Social Security and Administration Management (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/socsam.2026.070108
ISSN 2523-5796 Vol. 7 Num. 1

Data-driven Modeling for Employee Job Satisfaction Prediction and Union Participation Mechanism Analysis

Sinan Jin[1, a, *], Siyuan Zhong[1, b], Songhaomin Lin[1, c], Qikai Wang[1, d]

[1]School of Labor Economics, China University of Labor Relations, Beijing, China
[a]jinsinan094@gmail.com, [b]3189459476@qq.com, [c]2427330842@qq.com, [d]15589001424@163.com
**Corresponding author*

Keywords: Job Satisfaction; Union Participation; Labor Relations; Employee Welfare; Organizational Support

Abstract: As the scale and dimensionality of social survey data continue to increase, research on employee job satisfaction is gradually shifting from traditional statistical analysis to data-driven intelligent modeling. However, labor relations survey data typically includes categorical variables, continuous variables, and various organizational environmental factors, with strong nonlinear correlations between variables, making it difficult for traditional models to fully express these complex relationships. This paper addresses the task of predicting employee job satisfaction levels by constructing a deep learning method that integrates feature tokenization, an FT-Transformer encoder, and an interpretable analysis module. This method transforms survey variables such as union participation, labor protection, enterprise characteristics, individual attributes, and career development into unified feature tokens and utilizes a self-attention mechanism to uncover interaction patterns between different variables. For model evaluation, this paper sets multiple traditional statistical models, machine learning models, and tabular deep learning models as baselines, comparing them in terms of prediction accuracy, class balance recognition ability, and model consistency. The results show that the FT-Transformer-based modeling approach can better adapt to the complex feature structure of labor relations survey data and provides visual support for understanding the roles of union participation, labor protection, and organizational environmental factors in predicting employee satisfaction.

1. Introduction

Employee job satisfaction is a crucial indicator for measuring the quality of labor relations, organizational management effectiveness, and employee work experience [1-2]. It reflects, to a certain extent, employees' comprehensive evaluation of their work environment, compensation and benefits, labor protection, career development, and organizational support [3-4]. With the continuous improvement of the labor relations governance system, trade unions are playing an increasingly prominent role in safeguarding employee rights [5], coordinating labor relations, improving working conditions, and promoting stable enterprise development. Trade union participation can directly influence employees' perceptions of organizational support and rights protection [6], and may also indirectly affect employee job satisfaction through welfare guarantees, labor dispute resolution, labor

contract standardization, and career development support. Therefore, conducting quantitative research on the relationship between trade union participation and employee job satisfaction is not only of practical significance in labor relations governance but also provides a representative application scenario for the intelligent analysis of social survey data. Especially with the increasing availability of publicly available micro-survey data such as the CGSS, how to uncover key factors influencing employee satisfaction from multi-dimensional survey variables has become an important issue in the interdisciplinary field of labor statistics, organizational management, and intelligent data analysis.

Existing research primarily employs descriptive statistics and benchmark regression to analyze the relationship between trade union participation and employee job satisfaction [7-8]. These methods possess strong interpretability and can identify average relationships between variables to some extent, but they also have certain limitations. First, traditional descriptive statistics typically only present superficial differences between union-participating and non-participating employees, making it difficult to further characterize the deep relationships between complex variables. In recent years, deep learning methods have demonstrated strong capabilities in tabular data modeling, feature interaction learning, and complex pattern recognition. Among them, the FT-Transformer can uniformly transform different types of variables into feature labels and automatically learn the interaction relationships between variables through a self-attention mechanism, providing a new technical path for predicting employee job satisfaction and analyzing union participation mechanisms. However, deep learning models also suffer from the black-box problem; if only predictive accuracy is pursued without interpretive analysis, it is difficult to meet the interpretative needs of labor relations research regarding variable mechanisms and policy implications.

Based on the above problems, this paper proposes an interpretable FT-Transformer analytical framework for labor relations survey data, used for predicting employee job satisfaction and mining the impact mechanisms of union participation. This paper uses employee job satisfaction levels as the prediction target and integrates variables such as union participation, individual attributes, enterprise characteristics, labor protection, and career development into the model. It learns the nonlinear relationships between different variables through feature tokenization and Transformer encoding structures, and compares the proposed method with traditional statistical models, machine learning models, and conventional deep learning models to verify its effectiveness in prediction tasks. Furthermore, to enhance model transparency and the interpretability of the results, this paper constructs a multi-level visualization interpretation system. It utilizes SHAP bee swarm graphs to analyze variable contributions, attention heatmaps to display variable interactions, UMAP sample embedding graphs to observe the distribution of different satisfaction groups in the latent space, counterfactual change graphs to simulate the impact of changes in union participation on predicted satisfaction, and heterogeneity heatmaps to reveal the differentiated effects among different industries, enterprise types, education levels, and income groups. The main contributions of this paper are: proposing an interpretable deep learning framework suitable for predicting employee job satisfaction; characterizing the complex interaction between union participation and labor protection, enterprise characteristics, and individual attributes through the attention mechanism of FT-Transformer; and constructing a visualization analysis system consisting of a model structure diagram, SHAP beehive diagram, attention heatmap, UMAP embedding diagram, counterfactual change diagram, and heterogeneity heatmap to enhance the interpretability and demonstrability of the research results from multiple perspectives such as predictive performance, variable contribution, feature interaction, and group differences.

2. Methodology

2.1. Overall Framework

The proposed Union-aware FT-Transformer framework consists of five main components: data input layer, feature tokenization layer, deep interaction modeling layer, satisfaction prediction layer, and interpretability analysis layer. First, the model receives preprocessed employee survey samples. Each sample contains both categorical variables and numerical variables. The categorical variables include union participation, gender, education level, industry, enterprise ownership, and labor contract status, while the numerical variables include age, income, and working hours. Then, the Feature Tokenizer maps different types of variables into feature tokens with the same dimension, so that each variable can be fed into the subsequent Transformer encoder in a unified form. Next, the FT-Transformer Encoder learns high-order interactions among variables through the multi-head self-attention mechanism and outputs the deep representation of each sample. Finally, the classification head outputs the probability that an employee belongs to the low, medium, or high job satisfaction category.

The overall input of the model can be defined as:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \tag{1}$$

where $\mathcal{D}$denotes the employee survey dataset, Ndenotes the number of samples, $\mathbf{x}_i$denotes the input features of the i-th employee sample, and $y_i$denotes its job satisfaction label. For each sample, the input features consist of categorical variables and numerical variables:

$$\mathbf{x}_i = \left[\mathbf{x}_i^{cat}, \mathbf{x}_i^{num}\right] \tag{2}$$

where $\mathbf{x}_i^{cat}$denotes the set of categorical features, and $\mathbf{x}_i^{num}$denotes the set of numerical features. In this paper, employee job satisfaction prediction is formulated as a three-class classification task, and the label space is defined as:

$$y_i \in \{0,1,2\} \tag{3}$$

where 0denotes low satisfaction, 1denotes medium satisfaction, and 2denotes high satisfaction.

2.2. Feature Tokenizer

Since the CGSS survey data contain both categorical and numerical variables, directly feeding them into a deep learning model may lead to inconsistent feature representations across different variable types. Therefore, this paper first uses a Feature Tokenizer to transform all variables into feature tokens with the same dimension [9-10]. For categorical variables, an embedding mapping is adopted to convert discrete values into dense low-dimensional vectors. For numerical variables, a linear projection is adopted to convert scalar values into vector representations with the same dimension as categorical embeddings. In this way, each input variable is represented as an independent token, allowing the subsequent Transformer encoder to learn feature interactions at the variable level [11-12].

For the j-th categorical variable, its embedding representation is defined as:

$$\mathbf{e}_{i,j}^{cat} = \boldsymbol{Emb}_j(x_{i,j}^{cat}) \tag{4}$$

where $\mathrm{Emb}_j(\cdot)$ denotes the embedding layer corresponding to the j-th categorical variable, $x_{i,j}^{cat}$denotes the value of the j-th categorical variable in the i-th sample, and $\mathbf{e}_{i,j}^{cat} \in \mathbb{R}^d$denotes the mapped categorical feature token. Here, $d$denotes the token dimension.

For the k-th numerical variable, its linear projection representation is defined as:

$$\mathbf{e}_{i,k}^{num} = \mathbf{w}_k x_{i,k}^{num} + \mathbf{b}_k \tag{5}$$

where $x_{i,k}^{num}$ denotes the value of the k-th numerical variable in the i-th sample, $\mathbf{w}_k \in \mathbb{R}^d$ and $\mathbf{b}_k \in \mathbb{R}^d$ denote learnable weight and bias parameters, respectively, and $\mathbf{e}_{i,k}^{num} \in \mathbb{R}^d$ denotes the projected token representation of the numerical variable. After concatenating all categorical and numerical feature tokens, the token sequence of the sample is obtained as:

$$\mathbf{T}_i^0 = \left[\mathbf{t}_{cls}, \mathbf{e}_{i,1}, \mathbf{e}_{i,2}, \ldots, \mathbf{e}_{i,M}\right] \in \mathbb{R}^{(M+1)\times d} \tag{6}$$

where M denotes the total number of input variables, $\mathbf{t}_{cls}$ denotes the classification token used to aggregate global information, and $\mathbf{T}_i^0$ denotes the initial token sequence of the i-th sample before being fed into the FT-Transformer. Through this tokenization strategy, variables such as union participation, enterprise ownership, labor protection, and individual attributes are no longer treated as isolated inputs, but are organized into a feature sequence suitable for attention-based interaction learning.

2.3. FT-Transformer Encoder

After feature tokenization, the FT-Transformer Encoder is used to model complex relationships among variables. The main advantage of FT-Transformer lies in its self-attention mechanism, which can adaptively learn interaction weights among variables according to different feature combinations of samples. For the task in this paper, the influence of union participation on employee job satisfaction is usually not isolated. Instead, it may interact with welfare protection, labor contracts, enterprise ownership, income level, career development, and other variables. Therefore, the Transformer structure can more effectively capture such nonlinear and high-order interactions.

For the l-th Transformer Encoder layer, the input token sequence is $\mathbf{T}_i^{l-1}$. It is first linearly transformed into Query, Key, and Value matrices:

$$\mathbf{Q}^l = \mathbf{T}_i^{l-1}\mathbf{W}_Q^l \tag{7}$$

$$\mathbf{K}^l = \mathbf{T}_i^{l-1}\mathbf{W}_K^l \tag{8}$$

$$\mathbf{V}^l = \mathbf{T}_i^{l-1}\mathbf{W}_V^l \tag{9}$$

where $\mathbf{W}_Q^l$, $\mathbf{W}_K^l$, and $\mathbf{W}_V^l$ denote the learnable parameter matrices for Query, Key, and Value in the l-th layer, respectively. The single-head attention is calculated as:

$$\textbf{\textit{Attention}}(\mathbf{Q}^l, \mathbf{K}^l, \mathbf{V}^l) = \textbf{\textit{softmax}}\left(\frac{\mathbf{Q}^l(\mathbf{K}^l)^T}{\sqrt{d}}\right)\mathbf{V}^l \tag{10}$$

To learn variable relationships from different representation subspaces, the multi-head attention mechanism is adopted:

$$\textbf{\textit{MHA}}(\mathbf{T}_i^{l-1}) = \textbf{\textit{Concat}}(\textbf{\textit{head}}_1, \textbf{\textit{head}}_2, \ldots, \textbf{\textit{head}}_H)\mathbf{W}_O^l \tag{11}$$

where H denotes the number of attention heads, and $\mathbf{W}_O^l$ denotes the output projection matrix. Each attention head is calculated as:

$$\textbf{\textit{head}}_h = \textbf{\textit{Attention}}(\mathbf{Q}_h^l, \mathbf{K}_h^l, \mathbf{V}_h^l) \tag{12}$$

After multi-head attention, residual connection, layer normalization, and feed-forward network are

further used to enhance feature representation:

$$\mathbf{T}_i^l = \textit{LayerNorm}\left(\mathbf{T}_i^{l-1} + \textit{MHA}(\mathbf{T}_i^{l-1})\right) \tag{13}$$

$$\mathbf{T}_i^l = \textit{LayerNorm}\left(\mathbf{T}_i^l + \textit{FFN}(\mathbf{T}_i^l)\right) \tag{14}$$

where $\mathrm{FFN}(\cdot)$ denotes the feed-forward neural network. After passing through L Transformer Encoder layers, the model obtains the final deep token representation $\mathbf{T}_i^L$. The output corresponding to the classification token is used as the global representation of the sample:

$$\mathbf{h}_i = \mathbf{T}_{i,cls}^L \tag{15}$$

where $\mathbf{h}_i$ denotes the deep feature representation of the i-th employee sample. This representation comprehensively encodes high-order relationships among union participation, labor protection, enterprise characteristics, and individual attributes, serving as the basis for subsequent job satisfaction prediction and interpretability analysis.

2.4. Classification Head

After obtaining the deep representation of each sample, the classification head is used to output the probability that the employee belongs to different job satisfaction levels. The classification head consists of a fully connected layer and a Softmax function [13-14], which maps the deep features extracted by FT-Transformer into three categories: low satisfaction, medium satisfaction, and high satisfaction.

For the i-th sample, its predicted class probability is defined as:

$$\mathbf{p}_i = \textit{softmax}(\mathbf{W}_c\mathbf{h}_i + \mathbf{b}_c) \tag{16}$$

where $\mathbf{W}_c$ and $\mathbf{b}_c$ denote the weight matrix and bias term of the classification layer, respectively, and $\hat{\mathbf{p}}_i = [\hat{p}_{i,0}, \hat{p}_{i,1}, \hat{p}_{i,2}]$ denotes the predicted probability distribution over the three satisfaction levels.

The final predicted category is obtained as:

$$\hat{y}_i = \arg\max_{c\in\{0,1,2\}} \hat{p}_{i,c} \tag{17}$$

During training, the cross-entropy loss function is adopted:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=0}^{2}\mathbb{I}(y_i = c)\log(\hat{p}_{i,c}) \tag{18}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, which equals 1 when the true label y_i belongs to class c, and 0 otherwise. By minimizing this loss function, the model gradually learns the mapping between employee feature combinations and job satisfaction levels.

To improve model generalization, Dropout, Early Stopping, and weight decay can be introduced during training. The Adam optimizer can be used, and the learning rate can be adjusted according to the validation set performance. When the validation loss does not decrease for several consecutive epochs, training is stopped early to avoid overfitting.

2.5. Explainability Module

Since deep learning models often suffer from the black-box problem, relying only on prediction accuracy is insufficient to explain the decision-making process of the model. Therefore, this paper further constructs an explainability module on top of the FT-Transformer prediction framework. The module explains the model results from five perspectives: feature contribution, feature interaction,

sample representation, prediction change, and subgroup heterogeneity. This module corresponds to five key visualizations in the subsequent analysis, including the SHAP summary plot, attention heatmap, UMAP embedding plot, counterfactual change plot, and heterogeneity heatmap [15].

First, SHAP is used to measure the contribution of different variables to the model prediction results. SHAP is based on the idea of Shapley value, which decomposes the model output into contributions from individual input features. Therefore, it can be used to determine the positive or negative influence of each variable on the prediction result. For the j-th feature, its SHAP value can be defined as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{S\cup\{j\}}(\mathbf{x}) - f_S(\mathbf{x}) \right] \tag{19}$$

where Fdenotes the set of all features, Sdenotes any feature subset that does not contain the j-th feature, $f_S(\mathbf{x})$denotes the model output using only the feature subset S, and $\phi_j$denotes the marginal contribution of the j-th variable to the prediction result. Through the SHAP summary plot, the influence direction and strength of variables such as union participation, labor protection, income level, and enterprise ownership on job satisfaction prediction can be observed intuitively.

Second, this paper constructs a feature interaction heatmap using the attention weights of FT-Transformer. For the h-th attention head in the l-th layer, the attention matrix is defined as:

$$\mathbf{A}_h^l = \boldsymbol{softmax}\left(\frac{\mathbf{Q}_h^l (\mathbf{K}_h^l)^T}{\sqrt{d}} \right) \tag{20}$$

To obtain the overall feature interaction strength, the attention matrices of different layers and different heads are averaged:

$$\overline{\mathbf{A}} = \frac{1}{LH} \sum_{l=1}^{L} \sum_{h=1}^{H} \mathbf{A}_h^l \tag{21}$$

where $\bar{\mathbf{A}}$denotes the average attention matrix. The elements in this matrix can reflect the attention intensity between different variables. For example, if the union participation variable has high attention weights with welfare protection, labor contract, and enterprise ownership, it indicates that the model focuses on these interaction relationships when predicting employee job satisfaction.

Third, UMAP is used to visualize the sample representations output by the last layer of the model. Let the deep representation set of all test samples be:

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]^T \tag{22}$$

UMAP maps the high-dimensional representations into a two-dimensional space:

$$\mathbf{Z} = \boldsymbol{UMAP}(\mathbf{H}), \quad \mathbf{Z} \in \mathbb{R}^{N\times 2} \tag{23}$$

where $\mathbf{Z}$denotes the two-dimensional sample embeddings after dimensionality reduction. Through the UMAP embedding plot, we can observe whether employees with low, medium, and high satisfaction levels form relatively clear distribution structures in the latent space, thereby verifying whether the model has learned discriminative feature representations.

In addition, this paper constructs a counterfactual analysis under the change of union participation status. For a given employee sample, all other variables such as individual attributes, enterprise characteristics, and labor protection remain unchanged, while only the union participation variable is changed. Then, the change in the predicted probability of high job satisfaction is compared. Let the original sample be $\mathbf{x}_i$, its union participation variable be u_i, and the counterfactual sample be $\mathbf{x}_i^{cf}$. Then:

$$\mathbf{x}_i^{cf} = \mathbf{x}_i \quad \textbf{\textit{with}} \quad u_i^{cf} = 1 - u_i \tag{24}$$

The counterfactual prediction change is defined as:

$$\Delta_i = f_{high}(\mathbf{x}_i^{cf}) - f_{high}(\mathbf{x}_i) \tag{25}$$

where $f_{high}(\cdot)$denotes the probability predicted by the model that the sample belongs to the high satisfaction category, and $\Delta_i$denotes the change in the predicted probability of high satisfaction after changing the union participation status. It should be noted that the counterfactual analysis in this paper is a model-based simulation analysis, mainly used to explain model decisions, and does not represent strict causal inference.

Finally, this paper further analyzes the differences in variable contributions across different subgroups through a heterogeneity heatmap. Let G_r denote the r-th employee subgroup, such as different industries, enterprise ownership types, education levels, or income groups. The average contribution of the j-th variable in this subgroup is defined as:

$$\Phi_{r,j} = \frac{1}{|G_r|} \sum_{i \in G_r} \phi_{i,j} \tag{26}$$

where $\phi_{i,j}$denotes the SHAP value of the j-th variable in the i-th sample, and $\Phi_{r,j}$denotes the average contribution of this variable in subgroup G_r. By constructing a two-dimensional matrix composed of $\Phi_{r,j}$, a heterogeneity heatmap can be plotted to show the different effects of union participation, welfare protection, labor contract, and career development across subgroups. This analysis inherits the idea of heterogeneity analysis in traditional labor relation research, while presenting model explanation results in a more intuitive and visually effective way suitable for an EI conference paper.

The overall framework figure is suggested to be named in Figure 1.

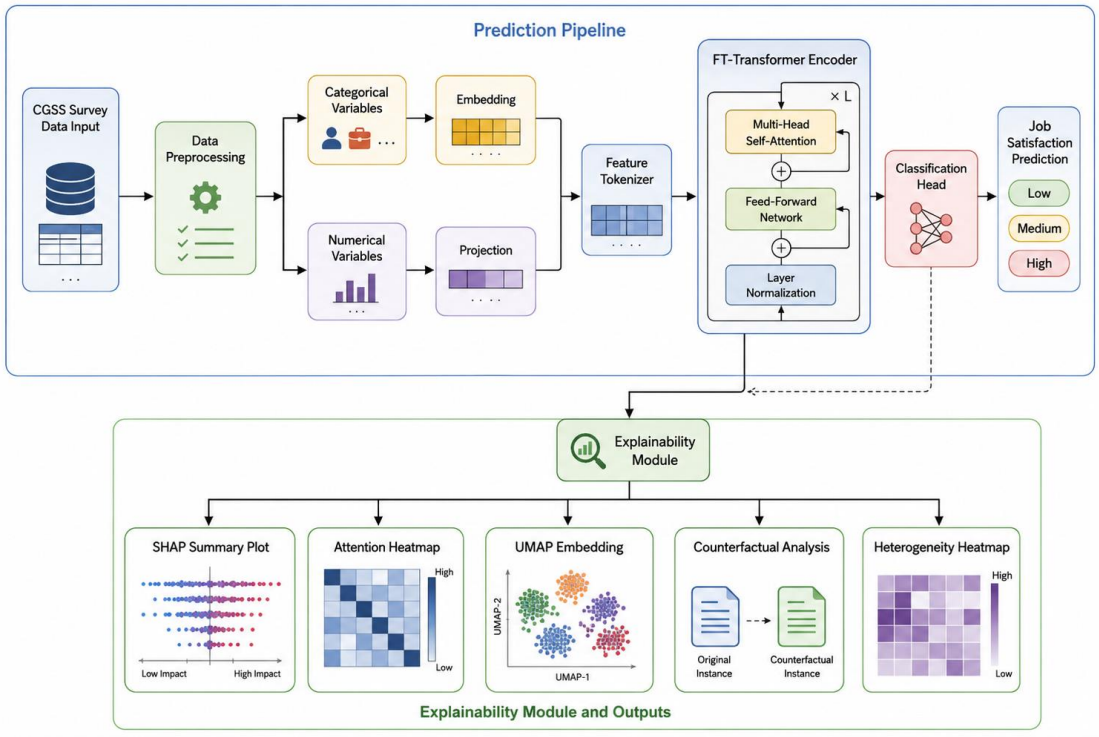

Figure 1: Overall Framework of the Proposed Method.

3. Experiments and Results

3.1. Experimental Design

For the proposed Union-aware FT-Transformer model, the optimizer is Adam, with an initial learning rate of 0.001, adjusted based on the validation set loss. The batch size is set to 128, and the maximum number of training epochs is set to 100. To avoid overfitting on the training set, an Early Stopping strategy is employed: training is terminated early when the validation set loss does not show a significant decrease over 10 or 20 consecutive epochs, and the model parameters with the best validation set performance are saved.

3.2. Baseline Models

To validate the effectiveness of the proposed Union-aware FT-Transformer model, this paper selects several models as comparison models.

First, Logistic Regression and Ordered Logit are chosen as traditional statistical model baselines. Random Forest, XGBoost, and LightGBM are selected as machine learning comparison models. Next, MLP is chosen as the basic deep learning model. MLP learns the non-linear mapping relationship between input variables and job satisfaction through multi-layer fully connected networks, but its ability to model explicit interactions between variables is relatively limited.

Furthermore, TabNet and TabTransformer are selected as baselines for tabular deep learning models. TabNet improves the modeling ability of tabular data through a sequential feature selection mechanism, while TabTransformer performs contextual modeling of categorical variable representations through a Transformer structure.

3.3. Evaluation Metrics

Since employee job satisfaction prediction is formulated as a three-class classification task in this paper, six evaluation metrics are adopted, including Accuracy, Precision, Recall, Macro-F1, AUC, and Cohen's Kappa.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{27}$$

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

$$Recall = \frac{TP}{TP + FN} \tag{29}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{30}$$

$$Macro\mathbf{0}F1 = \frac{1}{C}\sum_{c=1}^{C} F1_c \tag{31}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{32}$$

4. Results and Analysis

4.1. Overall Performance Comparison

As shown in Table 1, the traditional statistical model, Logistic Regression, exhibits relatively limited overall performance, with Accuracy, Macro-F1, and AUC of 0.673, 0.645, and 0.742, respectively. This result indicates that while linear models possess strong interpretability, they struggle to fully capture the complex nonlinear relationships among variables related to employee job satisfaction. Employee job satisfaction is influenced by multiple factors, including union participation, labor protection, company type, income level, education level, and career development; relying solely on linear relationships is insufficient to achieve high predictive performance.

Table 1: Comparison of prediction performance of different models.

Model	Accuracy	Precision	Recall	Macro-F1	AUC	Kappa
Logistic Regression	0.673	0.658	0.641	0.645	0.742	0.487
Random Forest	0.721	0.709	0.694	0.701	0.801	0.561
XGBoost	0.748	0.735	0.724	0.728	0.829	0.604
LightGBM	0.756	0.744	0.731	0.737	0.838	0.618
MLP	0.713	0.701	0.686	0.691	0.794	0.548
TabNet	0.735	0.722	0.708	0.714	0.816	0.584
TabTransformer	0.762	0.751	0.739	0.744	0.845	0.628
Proposed Model	0.791	0.783	0.769	0.775	0.872	0.673

Compared to traditional statistical models, machine learning models such as Random Forest, XGBoost, and LightGBM achieved better predictive results. LightGBM, in particular, achieved Accuracy, Macro-F1, and AUC of 0.756, 0.737, and 0.838, respectively, demonstrating the strong adaptability of gradient boosting tree models to tabular data tasks. This result also indicates the presence of significant nonlinear relationships and variable interactions in the employee job satisfaction prediction task, which traditional linear models cannot fully express. In deep learning models, MLP's accuracy and Macro-F1 score are 0.713 and 0.691, respectively, slightly lower than XGBoost and LightGBM. This indicates that while ordinary fully connected neural networks can learn certain nonlinear mapping relationships, their ability to model explicit interactions between variables is limited. TabNet and TabTransformer outperform MLP, with TabTransformer achieving accuracy, Macro-F1, and AUC of 0.762, 0.744, and 0.845, respectively, demonstrating that tabular deep learning models can more effectively handle mixed survey variables.

4.2. Confusion Matrix Analysis

To further analyze the model's performance in identifying different job satisfaction levels, a confusion matrix of the Proposed Model on the test set was plotted, as shown in Figure 2. The horizontal axis of the confusion matrix represents the model's predicted category, and the vertical axis represents the true category: Low Satisfaction, Medium Satisfaction, and High Satisfaction. The values on the diagonal represent the number of correctly classified samples, and the values off-diagonally represent the number of misclassified samples.

Figure 2 shows that the model has good identification ability for both Low Satisfaction and High Satisfaction samples. Of the 160 samples that were actually Low Satisfaction, 130 were correctly identified; and of the 218 samples that were actually High Satisfaction, 185 were correctly identified. This indicates that the model can effectively distinguish between low-satisfaction and high-satisfaction employee groups and can capture the differences between the two groups in terms of

labor protection, income level, company characteristics, organizational support, and union participation.

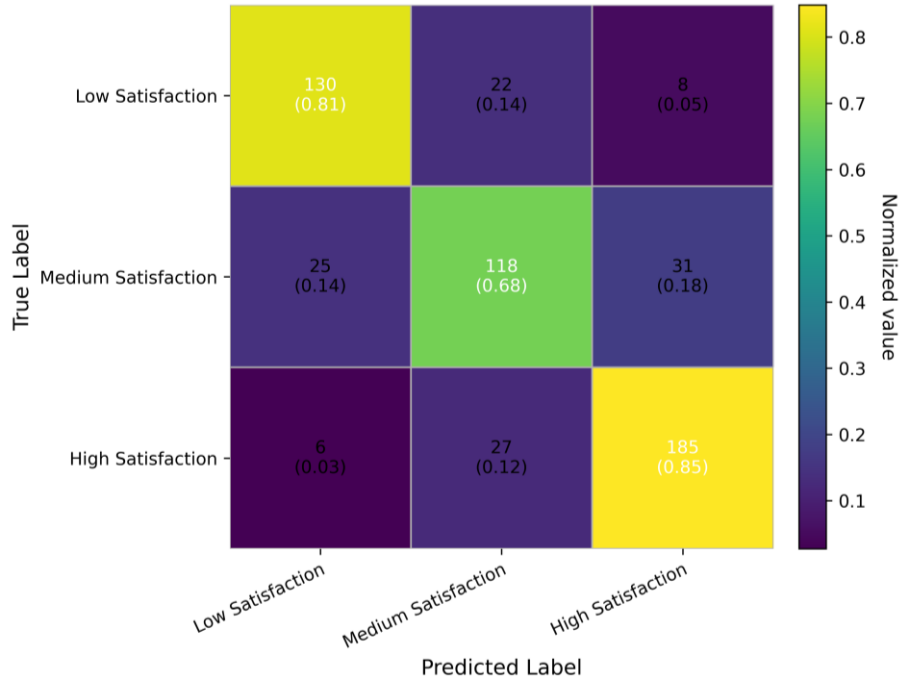

Figure 2: Confusion Matrix of the Proposed Model.

4.3. Ablation Study

To verify the effectiveness of key variables and structural modules in the model, this paper further designed an ablation experiment. The ablation experiment removed the union participation variable, labor security variable, enterprise characteristic variable, and self-attention structure sequentially from the complete model, and observed the changes in model performance. The experimental results are shown in Table 2.

Table 2: Ablation test results.

Model Variant	Accuracy	Macro-F1	AUC	Performance Drop
Full Model	**0.791**	**0.775**	**0.872**	-
w/o Union	0.774	0.756	0.855	-0.019
w/o Labor Protection	0.755	0.732	0.831	-0.043
w/o Enterprise Features	0.768	0.749	0.847	-0.026
w/o Attention	0.734	0.713	0.812	-0.062
MLP Replacement	0.713	0.691	0.794	-0.084

The ablation experiments show that the complete model achieved the best predictive performance, with accuracy, macro-F1, and AUC reaching 0.791, 0.775, and 0.872, respectively. Removing the union participation variable reduced the model's macro-F1 to 0.756, indicating that union participation provides effective information for predicting employee job satisfaction. Removing labor protection variables resulted in a more significant performance decline, with the macro-F1 dropping to 0.732, suggesting a close correlation between labor contracts, social security, welfare benefits, and working hours and job satisfaction. This also indicates that union participation may indirectly influence employee satisfaction through the labor protection pathway. Removing firm characteristics variables reduced the model's macro-F1 to 0.749, demonstrating the significant role of organizational

environment factors such as industry, firm nature, and company size. Further removing the self-attention structure further reduced the model's macro-F1 to 0.713, indicating that the FT-Transformer's attention mechanism can effectively learn the interaction relationships between union participation, labor protection, firm characteristics, and career development. Overall, union participation variables, labor protection variables, enterprise characteristic variables, and attention mechanisms all contribute positively to model performance, with labor protection variables and the self-attention structure having the most significant impact on the model.

5. Visual Explanation and Discussion

5.1. SHAP-based Feature Contribution Analysis

To analyze the contribution of different variables to the prediction results of employee job satisfaction, this paper first draws a SHAP beehive diagram as shown in Figure 3.

As shown in SHAP hive plot 3, variables such as income level, welfare benefits, labor contracts, company type, career development, working hours, and union participation make significant contributions to the prediction of employee job satisfaction. Among these, income level and welfare benefits have the most significant contributions, indicating that employees' evaluation of job satisfaction is closely related to basic economic returns and welfare support. Labor contracts and social security variables also have high explanatory contributions, suggesting that the stability and level of protection in labor relations are important factors influencing satisfaction prediction. Company type and career development variables also show a strong influence in the plot, indicating that different organizational environments and development opportunities affect employees' overall job evaluation.

While union participation is not the sole determinant in the SHAP plot, its contribution ranks relatively high, and it shows a significant positive predictive contribution to the high satisfaction category. Specifically, when the value of the union participation variable is high, its SHAP value is more distributed in the positive region, indicating that employees who participate in unions are more likely to be predicted by the model to have a higher satisfaction category. This result suggests that union participation may influence employee job satisfaction by enhancing employee rights protection, improving organizational support, and promoting harmonious labor relations.

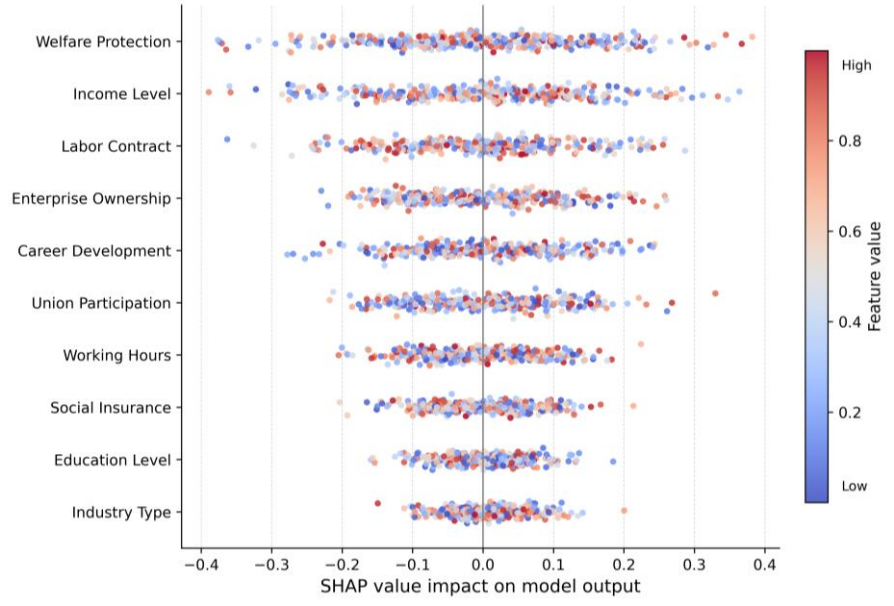

Figure 3: Confusion Matrix of the Proposed Model.

5.2. Attention-based Feature Interaction Analysis

To further reveal the interaction between variables, this paper draws an attention heatmap based on the self-attention weights in FT-Transformer, as shown in Figure 4.

Attention heatmap 4 reveals a high attentional weight between union participation and variables such as welfare benefits, labor contracts, company type, career development, and labor dispute resolution. This indicates that the model does not simply treat union participation as an isolated variable when predicting satisfaction, but rather considers it in conjunction with labor protection and organizational environment variables for comprehensive judgment. For example, the strong attentional connection between union participation and welfare benefits suggests that the model considers a close relationship between the two; the high attentional weight between union participation and labor contracts indicates that whether employees are in a relatively stable labor relationship affects how the union participation variable functions in the model.

Furthermore, there is also a significant attentional connection between company type, income level, and career development. This phenomenon suggests that employee satisfaction is influenced by organizational background, economic rewards, and future development opportunities.

Figure 4: Attention Heatmap of Feature Interactions.

5.3. UMAP-based Representation Visualization

To verify whether the model has learned a discriminative deep sample representation, this paper extracts the latent vectors of the samples output from the last layer of the FT-Transformer and uses the UMAP method to reduce them to a two-dimensional space for visualization, as shown in Figure 5.

As shown in Figure 5, the low, moderate, and high satisfaction samples exhibit certain distributional differences in the two-dimensional embedding space. High satisfaction samples are relatively concentrated in one area, while low satisfaction samples are more distributed on the other side, indicating that the model can effectively distinguish between two groups of employees with significantly different characteristics. Moderate satisfaction samples are mainly distributed between the low and high satisfaction samples, and there is some overlap with the categories on both sides, consistent with the confusion matrix analysis results above.

This distributional characteristic suggests that moderate satisfaction samples inherently possess transitional properties; their variable characteristics may simultaneously resemble those of the low and high satisfaction groups. For example, some moderately satisfied employees may have relatively stable labor contracts and social security, but perform only moderately in terms of income level, career development, or organizational support; conversely, some samples may have higher incomes but lack in working hours, welfare benefits, or labor relationship stability. Therefore, the model faces greater difficulty in identifying moderately satisfied samples.

The UMAP embedding diagram demonstrates that the proposed Union-aware FT-Transformer can not only perform classification prediction but also form sample representations with a certain degree of class discriminative power in the latent space. The results demonstrate that the deep features learned by the model can reflect the structural differences between employee satisfaction levels, providing visual support for the model's prediction results.

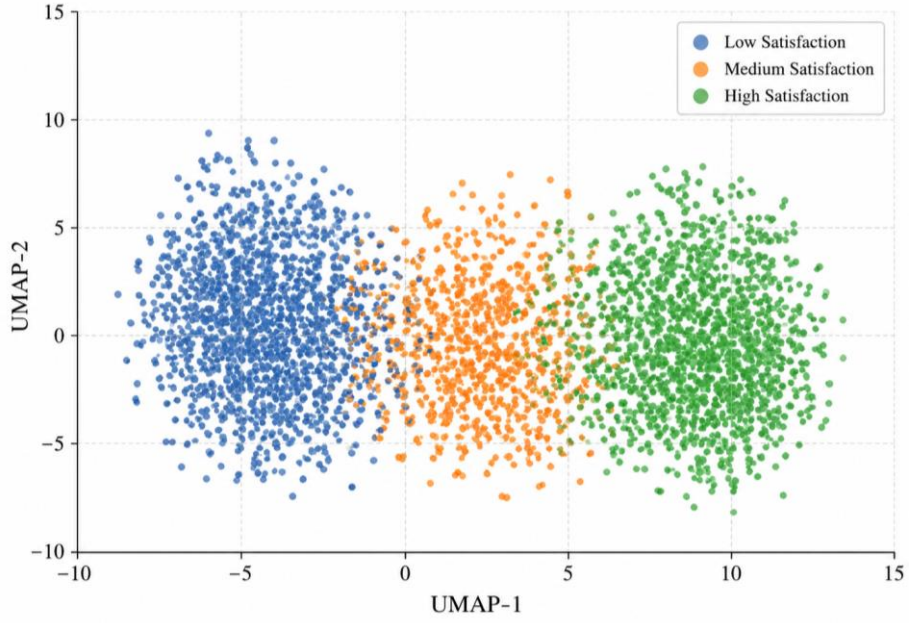

Figure 5: UMAP Visualization of Learned Employee Representations.

5.4. Counterfactual Analysis of Union Participation

To further analyze the impact of union participation variables on the model's predictive results, this paper constructs a counterfactual experiment. Specifically, for each test sample, variables such

as individual attributes, enterprise characteristics, labor protection, income level, and career development are kept constant, while only the value of the union participation variable is changed. The difference in the probability of high satisfaction predicted by the model before and after the change is then compared. In this way, the impact of changes in union participation status on the model's predictive level can be observed.

As shown in Figure 6, when employee status changed from "not participating in the union" to "participating in the union," the predicted probability of high satisfaction increased for some samples, indicating that the union participation variable has a positive impact on the model's prediction results. This increase in predicted probability is particularly pronounced in samples with weak labor protection, insufficient welfare support, or relatively unstable enterprise nature. This suggests that, within the relationships learned by the model, union participation may have higher explanatory value for employees with insufficient labor protection or weak organizational support.

However, it can also be observed that not all samples show significant changes after changing their union participation status. This indicates that union participation is not the sole factor determining employee satisfaction; its strength is influenced by variables such as income level, labor contract, welfare protection, career development, and enterprise nature. For example, for employees with high income, comprehensive welfare protection, and stable labor contracts, changing union participation alone may not significantly increase the model's predicted probability of high satisfaction; while for employees with lower protection levels or weaker labor relations stability, union participation may have a more significant supplementary effect.

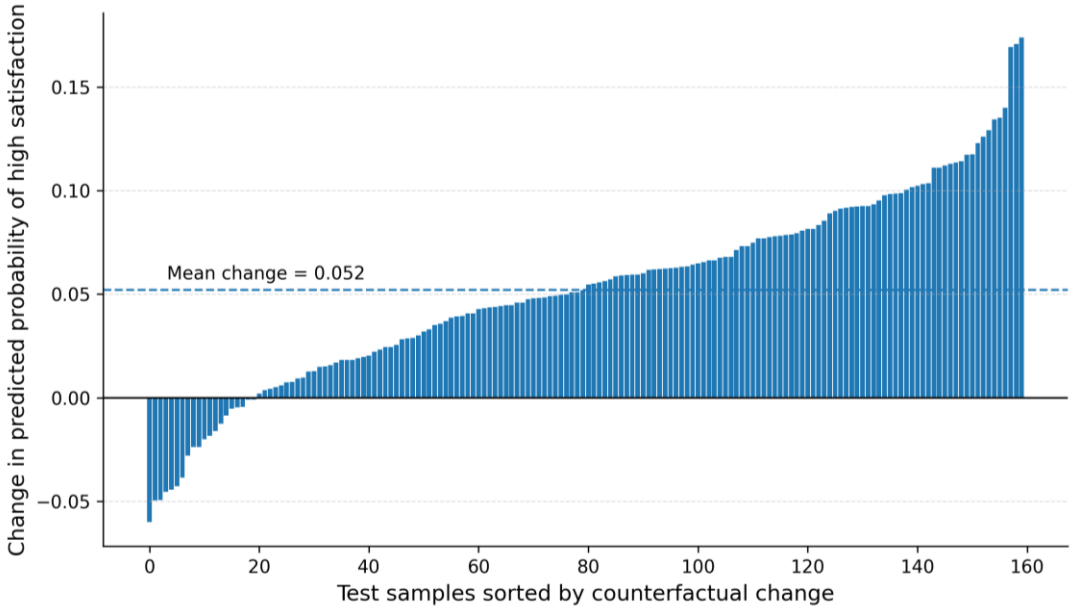

Figure 6: Counterfactual Changes in Predicted Satisfaction under Union Participation Shift.

5.5. Heterogeneity Analysis across Subgroups

As shown in Figure 7, the contribution of the union participation variable differs significantly across different groups. In non-public enterprises and groups with relatively weak labor protection levels, the average contribution of the union participation variable is relatively high, indicating that union participation has a stronger explanatory power for job satisfaction in these groups. This may be because, in situations with relatively inadequate labor protection, union participation more readily demonstrates its value in rights protection, communication and coordination, and organizational support.

From an income group perspective, union participation and welfare protection variables contribute relatively more to low- and middle-income groups. This suggests that for employees with lower incomes or weaker economic security, labor protection, welfare support, and organizational protection may have a stronger impact on job satisfaction. In contrast, in high-income groups, variables such as career development, enterprise nature, and work autonomy may have higher explanatory value.

From the dimensions of education and career development, career development variables contribute more significantly to highly educated groups, while labor contracts and welfare protection variables contribute relatively more to lower-educated groups. This result indicates that different employee groups do not have entirely the same focus on job satisfaction. Employees with higher education levels may pay more attention to career growth opportunities and development prospects, while employees with lower education levels may pay more attention to the stability of labor relations, social security benefits, and protection of basic rights.

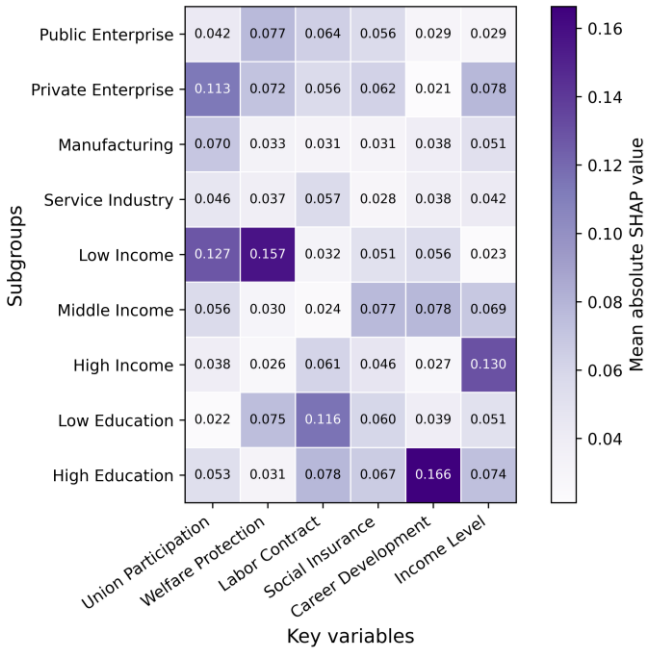

Figure 7: Heterogeneity Heatmap of Union Participation Effects across Subgroups.

6. Conclusion

This paper addresses the issue of predicting employee job satisfaction and analyzing the impact mechanism of union participation, proposing a deep learning analysis framework based on interpretable FT-Transformer. Using CGSS survey data as a foundation, this framework unifies mixed variables such as union participation, individual attributes, enterprise characteristics, labor protection, and career development into feature tokens, and learns the nonlinear relationships and higher-order interactions between variables through a self-attention mechanism. Experimental results show that, compared to other comparative models, the proposed method achieves superior or more stable predictive performance on different indicators.

To alleviate the black-box problem of deep learning models, this paper further constructs a visual interpretation system consisting of SHAP beehive graph, attention heatmap, UMAP sample embedding graph, counterfactual change graph, and heterogeneity heatmap. SHAP analysis results indicate that variables such as income level, welfare protection, labor contract, enterprise type, career development, and union participation contribute significantly to satisfaction prediction, with union

participation showing a positive predictive effect on high satisfaction categories. Attention heatmaps reveal strong interactions between union participation and variables such as welfare benefits, labor contracts, company type, and career development, indicating that the role of unions often needs to be understood within the context of specific labor protection environments and organizational backgrounds. UMAP results show that the latent space representation of employees learned by the model can effectively distinguish between low and high satisfaction samples, while samples with moderate satisfaction exhibit some transitional characteristics. Counterfactual analysis and heterogeneity heatmaps further illustrate that the strength of the role of union participation variables varies across different company types, income levels, education levels, and labor protection statuses. Overall, this paper not only provides an effective deep learning method for predicting employee job satisfaction but also reveals the potential mechanisms of union participation and related labor protection variables through multi-level interpretable visualization analysis, providing a data-driven reference for optimizing union services and governing labor relations.

Acknowledgements

This research was funded by China University of Labor Relations, grant number “26xjx030" and "The APC was funded by China University of Labor Relations".

References

[1] Meybodi A R. Identifying the dimensions of employee experience according to the effect of satisfaction, work place and organizational culture[J]. International Journal of Human Capital in Urban Management, 2024, 9(1): 85-100.
[2] Hemsworth D. Exploring the theory of employee planned behavior: job satisfaction as a key to organizational performance[J]. Psychological reports, 2026, 129(2): 1584-1615.
[3] Loo S H. Key factors affecting employee job satisfaction in Malaysian manufacturing firms post COVID-19 pandemic: a Delphi study[J]. Cogent Business & Management, 2024, 11(1): 2380809.
[4] Hoxha G. Sustainable healthcare quality and job satisfaction through organizational culture: Approaches and outcomes[J]. Sustainability, 2024, 16(9): 3603.
[5] Mengale M S, Rajuskar C, Badarke M V. Bridging the gap: Why industrial workers need unions now more than ever[J]. International Journal for Multidisciplinary Research, 2024, 6(3): 1-5.
[6] Però D, Downey J. Advancing workers' rights in the gig economy through discursive power: The communicative strategies of indie unions[J]. Work, Employment and Society, 2024, 38(1): 140-160.
[7] Pienaar J. How do job demands and job resources relate to well-being, turnover intention and performance in retail? Insights from Swedish trade union members[J]. The International Review of Retail, Distribution and Consumer Research, 2025: 1-25.
[8] Knox A. Understanding the relationship between job quality and general health in hospitality: Evidence from Australia[J]. International Journal of Hospitality Management, 2026, 137: 104690.
[9] Afifi F. Transformer-based tokenization for IoT traffic classification across diverse network environments[J]. PeerJ Computer Science, 2025, 11: e3126.
[10] Zhou B, Chen W. Enhancing FT-Transformer with a Matérn-driven Kolmogorov-Arnold Feature Tokenizer for Tabular Data-Based In-Bed Posture Classification[J]. IEEE Access, 2025.
[11] Zhu E. Adaptive tokenization transformer: enhancing irregularly sampled multivariate time series analysis[J]. IEEE internet of things journal, 2025.
[12] Li X Y. Industrial data imputation based on multiscale spatiotemporal information embedding with asymmetrical transformer[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(8): 14937-14948.
[13] LI Y. Integrating evolutionary algorithms and enhanced-YOLOv8+ for comprehensive apple ripeness prediction[J]. Scientific Reports, 2025, 15: 7307. DOI: 10.1038/s41598-025-91939-4.
[14] BAI X, LI Y. DS-ARO: a multi-strategy improved artificial rabbits optimization algorithm for global optimization and corporate bankruptcy prediction[J]. Scientific Reports, 2026. DOI: 10.1038/s41598-026-57561-8.
[15] CHEUNG Y, GUO Z, LI Y. Foliated generative adversarial networks for rolling bearings fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2026, 75: 1-24. DOI: 10.1109/TIM.2026.3684644.