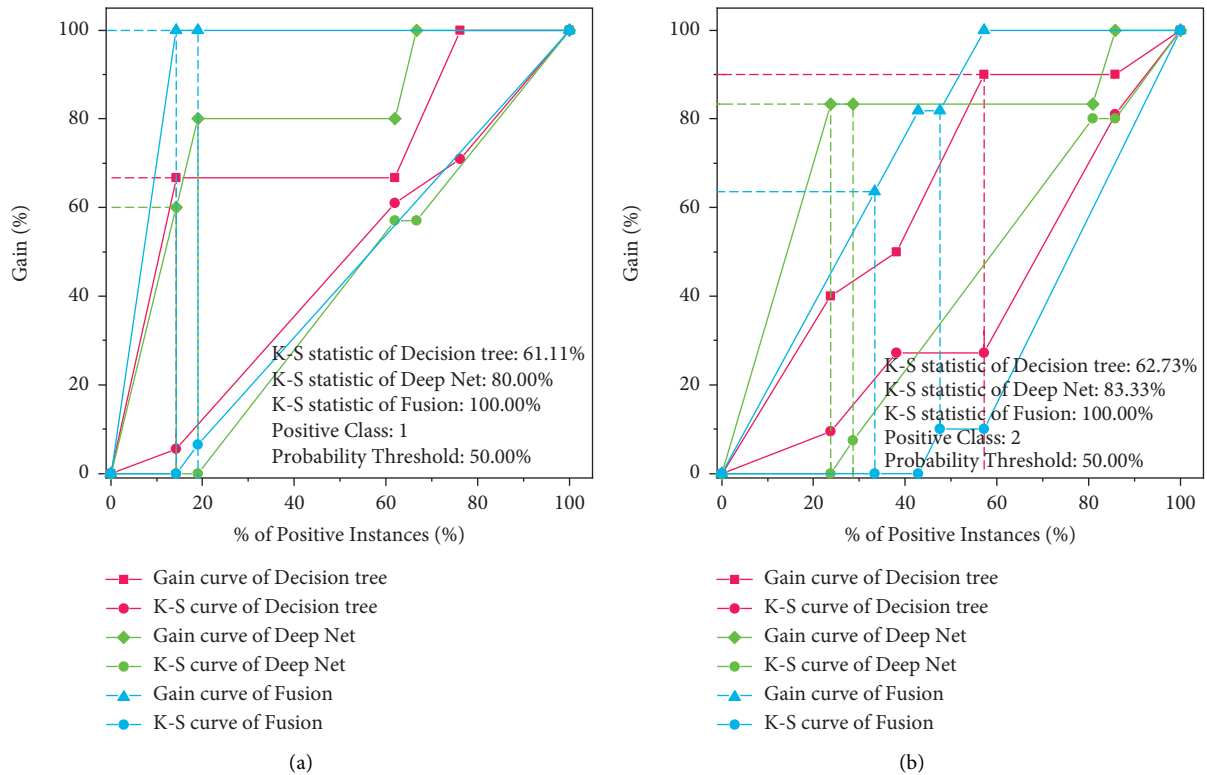


FIGURE 20: Continued.

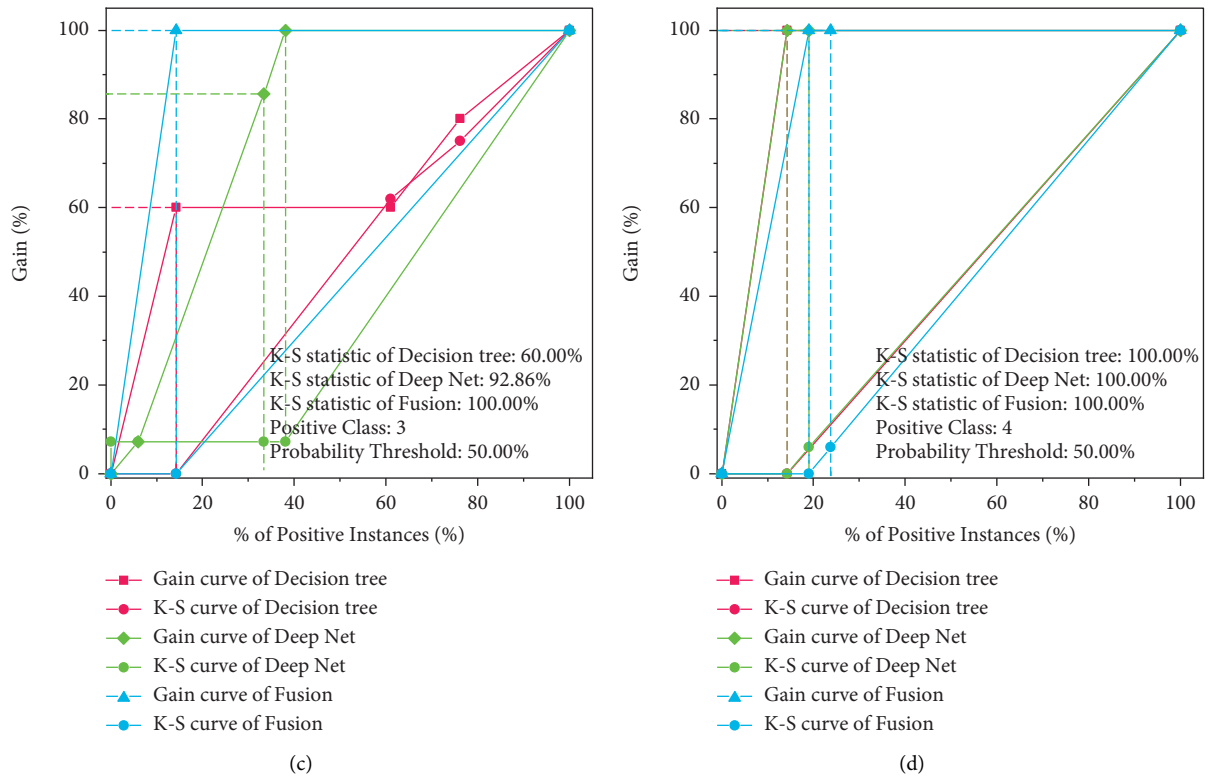


FIGURE 20: K-S curve and gain curve of each classification test of the fusion model.

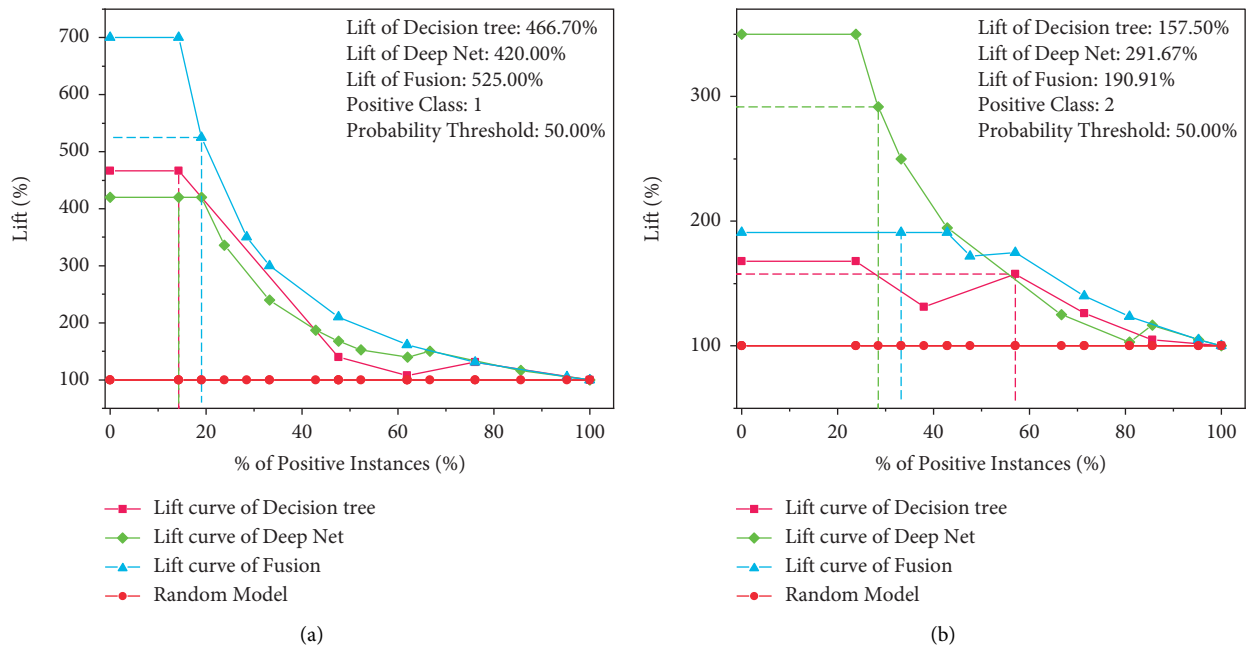


FIGURE 21: Continued.

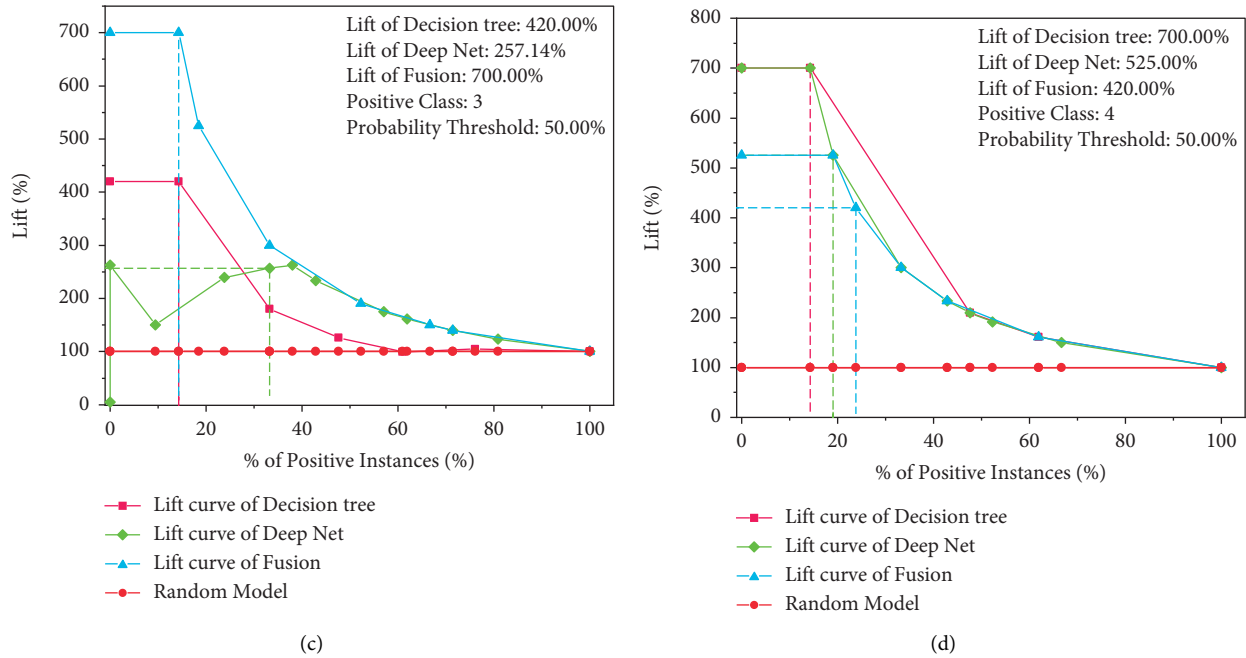


FIGURE 21: Lift curve of each classification test of the fusion model.

TABLE 10: Confusion matrix of fusion model test results and model evaluation related indicators.

Actual vs. predicted	1	2	3	4	Actual	Recall (%)
1	3	0	0	0	3	100.00
2	1	9	0	1	11	81.82
3	0	0	3	0	3	100.00
4	0	0	0	4	4	100.00
Predicted	3	7	7	4	21	96.45
Precision (%)	75.00	100.00	100.00	80.00	88.75	90.45
					AVG.PRECISION	ACCURACY

TABLE 11: Performance evaluation results of the fusion model.

Grade	F-measure	Phi coefficient
1	0.86	0.84
2	0.90	0.83
3	1.00	1.00
4	0.89	0.87
	0.9115	0.8838
	AVG. F	AVG. Phi

can be seen that the model has good prediction ability for the top coal caving of each grade. The ROC AUC of each model grade is greater than 0.9, the PR AUC is basically greater than or equal to 0.9, and the K-S value is 100%. The above analysis shows that under the probability threshold of 50%, the classification evaluation model of top coal caving established by fusion is perfect and robust and has fully met the prediction needs, whether from the global or local point of view.

7. Practical Application of Prediction Model in Engineering

Because the evaluation model of top coal caving classification is optimized by the fusion method, the model is perfect and robust no matter from the global or local view and has fully met the demand of prediction, so the model is applied to the evaluation of top coal caving of No. 3 coal seam in Gucheng Coal Mine. Gucheng Coal Mine belongs to Lu'an

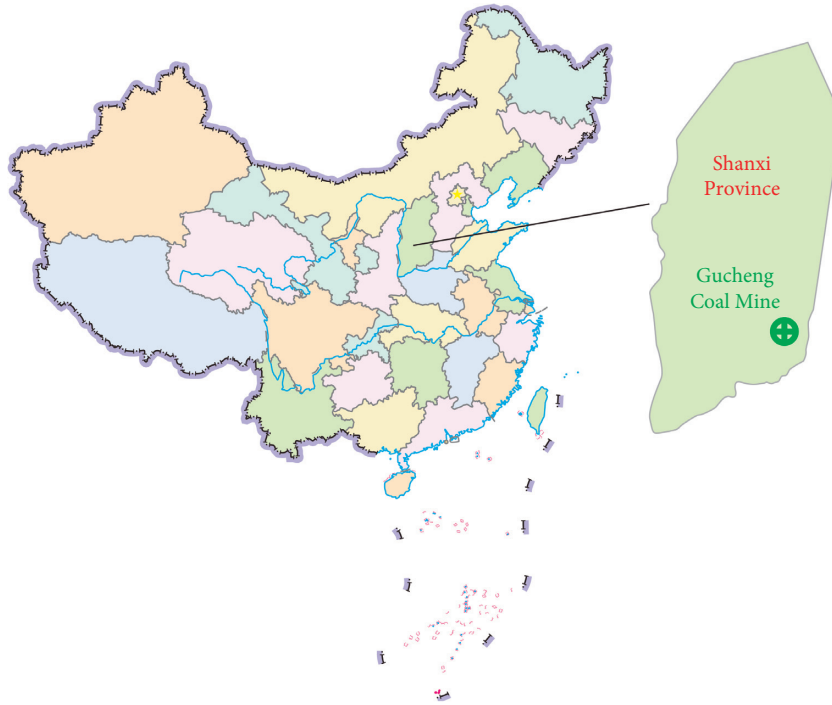


FIGURE 22: Geographical location of Gucheng Mine.

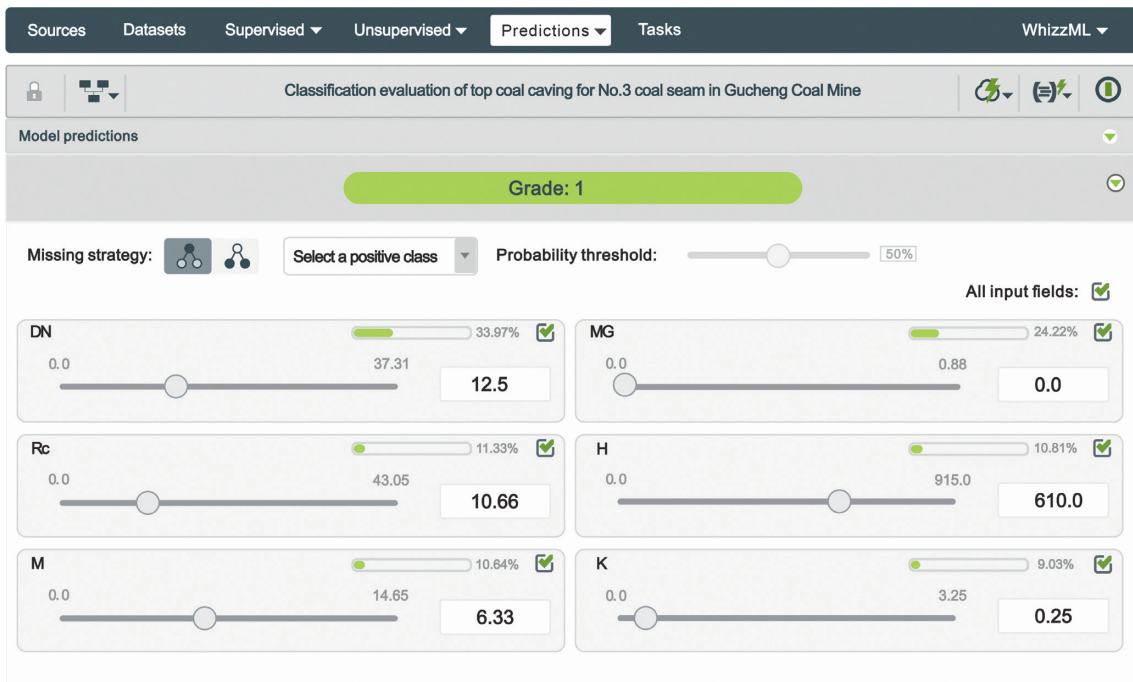


FIGURE 23: Top coal caving prediction results of No. 3 coal seam in Gucheng Coal Mine.

(Group) Co., Ltd., located west of Changzhi City, Shanxi Province, China. Its geographical location is shown in Figure 22. The average thickness of No.3 Coal Seam in Gucheng Coal Mine is 6.33 m, without gangue; The buried depth of coal seam is 583–637 m, with an average of 610 m; the average uniaxial compressive strength of coal is

10.66 MPa, which belongs to medium-hard coal. The top coal can be mined with caving, and the filling coefficient of the direct roof is 0.25; the development degree of coal seam joints is good ($DN = 12.5$). The prediction result shows that the top coal caving grade of No.3 Coal Seam in Gucheng Coal Mine is grade 1, as shown in Figure 23. From the

engineering side, the predicted results are consistent with the actual situation. In the mining of No. 3 coal seam in Gucheng Coal Mine, the top coal can cave well without extra measures.

8. Conclusion

Because of the current, most of the evaluation and prediction models of top coal caving established by experts and scholars are established by computer programming, which makes it difficult for people who do not understand the calculation language to use or modify the models and makes it difficult for these models to be widely applied in the actual application process. This article introduces a method to establish the evaluation and prediction model of top coal caving without programming. At the same time, the model can be used to predict and evaluate top coal caving and modify the model according to its own needs without programming. This method establishes the prediction model of top coal caving by using the machine learning platform BigML based on the cloud. At the same time, this paper establishes the prediction model of top coal caving evaluation by using BigML and applies it to evaluating top coal caving of No. 3 coal seam in Gucheng Coal Mine. The evaluation result is grade 1, which is consistent with the engineering practice. It fully proves that the application of BigML in evaluating top coal caving is successful and feasible and provides another more convenient method for the classification evaluation and prediction of top coal caving. In addition, it provides another way to realize the classification evaluation of top coal caving properties and the establishment of other evaluation predictions using machine learning without programming.

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 work was supported by the discipline innovation team of Liaoning Technical University (LNTU20TD-01).

References

- [1] By Fuel Type-Exajoules, Consumption Emissions and C. Dioxide, "BP statistical review of world energy 2020," 2020.
- [2] M. D. Leonard, E. E. Michaelides, and D. N. Michaelides, "Substitution of coal power plants with renewable energy sources - shift of the power demand and energy storage," *Energy Conversion and Management*, vol. 164, pp. 27–35, 2018.
- [3] B. Dudley, "BP statistical review of world energy 2016," 2016.
- [4] J. Wang, B. Yu, H. Kang et al., "Key technologies and equipment for a fully mechanized top-coal caving operation with a large mining height at ultra-thick coal seams," *International Journal of Coal Science & Technology*, vol. 2, no. 2, pp. 97–161, 2015.
- [5] N. Zhang, C. Liu, and M. Pei, "Effects of caving-mining ratio on the coal and waste rocks gangue flows and the amount of cyclically caved coal in fully mechanized mining of super-thick coal seams," *International Journal of Mining Science and Technology*, vol. 25, no. 1, pp. 145–150, 2015.
- [6] M. Bhattacharya, S. Rafiq, and S. Bhattacharya, "The role of technology on the dynamics of coal consumption-economic growth: new evidence from China," *Applied Energy*, vol. 154, pp. 686–695, 2015.
- [7] Z. Wang, G. Zhang, and L. Zhao, "Recognition of rock-coal interface in top coal caving through tail beam vibrations by using stacked sparse autoencoders," *Journal of Vibroengineering*, vol. 18, no. 7, pp. 4261–4275, 2016.
- [8] J. Wu, Y. Qin, and M. Zhai, "Mining safety of longwall top-coal caving in China," in *Proceedings of the 8th U.S. Mine Ventilation Symposium*, Rolla, Missouri, June 1999.
- [9] T. D. Le, R. Mitra, J. Oh, and B. Hebblewhite, "A review of cavability evaluation in longwall top coal caving," *International Journal of Mining Science and Technology*, vol. 27, no. 6, pp. 907–915, 2017.
- [10] N. E. Yasitli and B. Unver, "3D numerical modeling of longwall mining with top-coal caving," *International Journal of Rock Mechanics and Mining Sciences*, vol. 42, no. 2, pp. 219–235, 2005.
- [11] H. Alehossein and B. A. Poulsen, "Stress analysis of longwall top coal caving," *International Journal of Rock Mechanics and Mining Sciences*, vol. 47, no. 1, pp. 30–41, 2010.
- [12] G. Si, S. Jamnikar, J. Lazar et al., "Monitoring and modelling of gas dynamics in multi-level longwall top coal caving of ultra-thick coal seams, part I: borehole measurements and a conceptual model for gas emission zones," *International Journal of Coal Geology*, vol. 144–145, pp. 98–110, 2015.
- [13] G. Si, J.-Q. Shi, S. Durucan et al., "Monitoring and modelling of gas dynamics in multi-level longwall top coal caving of ultra-thick coal seams, Part II: numerical modelling," *International Journal of Coal Geology*, vol. 144–145, pp. 58–70, 2015.
- [14] V. Klishin, S. Nikitenko, and G. Opruk, "Longwall top coal caving (LTCC) mining technologies with roof softening by hydraulic fracturing method," *IOP Conference Series: Materials Science and Engineering*, vol. 354, Article ID 012015, 2018.
- [15] V. I. Klishin, V. N. Fryanov, L. D. Pavlova, and G. Y. Opruk, "Modeling top coal disintegration in thick seams in longwall top coal caving," *Journal of Mining Science*, vol. 55, no. 2, pp. 247–256, 2019.
- [16] Z. Ti, J. Li, M. Wang, K. Wang, Z. Jin, and C. Tai, "Fracture mechanism in overlying strata during longwall mining," *Shock and Vibration*, vol. 2021, Article ID 4764732, 15 pages, 2021.
- [17] Y. K. Shi, T. T. Sun, J. Hao, and S. Hao, "Prediction of top coal caving ability based on support vector machine," *Applied Mechanics and Materials*, vol. 217–219, pp. 2702–2705, 2012.
- [18] A. Vakili and B. K. Hebblewhite, "A new cavability assessment criterion for longwall top coal caving," *International Journal of Rock Mechanics and Mining Sciences*, vol. 47, no. 8, pp. 1317–1329, 2010.
- [19] P. Humphries, B. Poulsen, and T. Ren, "Longwall top coal caving application assessment in Australia," *ACARP Report for project C*, vol. 130187, 2007.
- [20] B. Singh and R. Goel, *Engineering Rock Mass Classification*, Elsevier, Amsterdam, Netherlands, 2011.

- [21] Y.-S. Xie and Y.-S. Zhao, "Numerical simulation of the top coal caving process using the discrete element method," *International Journal of Rock Mechanics and Mining Sciences*, vol. 46, no. 6, pp. 983–991, 2009.
- [22] J. Wang, S. Yang, Y. Li, L. Wei, and H. Liu, "Caving mechanisms of loose top-coal in longwall top-coal caving mining method," *International Journal of Rock Mechanics and Mining Sciences*, vol. 71, pp. 160–170, 2014.
- [23] Y. Xin, P. Gou, and F. Ge, "Analysis of stability of support and surrounding rock in mining top coal of inclined coal seam," *International Journal of Mining Science and Technology*, vol. 24, no. 1, pp. 63–68, 2014.
- [24] K. Yu, W. Qiang, M. Elhoseny, and X. Yuan, "Application of ant colony clustering algorithm in coal mine gas accident analysis under the background of big data research," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 2, pp. 1381–1390, 2020.
- [25] Z. Ruilin and I. S. Lowndes, "The application of a coupled artificial neural network and fault tree analysis model to predict coal and gas outbursts," *International Journal of Coal Geology*, vol. 84, no. 2, pp. 141–152, 2010.
- [26] H. Jiang, Q. Song, K. Gao, Q. Song, and X. Zhao, "Rule-based expert system to assess caving output ratio in top coal caving," *PloS One*, vol. 15, no. 9, Article ID e0238138, 2020.
- [27] S. Mohammadi, M. Ataei, R. Kakaie, A. Mirzaghobanali, Z. F. Rad, and N. Aziz, "A roof cavability classification system and its use for estimation of main caving interval in longwall mining," in *Proceedings of the 2020 Coal Operators' Conference*, pp. 104–115, Wollongong, Australia, February 2020.
- [28] S. Yongkui, L. Pengrui, W. Ying, Z. Jingyu, and L. Meijie, "The prediction of the caving degree of coal seam roof based on the naive bayes classifier," *Electronic Journal of Geotechnical Engineering*, vol. 19, no. Z2, p. 201, 2014.
- [29] K. Oraee and M. Rostami, "Qualitative and quantitative analysis of hangingwall caving in longwall mining method using a fuzzy system," in *Proceedings of the 21st world mining congress & expo*, Poland, Kraków-Katowice-Sosnowiec, September 2008.
- [30] J. H. Yu and D. B. Mao, "Prediction of top-coal caving and drawing characteristics using artificial neural networks in extremely thick coal seam," *Applied Mechanics and Materials*, vol. 743, pp. 612–616, 2015.
- [31] G. A. Rao, R. Nagaswetha, and D. N. Singh, "Voice based virtual agri farming analyzer with BigML algorithms," in *Proceedings of the 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, Moradabad, India, November 2019.
- [32] Y. Abdul Ghffar, M. Osman, S. Shrestha et al., "Usefulness of semisupervised machine-learning-based phenogrouping to improve risk assessment for patients undergoing transcatheter aortic valve implantation," *The American Journal of Cardiology*, vol. 136, pp. 122–130, 2020.
- [33] X. Tian and L. Liu, "Does big data mean big knowledge? integration of big data analysis and conceptual model for social commerce research," *Electronic Commerce Research*, vol. 17, no. 1, pp. 169–183, 2017.
- [34] M. Khanal, D. Adhikary, and R. Balusu, "Prefeasibility study-geotechnical studies for introducing longwall top coal caving in Indian mines," *Journal of Mining Science*, vol. 50, no. 4, pp. 719–732, 2014.
- [35] P. Humphries and B. Poulsen, "Geological and geotechnical influences on the caveability and drawability of top coal in longwalls," *Sociological Research Online*, pp. 56–66, 2008.
- [36] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," *2008 Eighth IEEE International Conference on Data Mining*, in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, Washington, DC, USA, December 2008.
- [37] T. B. Team, *Anomaly Detection with the BigML Dashboard*, BigML, Inc, Corvallis, OR, USA, 2021.
- [38] H. Mohammad and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 1–11, 2015.
- [39] K. Stapor, "Evaluating and comparing classifiers: review, some recommendations and limitations," in *Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017*, pp. 12–21, Polanica Zdroj, Poland, May 2017.
- [40] Y. Sasaki, "The truth of the F-measure," *Teach Tutor Mater*, pp. 1–5, 2007.
- [41] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using matthews correlation coefficient metric," *PloS One*, vol. 12, no. 6, Article ID e0177678, 2017.
- [42] J. Lever, M. Krzywinski, and N. Altman, "Classification evaluation," *Nature Methods*, vol. 13, no. 8, pp. 603–604, 2016.
- [43] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [44] C. X. Ling, J. Huang, and H. Zhang, "AUC: a better measure than accuracy in comparing learning algorithms," *Advances in Artificial Intelligence*, in *Proceedings of the Conference of the canadian society for computational studies of intelligence*, pp. 329–341, Calgary, AB, Canada, May 2003.
- [45] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, Pittsburgh, Pennsylvania, June 2006.
- [46] T. Brandenburger and A. Furth, "Cumulative gains model quality metric," *Journal of Applied Mathematics and Decision Sciences*, vol. 2009, Article ID 868215, 14 pages, 2009.
- [47] T. M. Khoshgoftaar, M. Golawala, and J. V. Hulse, "An empirical study of learning from imbalanced data using random forest," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, vol. 2, Patras, Greece, October 2007.
- [48] M. Kuhn and K. Johnson, "Measuring performance in classification models," *Applied Predictive Modeling*, Springer, New York, NY, pp. 247–273, 2013.
- [49] M.-H. Horng, "Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ML, RBFNN and SVM classifiers," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4146–4155, 2010.
- [50] T. Srivastava, *11 Important Model Evaluation Metrics for Machine Learning Everyone Should Know*, <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>, 2019.
- [51] N. Seliya, T. M. Khoshgoftaar, and J. Van Hulse, "A study on the relationships of classifier performance metrics," in *Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence*, Washington, DC, USA, November 2009.
- [52] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "A survey on graphical methods for classification predictive performance evaluation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1601–1618, 2011.
- [53] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, "Finding the best classification threshold in imbalanced classification," *Big Data Research*, vol. 5, pp. 2–8, 2016.