International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3214-3224

https://doi.org/10.70517/ijhsa463270

Research on Corporate Culture Remodeling and Organizational Structure Adaptation Driven by Digital Transformation

Ye Wu[1,*], Chuan Tian[1] and Huan Wan[1]
[1] Department of Economic Management, Jiangxi Tourism & Commerce Vocational College, Nanchang, Jiangxi, 330100, China
Corresponding authors: (e-mail: 15180560090@163.com).

Abstract Organizational structure is the most basic framework of an enterprise and an important part of enterprise management, and excellent enterprise culture is a huge intangible asset of an enterprise, which is a valuable spiritual wealth to unite all the employees and achieve excellence in the enterprise. The study summarizes the influencing factors of corporate culture reshaping and organizational structure adaptation, uses Spearman correlation coefficient to analyze the correlation of independent variables, and finally identifies 12 characteristic variables with insignificant correlation as independent variables. In order to improve the prediction accuracy of corporate culture remodeling and organizational structure fitness, this paper proposes a combined machine learning model based on SMOTE-XGBoost, comparing with logistic regression, KNN, decision tree, and random forest in machine learning, the SMOTE-XGBoost model for this paper has a better prediction effect in terms of the fitness precision rate, recall rate, and F1 score increased by 50%, 18%, and 38%, respectively. , with better prediction effect. The SHAP model is used to analyze the important influencing factors of corporate culture remodeling and organizational structure fitness, and it is concluded that behavioral performance, information technology enhancement, enterprise scale development strategy, market competition and skill mastery are the five characteristic indicators with the greatest influence.

Index Terms corporate culture, SMOTE-XGBoost algorithm, fitness, enterprise development, SHAP

I. Introduction

With the rapid development of digital technology, more and more enterprises are accelerating the pace of digital transformation [1], [2]. Digital transformation refers to the transformation of traditional business processes into new models and approaches driven by digital technologies in order to achieve enterprise innovation and growth [3], [4]. Through digitalization, companies can make better use of data and technology to improve business efficiency, quality, and customer satisfaction, which in turn contributes to the overall development of the organization, and at the same time, this transformation puts forward new requirements for cultural reshaping and organizational structure adaptation [5]-[8].

Digital transformation has had a profound impact on corporate culture [9]. By reshaping corporate culture, companies can stimulate innovation and change, and improve organizational adaptability and competitiveness [10], [11]. In order to achieve digital transformation and reshape corporate culture, organizations need to build the role of leaders, provide training and education, create an environment for innovation, build a data-driven culture, implement agile methods and cross-functional collaboration, and build a customer-oriented culture [12]-[15]. Through these practical approaches, organizations can successfully achieve digital transformation and sustained business success in the digital era [16], [17]. Enterprise digital transformation often requires organizational restructuring and redesign to enable rapid advancement of digital processes and digital business [18]-[20]. Enterprises need to establish a team dedicated to digital transformation, including technologists and strategic planners [21], [22]. And to realize the organizational structure should be flattened to improve the efficiency of information flow and break down the barriers between departments [23]. In addition, internal and external synergies should be strengthened to promote digital transformation with partners [24].

This paper constructs the influence factor indicators of corporate culture remodeling and organizational structure fitness, adopts Spearman correlation coefficient for data preprocessing, and determines 12 influence factor indicators as independent variables. This paper proposes the SMOTE-XGBoost algorithm applicable to the prediction method of corporate culture reinvention and organizational structure fitness, and compares it with machine learning models such as linear logistic regression, K-Nearest Neighbor, decision tree, etc., through the introduction of the precision, accuracy, recall, F1 value and prediction result of each model on the test set as the evaluation indexes, and by introducing the SHAP model to enhance the interpretability of the prediction models, and feature analysis to identify the main factors affecting the fitness of corporate culture remodeling and organizational structure.

II. Predictive model of corporate culture reshaping and organizational fit

II. A. Factors affecting the reshaping of corporate culture and the adaptation of organizational structure

The study concluded that corporate culture reshaping is mainly affected by changes in internal and external environments, such as leaders' behaviors, changes in property rights structure, corporate development strategies, as well as market and political factors. The change of enterprise organization structure is a kind of systematic adjustment and innovation with purpose, which is the effective embodiment of organization, division and arrangement of employees' tasks in the enterprise, and it is also affected by the influence of internal and external environments and thus produces changes, and the external factors are mainly the personal knowledge economy, the development of information technology, and the competition in the market, and the internal factors include the enterprise development strategy, the development of the scale of the enterprise, the technology of the enterprise process, the culture of the enterprise organization, etc. Internal factors include enterprise development strategy, enterprise scale development, enterprise technology, enterprise organization culture, etc. Combined with the above, scholars believe that the influence of corporate culture reshaping and organizational structure adaptability is mainly affected by the following factors.

(1) Internal factors: leaders' behavioral knowledge, information technology development, development strategy.

(2) External factors: market development trends, employee skills, and policy reforms.

II. B. Selection of independent variables

II. B. 1) Pre-processing of independent variables

This paper combines the above to select 12 influencing factors from the six factors as independent variables. Corporate culture reinvention and organizational structure fitness are taken as dependent variables. The coding of the influencing factors of corporate culture reinvention and organizational structure fitness is shown in Table 1.

Table 1: Impact factor encoding

Categories	Variable name	Coding
Leadership behavior knowledge	Behavioral expression	X1
	Knowledge storage	X2
Information technology development	Network development	X3
	Information technology improvement	X4
Development strategy	Enterprise development strategy	X5
	Enterprise scale development strategy	X6
Market trends	Market competition	X7
	Market product sales	X8
Employee technology	Technical technology	X9
	Technical skill mastery	X10
Policy reform	Government agency	X11
	Administrative management	X12

II. B. 2) Correlation analysis of independent variables

Correlation analysis is a statistical method that is mainly used to assess the linear relationship between two or more variables. Through this analysis, the researcher can understand whether there is some degree of association between the variables, as well as the strength and direction of this association. It is mainly represented through graphical and numerical representations. There are three types of numerical representations, namely the Pearson correlation coefficient, the Kendall τ correlation coefficient, and the Spearman correlation coefficient; whereas graphical representations are mainly used to visualize the relationship between two continuous variables in the form of scatter plots, which can help to identify the existence of a linear relationship between the variables, the strength of the relationship, and the potential trend through the pattern of the distribution of data points. Since Spearman's correlation coefficient is a nonparametric statistical method, it does not require any assumptions about the distribution of the data, which makes it more flexible and precise when dealing with non-normally distributed or ordered categorical data [25]. At the same time, compared with the Pearson correlation coefficient, the Spearman correlation coefficient is more suitable for dealing with data that satisfy the assumption of normal distribution and has less influence on outliers, so this paper chooses to use the Spearman correlation coefficient to analyze the independent variables.

The formula of Spearman correlation coefficient is shown in (1):

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}(R(x_i)-\overline{R(x)})\cdot(R(y_i)-\overline{R(y)})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(R(x_i)-\overline{R(x)})^2\right)\cdot\left(\frac{1}{n}\sum_{i=1}^{n}(R(y_i)-\overline{R(y)})^2\right)}} \tag{1}$$

where $R(x)$ and $R(y)$ - the rank of x and y.

$\overline{R(x)}$ and $\overline{R(y)}$ - the average rank of x and y.

The results of the Spearman's correlation coefficient made by matlab are plotted as a correlation matrix heat map1 as shown below. From the figure, it can be concluded that the correlation indexes of the independent variable influencing factors, enterprise scale development strategy and behavioral performance, market product sales and network development reached 0.72 and 0.75 respectively, which have strong correlation.

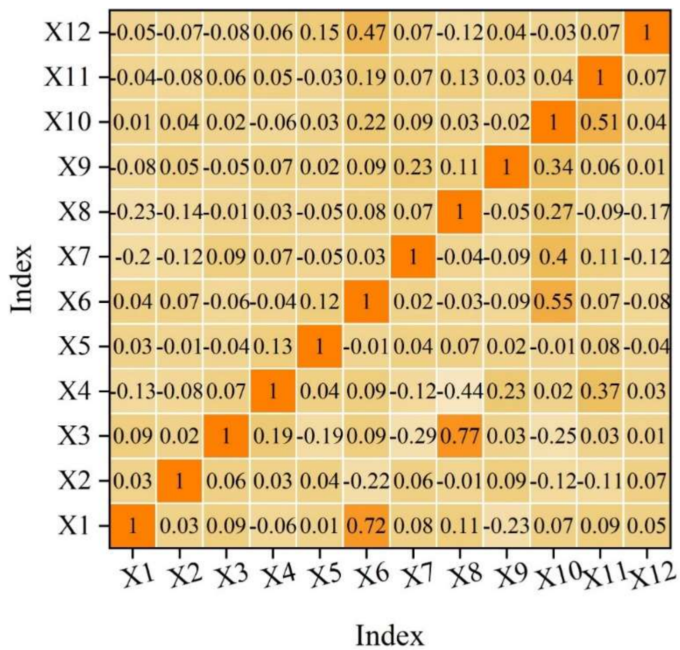


Figure 1: Spearman Correlation Coefficient Matrix Heatmap

II. C.SMOTE-XGBoost algorithm based fitness prediction

II. C. 1) SMOTE algorithm

In order to solve the problem of category imbalance, CHAWLA et al. proposed the SMOTE algorithm (i.e., oversampling technique) based on artificially synthesizing data from a few categories.The basic idea of the SMOTE algorithm is to analyze the samples of a few categories, and then artificially synthesize new samples to be added to the dataset based on the few samples of a few categories, the principle of which is shown in Fig. 2.

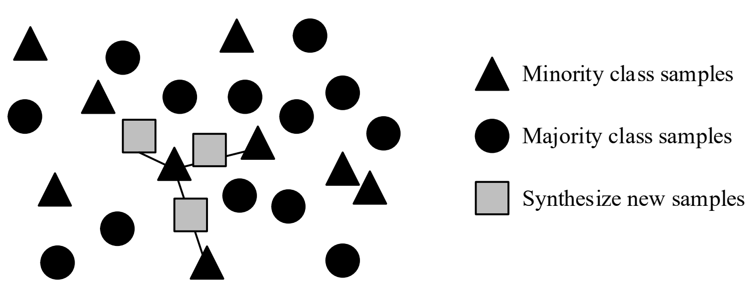


Figure 2: Schematic diagram of the SMOTE algorithm

The SMOTE algorithm can solve the imbalance problem of the dataset, compared with the random oversampling which is easy to make the model overfitting.The SMOTE algorithm samples for the feature space, and the accuracy is higher than the traditional sampling method.

The principle of SMOTE algorithm is as follows.

(1) For each sample x in the minority class, calculate the Euclidean distance between the point and other sample points in the minority class to get the nearest k nearest neighbors.

(2) Determine the sampling multiplicity according to the number of new samples required, and for each sample x in the minority class, randomly select N samples from its k nearest neighbors, denoted as $x_1^{'}, x_2^{'}, x_3^{'}, \cdots\cdots, x_N^{'}$.

(3) A new minority class sample x_{new} is constructed by performing a random linear difference between the minority class sample x and a randomly selected sample $x_j^{'}(j=1,2,\cdots\cdots N)$ in k -nearest neighbors, see equation (2).

$$x_{new} = x + rand(0,1) \times (x_j^{'} - x) \tag{2}$$

where: $rand(0,1)$ is a random number in the interval (0, 1).

The objective function $O_{bj}(\varphi)$ of the XGBoost algorithm is shown in equation (3).

$$O_{bj}(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{3}$$

where: $l(y_i, \hat{y}_i)$ is the loss function; y_i is the labeled value; $\hat{y}_i$ is the predicted value; $\Omega(f_t)$ is the regularization term, which is used to control the complexity of the model in order to prevent the occurrence of overfitting; f_t is the t th tree model.

The $\Omega(f_t)$ in the objective function $O_{bj}(\varphi)$ is shown in equation (4).

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

where: γ is the shrinkage coefficient; T is the number of leaf nodes; ω is the weight value of the leaf nodes; λ is the penalty term coefficient; ω_j is the weight of the j th leaf.

In the XGBoost algorithm, the newly added tree needs to fit the residual values of the last prediction, then the objective function $O_{bj}(\varphi)$ can be transformed into equation (5).

$$O_{bj}(\varphi) = \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i^{t-1}) + f_t(x_i) \right] + \sum_{i=1}^{t} \Omega(f_t) \tag{5}$$

II. C. 2) Principles of the XGBoost Algorithm

XGBoost algorithm has achieved better prediction results in many prediction fields, which is an optimization of Boosting algorithm, integrating multiple weak classifiers into one strong classifier.

The idea of XGBoost algorithm: first, a simple model is used to fit the data, and after a result is obtained, new trees are continuously added to the model; then, as the number of added trees increases, the accuracy of the prediction model of the XGBoost algorithm will gradually increase, and the second-order Taylor expansion of the residual values of the last prediction is continuously carried out until the accuracy of the prediction model is close to the data itself; Finally, the predicted values corresponding to each tree are added up to be the final predicted value for that sample [26]. The flow is shown in Fig. 3.

The objective function $O_{bj}(\varphi)$ of the XGBoost algorithm is shown in equation (6).

$$O_{bj}(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{6}$$

where: $l(y_i,\hat{y}_i)$ is the loss function; y_i is the labeled value; $\hat{y}_i$ is the predicted value; $\Omega(f_i)$ is the regularization term, which is used to control the complexity of the model in order to prevent the occurrence of overfitting; f_i is the t th tree model.

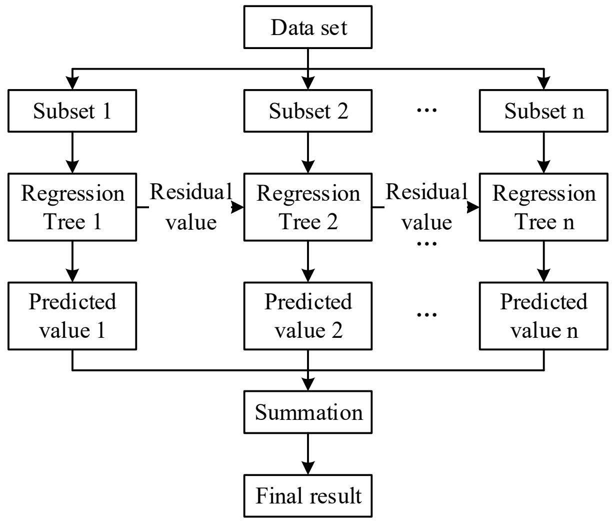


Figure 3: Schematic diagram of XGBoost algorithm

The $\Omega(f_i)$ in the objective function $O_{bj}(\varphi)$ is shown in equation (7).

$$\Omega(f_i)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2 \tag{7}$$

where: γ is the shrinkage coefficient; T is the number of leaf nodes; ω is the weight value of the leaf nodes; λ is the penalty term coefficient; ω_j is the weight of the j th leaf.

In the XGBoost algorithm, the newly added tree needs to fit the residual values of the last prediction, then the objective function $O_{bj}(\varphi)$ can be transformed into Equation (8).

$$O_{bj}(\varphi)=\sum_{i=1}^{n}\left[l(y_i,\hat{y}_i^{t-1})+f_t(x_i)\right]+\sum_{i=1}^{t}\Omega(f_t) \tag{8}$$

where: t is the number of iterations, i.e., the t th tree; $l(\hat{y}_i,\hat{y}_i^{t-1})$ is the loss function after the $t-1$ th round of iteration; $\hat{y}_i^{t-1}$ is the predicted value of the i th sample after the $t-1$ th round of iteration; $f_t(x_i)$ is the predicted value of the newly added DT for the i th sample in the t th round of iteration.

To find the minimized objective function, the loss function in the objective function $O_{bj}(\varphi)$ is subjected to a second-order Taylor expansion, which yields Eq. (9).

$$O_{bj}(\varphi)=\sum_{i=1}^{n}\left[l[y_i,\hat{y}_i^{t-1}]+g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\sum_{i=1}^{t}\Omega(f_i)+C \tag{9}$$

where: g_i, h_i are the first-order and second-order derivatives of $l(y_i,\hat{y}_i^{t-1})$, respectively; C is a constant term; and $l(y_i,\hat{y}_i^{t-1})$ is the predicted value of the first $t-1$ tree, which is the same as that of C None of the terms affects the solution of the minimization of the objective function.

Substituting Eqs. (6)~(7) into Eqs. (8)~(9) leads to Eq. (10).

$$
\begin{aligned}
O_{bj}(\varphi) &= \sum_{i=1}^{n}\left[g_i f_i(x_i)+\frac{1}{2}h_i f_i^2(x_i)\right]+\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{r}\omega_j^2 \\
&= \sum_{j=1}^{T}\left[G_j\omega_j+\frac{1}{2}(H_j+\lambda)\omega_j^2\right]+\gamma T
\end{aligned}
\tag{10}
$$

where $G_j = \sum_{i\in I_j} g_i$; and $H_j = \sum_{i\in I_j} h_i$.

Deriving $G_j\omega_j+\frac{1}{2}(H_j+\lambda)\omega_j^2$ in ω_j and making it equal to 0, the optimal weights w^* can be obtained, as shown in equation (11).

$$
w^* = -\frac{G_j}{H_j+\lambda} \tag{11}
$$

The final objective function $O_{bj}(\varphi)$ is obtained by substituting Eq. (10) into Eq. (11), see Eq. (12).

$$
O_{bj}(\varphi) = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda}+\gamma T \tag{12}
$$

The results of Eq. (6) to Eq. (12) show that when a tree structure is specified, it can reduce the maximum value on the objective.The XGBoost algorithm makes the prediction model less iterative and more accurate by fitting the second-order derivative expansion of the loss function from the previous round.

II. C. 3) SHAP Model Interpretation

Random Forest prediction models can be highly accurate, but their “black box” nature means that the interpretation of the results is weak, for example, it is difficult to explain why the algorithm accurately predicts whether or not a patient will develop a particular disease.

SHAP (Shapley Additive exPlanation) explains the individual predictions of the Random Forest model by observing the effect of each feature on the prediction of a particular sample. The principle of the SHAP model is to generate a prediction value for each individual prediction sample, and the value corresponding to the assignment of the feature to the individual sample is expressed as SHAP value [27]. Assuming that the j th feature of the i th sample is x_{ij} , the model's predicted value for that sample is y_i , and the model's baseline (which defaults to the mean of the target variable for all the samples) is y_{base} , then the SHAP value obeys the following equation:

$$
y_i = y_{base}+f(x_{i1})+f(x_{i2})+\cdots+f(x_{im}) \tag{13}
$$

where $f(x_{ij})$ denotes the contribution of the j th feature of the i th sample to the predicted value y_i of the sample. When $f(x_{ij})>0$, it means that the feature makes the prediction value higher and has a positive influence, and vice versa, it means that the feature makes the prediction value lower and has a negative influence. The advantage of SHAP value is that SHAP can reflect the influence of each feature in each sample as well as the positivity and negativity of the influence, and the features themselves also have interactions within the model. This paper utilizes SHAP to explain how the Random Forest algorithm internally predicts outcomes.

III. Reinventing corporate culture and improving the fit of organizational structures

III. A. Analysis of corporate culture reshaping and organizational structure suitability

III. A. 1) Model evaluation criteria

After completing the preprocessing of the original dataset, this paper divides the data into a training set, a test set, and a validation set in the ratio of 4:3:3 to test the warning accuracy of the model. In order to better illustrate the effect of SMOTE-XGBoost model, this paper adopts four evaluation indexes, precision rate, recall rate, F1 score, and ACC, which are common evaluation indexes for binary classification problems, to assess the prediction effect of the model. The original fitness is labeled as positive class, the fitness of this paper is labeled as negative class, and the four results true class, true negative class, false positive class, and false negative class are shown in Table 2.

Table 2: Confusion matrix table

Actual classification	Forecast for the adaptation of this article	The prediction is primitive
Original adaptation	TP:True Positive	FN:False Negative
This article is appropriate	FP:False Positive	TN:True Negative

The precision rate (also known as the check accuracy rate) describes the full sample size predicted to be fitness for this paper, while the recall rate (also known as the check completeness rate) describes how many of the samples that are actually fitness for this paper are correctly predicted. Since the first two evaluation metrics are sometimes contradictory, a score of F1 is added, which is a quantitative assessment criterion for model performance when equal importance is given to the check accuracy and check completeness rates.ACC can be simply understood as the ratio of the number of correctly predicted samples to the number of all the samples, i.e., the accuracy rate. Since this paper selects this paper's fitness and original fitness with a ratio of 1:2, the data distribution is not balanced, and simply using the check accuracy rate as an evaluation index is not scientific and comprehensive enough, so the ACC index is added. According to the categorization of the confusion matrix table in Table 2, the formulas for the four evaluation indicators of precision rate, recall rate, F1 and ACC are as follows:

$$P(\Pr ecision) = \frac{TP}{TP+FP} \tag{14}$$

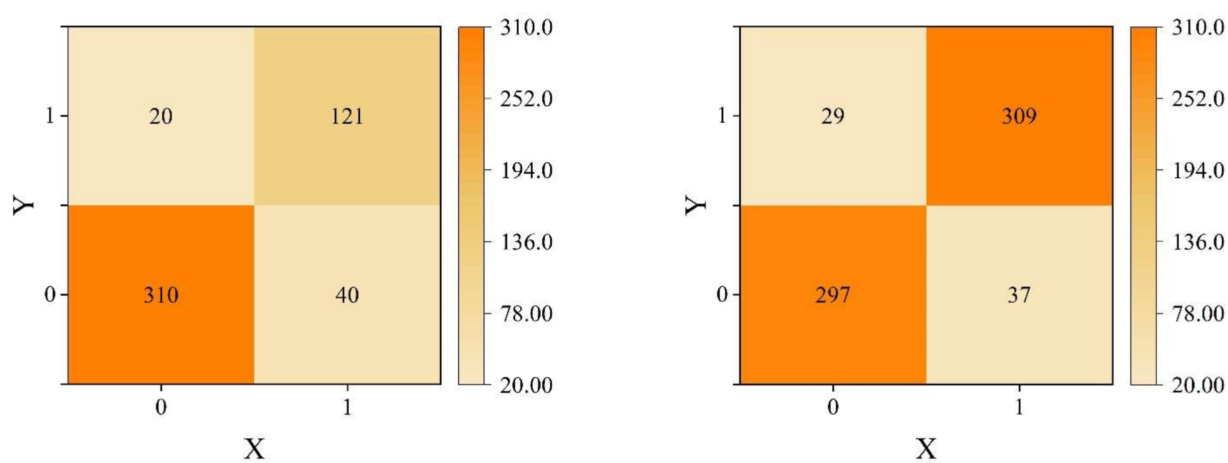
$$R(Recall) = \frac{TP}{TP+FN} \tag{15}$$

$$F_1 = \frac{2\times P\times R}{P+R} \tag{16}$$

$$ACC(Accuracy) = \frac{TP+TN}{TP+FP+FN+TN} \tag{17}$$

III. A. 2) Comparison of prediction accuracy with XGBoost model

In this paper, the SMOTE algorithm is used to oversample the label 1, i.e., generating a few class samples to increase the sample size, and the prediction ability of the XGBoost model before and after the addition of the SMOTE algorithm is first compared. In this paper, we visualize the classification results before and after the SMOTE algorithm is not added, as shown in Fig. 4, each column represents the predicted category, while each row represents the actual category, and there are four classifications: TP, FP, FN, and TN, which correspond to the four squares in the figure, and TP and TN correspond to the part of the prediction that is correct, and the opposite is true for FP and FN. By observing Fig. 4(a), it can be quickly judged that there is an imbalance of data distribution, and Fig. 4(b) the imbalance of data is improved after adding SMOTE, and it can be seen in conjunction with Table 2 that the accuracy rate of this paper's fitness improves by 6% after adding SMOTE, and the overall accuracy rate increases by 3%.



(a) XGBoost confusion matrix (b) SMOTE-XGboost confusion matrix

Figure 4: The classification results of the attack algorithm before and after optimization

The comparison of XGBoost model and SMOTE-XGBoost model is shown in Table 3. In the table, label 0 represents the original fitness, label 1 represents the fitness of this paper, and the accuracy of XGBoost model and SMOTE-XGBoost model are 0.91 and 0.94, respectively.After adding SMOTE, for the fitness of this paper, the precision rate, recall, and F1 scores are 0.94, 0.86, and 0.87, respectively.For the original fitness, the precision rate, recall, F1 scores for the original fitness are 0.88, 0.91, and 0.86, respectively. The precision, recall, and F1 scores for the original fitness are improved by 0.09, 0.12, and 0.07, respectively, compared to the original fitness without the addition of SMOTE. It can be concluded from the experiments that the balanced dataset is effective in improving the prediction accuracy of the XGBoost model.

Table 3: The XGBoost model is controlled by the SMOTE-XGBoost mode

Whether to join the SMOTE	Tags	Accuracy rate	Recall rate	F1 score	Class quantity	ACC
NO	0	0.84	0.93	0.90	311	0.91
	1	0.85	0.74	0.80	149	
YES	0	0.88	0.91	0.86	309	0.94
	1	0.94	0.86	0.87	331	

III. A. 3) Comparison of prediction accuracy with other models

As shown by the results of comparing the prediction accuracy of the SMOTE-XGBoost model with the traditional XGBoost model, the addition of the SMOTE algorithm can improve the classification performance of the model prediction in the processing data stage. In order to compare the effect of different models, this paper selects logistic regression, KNN, decision tree in the basic classification algorithms of machine learning, as well as random forest and XGBoost in the integrated classification algorithms as the representatives, all of which are based on the dataset of the data after adding SMOTE algorithm to carry out experiments, and the results obtained are evaluated by the four classification evaluation indexes of precision rate, recall rate, F1 score, and ACC, and the results are shown in Table 4. Shown.

From the evaluation index ACC, the accuracy rate from high to low is: XGBoost, Random Forest, Logistic Regression, KNN, Decision Tree, and the accuracy rate is 0.94, 0.86, 0.88, 0.61, 0.63. The best classification effect is XGBoost, the main parameters are better than the other classification algorithms, and it can be used for the better prediction of the fitness, and in this paper the precision rate, recall rate, and F1 score of the fitness Recall, F1 score are 0.95, 0.94, 0.96 respectively; followed by Random Forest, for this paper's fitness of the precision rate, recall, F1 score are 0.87, 0.93, 0.92 respectively. the less effective are the logistic regression, for this paper's fitness of the precision rate, recall, F1 score are 0.46, 0.77, 0.59 respectively; the precision rate, recall, F1 score of the original fitness of the rate, recall, and F1 score for the original fitness are 0.86, 0.63, and 0.74, respectively. Compared with the prediction results of the SMOTE-XGBoost model, the accuracy of the prediction of the SMOTE-XGBoost model is significantly higher, which is improved by 29%. For this paper, the fitness precision rate, recall rate, and F1 score increased by 50%, 18%, and 38%, respectively; for the original fitness precision rate increased by 7%, recall rate increased by 35%, and F1 score increased by 23%.

Table 4: The SMOTE-XGBoost model is compared to his model

Model	Tags	Accuracy rate	Recall rate	F1 score	Class quantity	ACC
XGBoost	0	0.92	0.96	0.95	311	0.94
	1	0.95	0.94	0.96	331	
Random forest	0	0.93	0.89	0.92	338	0.86
	1	0.87	0.93	0.92	301	
Logistic regression	0	0.86	0.63	0.74	465	0.61
	1	0.46	0.77	0.59	187	
KNN	0	0.46	0.77	0.59	314	0.63
	1	0.66	0.68	0.68	332	
Decision tree	0	0.68	0.68	0.70	327	0.88
	1	0.90	0.89	0.91	319	

III. B. Analysis of Impact Factors

III. B. 1) Interpretive analysis of single samples

When performing a single sample analysis, although the results may be clear and intuitive, it is important to be wary of errors that may be introduced by randomness. In order to concretely demonstrate the analysis process, the first sample in the dataset is chosen here as an example and the results are presented graphically as shown in Fig. 5. In the graph, the Y-axis represents different eigenvalues, the X-axis represents Shap-value, and $E\left[f(x)\right]$ represents the expectation of all samples $f(x)$, i.e., the average of all predicted values of the model. The $f(x)$ value represents the predicted value of the xth sample, where x=0, the first sample.

Based on the analysis results shown in the figure, it is easy to find that the length of the yellow feature bar representing the negative impact significantly exceeds that of the orange feature bar representing the positive impact, indicating that the negative impact of the feature on the model output is more prominent compared to the positive impact. Therefore, it can be inferred that the predicted probability of this sample will be lower than the baseline value y_{base}. The model predicts a negative indicator of corporate culture remodeling and organizational structure fit, i.e., a non-attrition indicator. By verifying the label of the indicator, it is confirmed that it is indeed a non-attrition indicator, thus verifying the accuracy of the model's prediction. Further analysis reveals that the three characteristics that have the greatest impact on the predicted value of this indicator are: behavioral performance, network development, and enterprise scale development strategy.

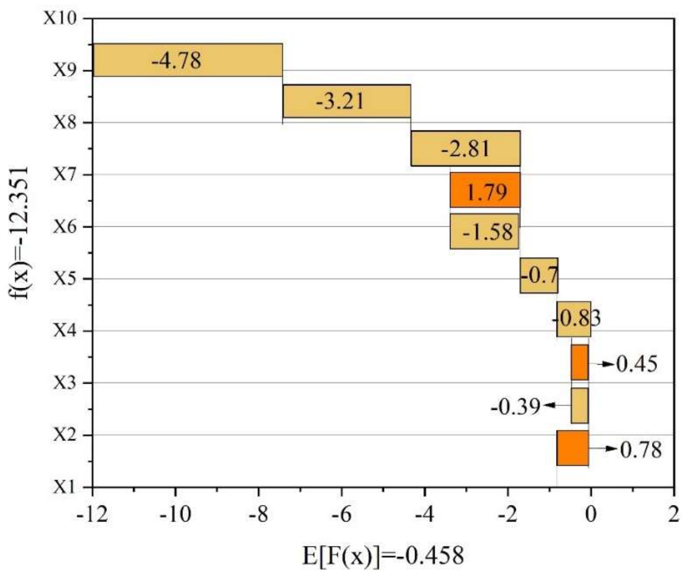


Figure 5: Individual sample analysis results

III. B. 2) Explanatory analysis of the overall model

By calculating and utilizing the SHAP values of the features across all samples, it was possible to plot the positive or negative impact of each indicator on corporate culture reinvention and organizational fitness, as shown in Figure 6. These SHAP values not only help us to gain a deeper understanding of the ways in which each feature contributes to the model's predictions, but also clearly reveal the direction of their influence on the predicted results. In this way, the prediction process of the model can be grasped more comprehensively, and then the role played by different features in the prediction can be interpreted in a targeted manner.

The distribution density in the figure visualizes the aggregation among the indicators. As for the color of the dots, it maps the numerical magnitude of the sample features. Specifically, the color tends to be light yellow indicates that the feature has a low value; on the contrary, the color tends to be orange means that the value is high. And when the value is close to the mean value of the feature, the color of the square will show a yellowish tone. The SHAP value on the horizontal coordinate is an important indicator used to quantify the degree of influence of a particular feature on the model's prediction results.

From the figure, it can be seen that behavioral performance, information technology enhancement, enterprise scale development strategy, market competition and skill mastery are the five characteristics that have the greatest impact on the fitness of corporate culture remodeling and organizational structure.

III. C. Path to Improve the Fitness of Corporate Culture Remodeling and Organizational Structure

By analyzing the fit relationship between corporate culture reshaping and organizational structure, it can provide a basis for enterprises to select, train and cultivate talents as well as corporate culture construction.

(1) Shape and cultivate suitable corporate culture and organizational atmosphere. Differences in leadership effectiveness caused by the behavior of leaders, in which the role of corporate culture is significant. Corporate culture has a guiding effect on the value orientation and behavioral orientation of the enterprise as a whole and its

employees, and it can constrain the thoughts, psychology and behavior of the employees. A strong and positive corporate culture is important for the formation of corporate cohesion and employee centripetal force, and can also have a positive impact on the performance of the leader's effectiveness, therefore, corporate leaders should pay attention to the shaping and refining of the corporate culture while continuously improving their own effectiveness, to form a corporate culture that matches their leadership behavior, and to promote the effective play of their leadership effectiveness.

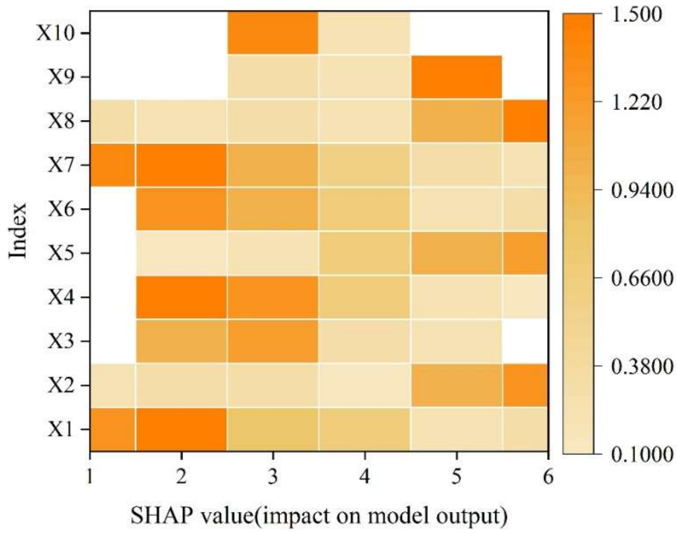


Figure 6: Profile of the output of the shap method

(2) Optimize the selection and evaluation mechanism of the enterprise. Corporate culture is generally deep-rooted in the hearts of employees, when the organizational structure contradicts the strong corporate culture, employee satisfaction will be reduced, and their job satisfaction and organizational commitment will be affected accordingly, which will inevitably affect their performance and thus affect the effectiveness of the team. Therefore, enterprises should focus on strengthening the construction of corporate culture, but also optimize the organizational structure and evaluation mechanism, adjust values, emotions, etc., in order to cultivate leadership behaviors compatible with corporate culture.

IV. Conclusion

This paper summarizes the influencing factors of corporate culture remodeling and organizational structure fitness, selects 12 influencing factors as independent variables, and analyzes the independent variables by using Spearman's correlation coefficient, and concludes that the correlation indexes of enterprise scale development strategy and behavioral performance, and market product sales and network development have reached 0.72 and 0.75, respectively, which are all strongly correlated. After the verification of experimental data, the SMOTE-XGBoost model has a high prediction accuracy for corporate culture remodeling and organizational structure fitness, which is better than other methods. According to the SHAP analysis, it can be concluded that behavioral performance, information technology enhancement, enterprise scale development strategy, market competition and skill mastery are the five characteristics that have the greatest influence on the fitness of corporate culture remodeling and organizational structure.

References

[1] Babar, Z., & Yu, E. (2019, October). Digital transformation–implications for enterprise modeling and analysis. In 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW) (pp. 1-8). IEEE.
[2] Luo, Y., Cui, H., Zhong, H., & Wei, C. (2023). Business environment and enterprise digital transformation. Finance Research Letters, 57, 104250.
[3] Lei, H., Tang, S., & Zan, A. (2023). Enterprise digital transformation under process constraints: investigation of the combinative effects of different constraints. Business Process Management Journal, 29(6), 1867-1891.
[4] Li, Z., Zhang, X., Tao, Z., & Wang, B. (2024). Enterprise digital transformation and supply chain management. Finance Research Letters, 60, 104883.
[5] Peng, Y., & Tao, C. (2022). Can digital transformation promote enterprise performance?—From the perspective of public policy and innovation. Journal of Innovation & Knowledge, 7(3), 100198.

[6] Shang, Y., Raza, S. A., Huo, Z., Shahzad, U., & Zhao, X. (2023). Does enterprise digital transformation contribute to the carbon emission reduction? Micro-level evidence from China. International Review of Economics & Finance, 86, 1-13.

[7] Su, Y., & Wu, J. (2024). Digital transformation and enterprise sustainable development. Finance Research Letters, 60, 104902.

[8] Wang, H., & Wang, S. (2023). A semantic model for enterprise digital transformation analysis. Journal of Computer Information Systems, 63(1), 133-148.

[9] Hamdani, N. A., Maulani, G. A. F., Nugraha, S., Mubarok, T. M. S., & Herlianti, A. O. (2021). Corporate culture and digital transformation strategy in universities in Indonesia. Studies of Applied Economics, 39(10).

[10] Stoianova, O., Lezina, T., & Ivanova, V. (2020, June). Corporate culture: Impact on companies' readiness for digital transformation. In International Conference on Digital Economy (pp. 13-26). Cham: Springer International Publishing.

[11] Herget, J. (2025). The Importance of Corporate Culture in Digital Transformation. In Digitalization and Cultural Change: Strategies for the Modern Work Environment–A Guide (pp. 51-77). Wiesbaden: Springer Fachmedien Wiesbaden.

[12] Steiber, A., & Alvarez, D. (2025). Culture and technology in digital transformations: how large companies could renew and change into ecosystem businesses. European Journal of Innovation Management, 28(3), 806-824.

[13] Asif, M., Yang, L., & Hashim, M. (2024). The role of digital transformation, corporate culture, and leadership in enhancing corporate sustainable performance in the manufacturing sector of China. Sustainability, 16(7), 2651.

[14] Deep, G. (2023). Digital transformation's impact on organizational culture. International Journal of Science and Research Archive, 10(2), 396-401.

[15] Nadkarni, S., & Prügl, R. (2021). Digital transformation: a review, synthesis and opportunities for future research. Management review quarterly, 71, 233-341.

[16] Vial, G. (2021). Understanding digital transformation: A review and a research agenda. Managing digital transformation, 13-66.

[17] Mukesh, H. V. (2022). Digital transformation taking centre stage: how is digital transformation reshaping entrepreneurial innovation?. The Journal of Entrepreneurship, 31(2), 364-401.

[18] Bozkus, K. (2023). Organizational culture change and technology: Navigating the digital transformation. In Organizational Culture-Cultural Change and Technology. IntechOpen.

[19] Chen, Y., Shen, L., Bian, Y., & Zhang, X. (2023). Effects of digital transformation on dynamic capital structure adjustment: Evidence from China. Systems, 11(7), 330.

[20] Zhang, Y., & Wang, J. (2024). Research on influencing factors and path of digital transformation of manufacturing enterprises. Kybernetes, 53(2), 752-762.

[21] Teo Piaw, L. L., Tiai, B. B., & Binti Hassan, E. M. (2024). Dynamic Capital Structure Adjustment: Does Digital Transformation Matter?. Global Business & Management Research, 16(2).

[22] Ktit, M. A., & Abu Khalaf, B. (2025). Does Digital Transformation Reflect the Adjustment of Capital Structure?. Journal of Risk and Financial Management, 18(4), 168.

[23] Zhu, X., & Zhang, H. (2024). Research on Organizational Change of Enterprises under the Background of Digital Transformation. Academic Journal of Business & Management, 6(9), 162-167.

[24] Feng, C., & Ali, D. A. (2024). The Impact of Digital Transformation on Optimising Organisational Efficiency. Accounting and Corporate Management, 6(2), 109-115.

[25] Xuli Li,Yingcang Ma,Qimin Zhou & Xiaohong Zhang. (2024). Sparse large-scale high-order fuzzy cognitive maps guided by spearman correlation coefficient. Applied Soft Computing,167(PA),112253-112253.

[26] Lizheng Wang,Jinling Kong,Qiutong Zhang,Lixin Dong & Yanling Zhong. (2025). Evaluation and Application of the PT-JPL Physical Model Optimized with XGBoost Algorithm in Latent Heat Flux Estimation. Water Resources Management,(prepublish),1-18.

[27] Yang Yonghong,Zheng Tao,Zhang Yu,Jiang Yi & Hu Yixi. (2025). Prediction and Analysis of Expressway Tunnels Crash Based on M-CNN and SHAP Techniques. Proceedings of the Institution of Civil Engineers - Transport,1-32.