Population, Resources & Environmental Economics (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/pree.2026.070101
ISSN 2616-2253 Vol. 7 Num. 1

Research on Income Inequality Measurement and Influencing Factor Identification among Flexible Workers Based on Interpretable Machine Learning

Sinan Jin

School of Labor Economics, China University of Labor Relations, Beijing, China
jinsinan094@gmail.com

Keywords: Flexible employment; Income inequality; Identification of influencing factors; Income forecasting; Interpretable analysis

Abstract: With the rapid development of the digital economy and platform economy, flexible employment has become an important form of employment in the labor market. However, problems such as widening income disparities, insufficient income stability, and inadequate social security coverage are gradually becoming prominent. Addressing the shortcomings of traditional statistical analysis methods in characterizing the nonlinear relationships and interactions of income-influencing factors, this paper proposes a method for measuring income inequality and identifying influencing factors among flexible workers based on the Extreme Gradient Boosting (XGBoost) algorithm and Shapley Additive Explanations (SHAP). First, the Gini coefficient, Theil index, and Lorenz curve are used to measure and visualize the income distribution of flexible workers to characterize the degree of income inequality within the sample. Second, XGBoost is constructed to predict the income level of flexible workers and compared with other models. Finally, the Shapley Additive Explanation method is used to interpret the prediction results of the Extreme Gradient Boosting model, identifying the contribution and direction of different variables to income disparities. This study provides a computational analysis method that combines predictive and explanatory power for measuring income inequality among flexible workers, identifying low-income risks, and analyzing factors influencing income.

1. Introduction

With the rapid development of the digital economy and platform economy, flexible employment has become an important form of employment in the labor market [1-2]. The number of jobs in food delivery [3], ride-hailing, live streaming, e-commerce operations, and various platform-based services is constantly expanding, providing workers with more diverse employment options. However, while flexible employment increases employment flexibility, it also brings about significant income distribution issues [4-5]. Due to platform algorithm scheduling, differences in worker skills, working hours, insufficient social security coverage, and varying industry entry barriers [6-7], the income levels of flexible workers exhibit significant differentiation. Some workers face problems such as large income fluctuations, low levels of social security, and limited career

development opportunities [8-9]. Therefore, accurately measuring the degree of income inequality among flexible workers and further identifying the key factors affecting income disparities is of significant practical importance and research value.

Existing research on income issues in flexible employment mainly focuses on income characteristic description, inequality index measurement, and identification of influencing factors. Commonly used methods include descriptive statistical analysis, which can reflect the overall characteristics of income distribution among flexible workers and analyze the sources of income disparities from the perspectives of education level, skill level, working hours, platform type, and social security participation. However, traditional statistical models typically rely on strong linear assumptions, failing to adequately characterize the potential nonlinear relationships and interactions between variables, thus making it difficult to fully reveal the formation mechanisms of income inequality in complex employment scenarios. Meanwhile, while some machine learning models possess strong predictive capabilities, without reasonable explanations, they can easily remain at the level of black-box predictions, failing to provide a clear basis for income distribution optimization and policy formulation.

To address these issues, this paper proposes a method for measuring income inequality and identifying influencing factors among flexible workers, based on the Extreme Gradient Boosting (XGBoost) algorithm and the Shapley Additive Explanations (SHAP) method. First, the Gini coefficient, Theil index, and Lorenz curve are used to measure the degree of income inequality among flexible workers, characterizing the overall differences in income distribution. Second, an extreme gradient boosting regression model is constructed and compared with different models to verify the fitting ability of the machine learning model to the mechanism of income impact of flexible employment. Finally, the Shapley additive interpretation method is introduced to interpret the model's prediction results, identifying the contributions of variables such as working hours, education level, skill level, platform type, social security participation, and regional differences to income inequality.

2. Methodology

2.1. Overall Framework

The overall framework proposed in this paper mainly consists of five parts: data preprocessing, income inequality measurement, income prediction modeling, model performance evaluation, and key influencing factor explanation. First, the original survey data are cleaned, missing values are processed, and categorical variables are encoded to form a structured dataset for modeling and analysis. The research sample is defined as:

$$D = \{(x_i, y_i)\}_{i=1}^{n} \tag{1}$$

where $x_i$denotes the feature vector of the i-th flexible employment worker, including individual characteristics, employment characteristics, and social security characteristics. $y_i$denotes the income level of the i-th sample, and $n$denotes the number of samples.

The overall technical route can be shown in Figure 1.

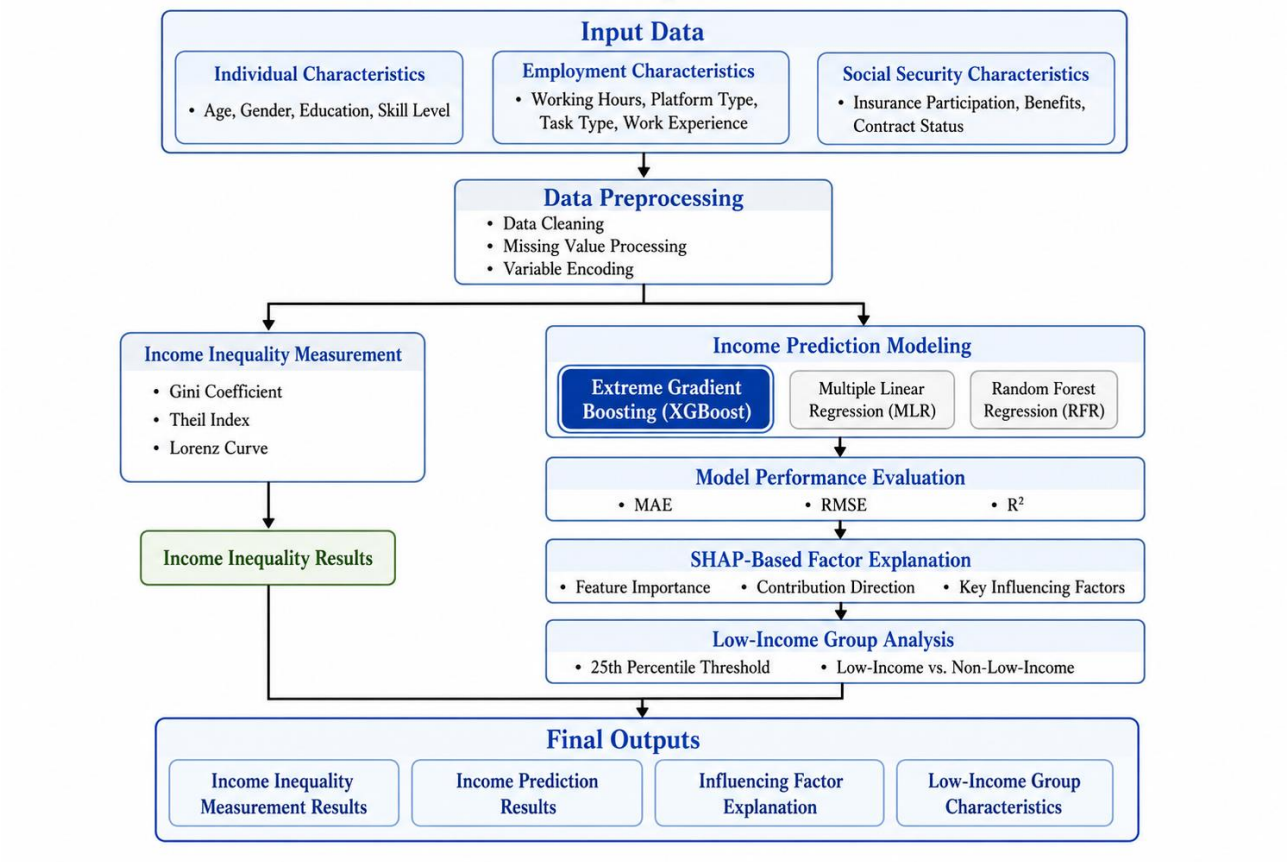

Figure 1: Overall Framework of the Proposed XGBoost-SHAP Method.

2.2. Income Inequality Measurement

To characterize the income inequality degree of flexible employment workers, this paper adopts the Gini Coefficient, Theil Index, and Lorenz Curve as the main measurement methods. The Gini Coefficient is used to measure the overall inequality degree of income distribution [10-11]. The Theil Index is used to reflect the dispersion degree of income distribution. The Lorenz Curve is used to visually describe the deviation between the cumulative population proportion and the cumulative income proportion.

First, suppose the sample incomes are $y_1, y_2, \cdots, y_n$, and the average income is $\bar{y}$. The Gini Coefficient is calculated as follows:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j|}{2n^2\bar{y}} \tag{2}$$

where G denotes the Gini Coefficient. A larger value of G indicates a more unequal income distribution, while a value closer to 0 indicates a more equal income distribution.

Second, the Theil Index is calculated as follows:

$$T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\bar{y}}\ln\left(\frac{y_i}{\bar{y}}\right) \tag{3}$$

where T denotes the Theil Index, y_i denotes the income level of the i-th sample, and $\bar{y}$ denotes the average income of all samples. A larger Theil Index indicates a more significant income distribution

difference.

To further visualize the income distribution difference, this paper draws the Lorenz Curve. All sample incomes are sorted in ascending order, and the sorted incomes are denoted as $y_{(1)}, y_{(2)}, \cdots, y_{(n)}$. When the cumulative population proportion is p_k, the corresponding cumulative income proportion is:

$$L(p_k) = \frac{\sum_{i=1}^{k} y_{(i)}}{\sum_{i=1}^{n} y_{(i)}}, \quad p_k = \frac{k}{n} \tag{4}$$

where $L(p_k)$denotes the cumulative income proportion of the lowest-income $k$samples. The more the Lorenz Curve deviates from the absolute equality line, the more unequal the income distribution is.

2.3. XGBoost Income Prediction Model

In the income prediction modeling part, this paper adopts Extreme Gradient Boosting as the main model. XGBoost is an ensemble learning method based on gradient boosting decision trees. It reduces prediction errors by continuously adding regression trees [12-13]. Compared with traditional Multiple Linear Regression, XGBoost does not require a strict linear relationship assumption and can effectively handle nonlinear relationships and interaction effects among variables. Therefore, it is suitable for the income prediction task of flexible employment workers.

Suppose the input feature is x_i, and the actual income is y_i. The prediction result of the XGBoost model can be expressed as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathrm{F} \tag{5}$$

where $\hat{y}_i$denotes the predicted income of the i-th sample, Kdenotes the number of regression trees, $f_k(x_i)$denotes the prediction result of the k-th regression tree, and $\mathcal{F}$denotes the function space composed of all regression trees.

The objective function of XGBoost consists of a loss function and a regularization term, which can be expressed as:

$$\mathrm{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where $l(y_i, \hat{y}_i)$ denotes the loss function between actual income and predicted income, and $\Omega(f_k)$denotes the regularization term of the k-th tree, which is used to control model complexity and prevent overfitting.

The regularization term of a regression tree is defined as:

$$\Omega(f) = \gamma Q + \frac{1}{2}\lambda \sum_{j=1}^{Q} w_j^2 \tag{7}$$

where Qdenotes the number of leaf nodes, $w_j$denotes the weight of the j-th leaf node, and $\gamma$and $\lambda$are regularization parameters. This term penalizes overly complex tree structures and improves the generalization ability of the model.

During model training, XGBoost adopts an additive modeling strategy to update the prediction result step by step. The prediction value in the t-th iteration can be expressed as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{8}$$

where $\hat{y}_i^{(t)}$ denotes the prediction result of the model in the t-th iteration, $\hat{y}_i^{(t-1)}$ denotes the prediction result of the previous iteration, and $f_t(x_i)$ denotes the newly added regression tree in the t-th iteration.

To verify the prediction performance of XGBoost, this paper selects Multiple Linear Regression and Random Forest Regression as comparison models. Multiple Linear Regression represents traditional statistical modeling methods, while Random Forest Regression represents general ensemble learning methods. By comparing these three models, the applicability of XGBoost in the flexible employment income prediction task can be more clearly demonstrated.

The model performance is evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). The Mean Absolute Error is calculated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

2.4. SHAP-Based Factor Explanation

To improve the interpretability of the machine learning model, this paper uses Shapley Additive Explanations to explain the XGBoost model. SHAP is derived from the Shapley Value concept in cooperative game theory. Its core idea is to decompose the model prediction result into the sum of the contributions of each feature variable [14-15], thereby identifying the positive or negative influence of each variable on predicted income.

For the i-th sample, SHAP represents the model prediction value as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{M}\phi_{ij} \tag{12}$$

where $\hat{y}_i$ denotes the predicted income of the i-th sample, ϕ_0 denotes the baseline prediction value of the model, M denotes the number of feature variables, and ϕ_{ij} denotes the contribution value of the j-th feature to the prediction result of the i-th sample. When $\phi_{ij} > 0$, the feature increases the predicted income. When $\phi_{ij} < 0$, the feature decreases the predicted income.

The SHAP value of a certain feature j is calculated as:

$$\phi_j = \sum_{S \subseteq F,\ \{j\}} \frac{|S|!(M-|S|-1)!}{M!}\left[f_{S\cup\{j\}}(x_{S\cup\{j\}}) - f_S(x_S)\right] \tag{13}$$

where F denotes the full feature set, S denotes a feature subset that does not contain feature j, M denotes the total number of features, $f_{S\cup\{j\}}(x_{S\cup\{j\}})$ denotes the model prediction result after adding feature j, and $f_S(x_S)$ denotes the model prediction result without feature j. The difference between the two terms represents the marginal contribution of feature j.

In this paper, SHAP is mainly used to explain the influence of working hours, education level, skill level, platform type, social security participation, and regional differences on the income level

of flexible employment workers. The SHAP feature importance ranking can identify which variables have the greatest influence on income prediction results. The SHAP summary plot can further show the positive and negative influence directions of different variable values on predicted income.

2.5. Low-Income Group Analysis

To further analyze the low-income risk characteristics of flexible employment workers, this paper divides the samples according to income quantiles. Specifically, the 25th percentile of income is used as the threshold for identifying low-income workers. Samples with income lower than this threshold are defined as the low-income group, while the remaining samples are defined as the non-low-income group.

The low-income group indicator variable is defined as:

$$g_i = \begin{cases} 1, & y_i < Q_{0.25} \\ 0, & y_i \geq Q_{0.25} \end{cases} \tag{14}$$

where $g_i = 1$indicates that the i-th sample belongs to the low-income group, $g_i = 0$indicates that the i-th sample belongs to the non-low-income group, and $Q_{0.25}$denotes the 25th percentile of income.

After income grouping, this paper further compares the differences between the low-income group and the non-low-income group in terms of individual characteristics, employment characteristics, and social security characteristics. For example, the low-income group may show characteristics such as lower education level, insufficient skill level, lower social security participation rate, longer working hours, and lower hourly income return. Through this analysis, the model interpretation results can be further transformed into the identification of low-income risk groups, providing a basis for income distribution optimization, social security coverage improvement, and vocational skill enhancement.

3. Experiments and Results

3.1. Dataset and Experimental Settings

The experimental data used in this study mainly come from public survey data and self-collected questionnaire data. The public data can be obtained from the China Family Panel Studies (CFPS) and the Chinese General Social Survey (CGSS), including samples related to employment status, income level, working hours, social security participation, and individual characteristics. The self-collected questionnaire data are used to supplement more detailed variables, such as platform type, task type, flexible employment form, and social security participation.

During sample screening, this study first identifies flexible employment workers according to employment type, and removes samples with missing income values, missing key variables, or obvious abnormal values. For continuous variables, such as age, working hours, work experience, and income level, numerical processing is adopted. For categorical variables, such as gender, education level, platform type, task type, urban-rural attribute, and social security participation, one-hot encoding or label encoding is used to transform them into features that can be recognized by the model. To reduce the influence of extreme values on model training and income inequality measurement, the income variable is checked for outliers, and necessary winsorization or smoothing is conducted according to the actual data distribution.

The input variables in this study mainly include three categories. The first category is individual characteristics, including age, gender, education level, and skill level. The second category is employment characteristics, including working hours, platform type, task type, and work experience. The third category is social security characteristics, including social insurance participation, welfare protection, and labor contract status. The output variable is the income level of flexible employment

workers. Depending on data availability, monthly income or annual income can be selected as the target variable. For the machine learning prediction experiment, the samples are divided into a training set and a test set at a ratio of 7:3. The relevant data settings are shown in Table 1.

Table 1: Variable Description.

Variable Type	**Variable Name**	**Description**	**Processing Method**
Dependent variable	Income level	Monthly or annual income of flexible employment workers	Continuous variable
Individual characteristics	Age	Age of the respondent	Continuous variable
Individual characteristics	Gender	Male or female	Categorical encoding
Individual characteristics	Education level	Educational attainment	Categorical encoding
Individual characteristics	Skill level	Occupational skill or skill grade	Categorical encoding
Employment characteristics	Working hours	Weekly or daily working hours	Continuous variable
Employment characteristics	Platform type	Food delivery, ride-hailing, e-commerce, live streaming, etc.	Categorical encoding
Employment characteristics	Task type	Delivery, order-taking, content creation, sales, etc.	Categorical encoding
Employment characteristics	Work experience	Duration of flexible employment	Continuous variable
Social security characteristics	Social security participation	Whether the worker participates in social insurance	Binary variable
Social security characteristics	Welfare protection	Whether the worker has platform subsidies, commercial insurance, or other welfare support	Binary variable
Regional characteristics	Urban-rural attribute	Urban or rural residence	Categorical encoding

3.2. Income Inequality Measurement Results

To analyze the degree of income inequality among flexible employment workers, this study first draws the Lorenz Curve and calculates the Gini Coefficient and Theil Index. The Lorenz Curve is used to show the relationship between the cumulative population share and the cumulative income share. When the Lorenz Curve is closer to the 45-degree line of perfect equality, the income distribution is more equal. When the curve bends downward and deviates further from the line of perfect equality, income concentration becomes higher, indicating a more obvious degree of income inequality.

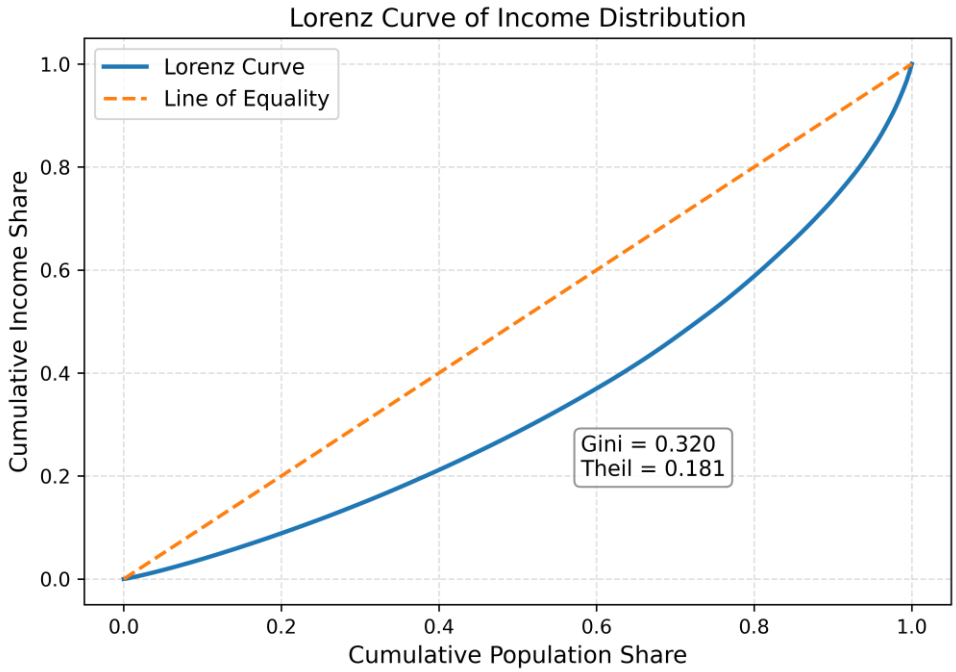

Figure 2: Lorenz Curve of Income Distribution.

As shown in Fig. 2, the Lorenz Curve of the income distribution of flexible employment workers lies below the line of perfect equality and shows a certain degree of deviation from it. This indicates that income is not evenly distributed within the sample, and there is a certain degree of income concentration. Specifically, low-income samples account for a relatively large proportion of the population, but their cumulative income share is relatively low. In contrast, high-income samples account for a smaller proportion of the population but hold a higher proportion of total income. This result suggests that there are clear income differences among flexible employment workers. Therefore, it is necessary to further use machine learning models to identify the key factors affecting income differences.

Table 2: Income Inequality Measurement Results.

Indicator	**Value**	**Meaning**	**Result Interpretation**
Gini Coefficient	0.356	Measures the overall degree of income distribution inequality	The value is greater than 0, indicating that income differences exist in the sample. The Lorenz Curve is clearly below the line of perfect equality, which is consistent with Fig. 2
Theil Index	0.218	Measures the dispersion degree of income distribution	The value is greater than 0, indicating income dispersion. Some high-income samples increase the overall income gap
Cumulative income share of the lowest-income 50% group	0.279	Measures the share of total income held by the lower-income half of the sample	Low-income samples account for a large population share but hold a relatively low income share
Cumulative income share of the highest-income 20% group	0.411	Measures the share of total income held by the top 20% income group	The high-income group accounts for a smaller population share but holds a higher income share, indicating a certain degree of income concentration

The results in Table 2 correspond to the Lorenz Curve in Fig. 2. Both the Gini Coefficient and the Theil Index are greater than 0, indicating that the income distribution among flexible employment workers is not completely equal. Meanwhile, the cumulative income share of the lowest-income 50% group is significantly lower than its population share, while the highest-income 20% group holds a relatively high proportion of total income. This further indicates that there is a certain degree of income inequality within the sample. Overall, income differences among flexible employment workers are reflected not only in the dispersion of income levels, but also in the income concentration across different income groups. Therefore, it is necessary to further use Extreme Gradient Boosting and Shapley Additive Explanations to identify the key factors causing income differences.

3.3. Income Prediction Results

After measuring income inequality, this paper further constructs an income prediction model to analyze the combined impact of personal attributes, employment characteristics, and social security factors on the income level of flexible workers.

As shown in Table 3, compared to the multiple linear regression model and the random forest regression model, the XGBoost model achieved lower errors in the MAE and RMSE indices, while also showing higher performance in the R^2 index. This indicates that the XGBoost model can better fit the nonlinear relationship between the income level of flexible workers and multiple influencing factors. Although the multiple linear regression model has good interpretability, its strong linear assumptions make it difficult to fully characterize the interaction effects between variables such as platform type, working hours, skill level, and social security participation. While the random forest model can handle nonlinear relationships, it is slightly weaker than XGBoost in terms of overall predictive stability and error control. Overall, XGBoost is more suitable as the basic model for the subsequent explanatory analysis of influencing factors in this paper.

Table 3: Model Performance Comparison.

Model	**MAE**	**RMSE**	**R^2**
Multiple Linear Regression	842.36	1168.42	0.512
Random Forest Regression	706.58	984.75	0.641
Extreme Gradient Boosting	638.27	895.31	0.718

3.4. SHAP Explanation Results

To further interpret the prediction results of the XGBoost model, this paper introduces SHAP to analyze the model's interpretability. Compared with traditional feature importance methods, SHAP can not only measure the importance of different variables to the model's prediction results, but also further determine the positive or negative impact of variables on income prediction results. Therefore, the SHAP method can compensate for the lack of interpretability in machine learning models, making the model results more intuitive and understandable.

Figure 3 shows the ranking of SHAP feature importance of each variable in the XGBoost model. This figure is ranked according to the average absolute SHAP value of different variables; the larger the value, the stronger the overall impact of the variable on the income prediction results. As can be seen from Figure 3, variables such as working hours, education level, skill level, platform type, and social security participation are generally highly important, indicating that these factors are key variables affecting income differences among flexible workers. Among them, working hours reflect labor input intensity, education level and skill level reflect differences in human capital, platform type and task type reflect differences in income structure among different flexible employment positions, and social security participation reflects the stability and level of protection of workers' employment.

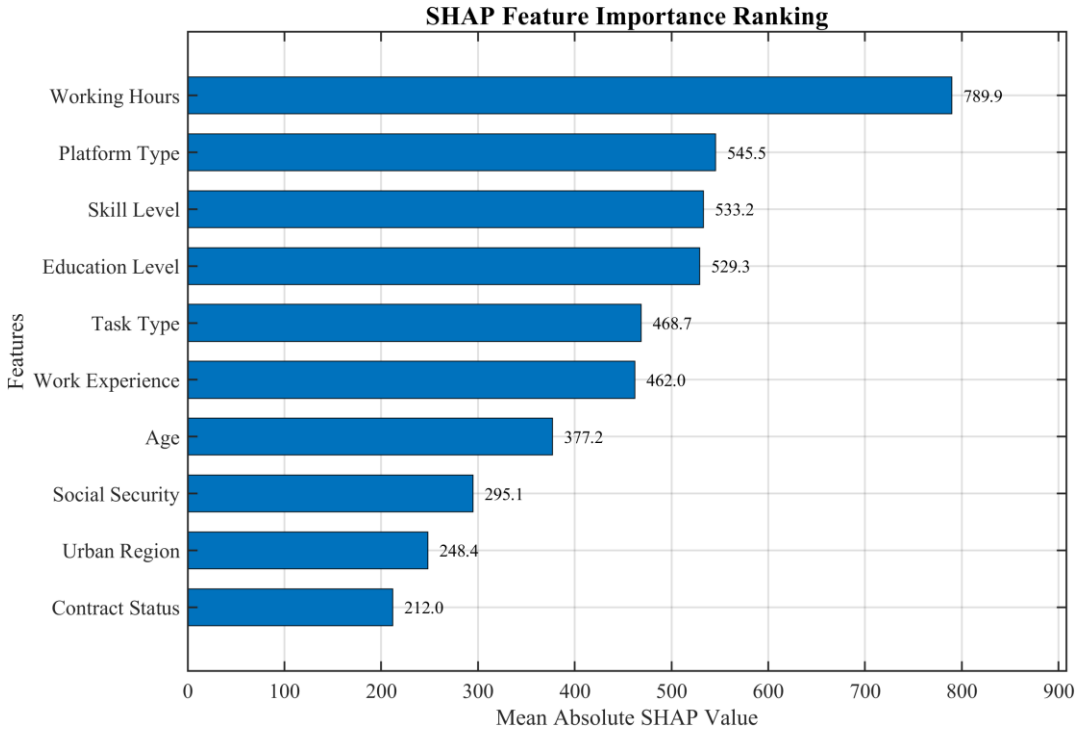

Figure 3: SHAP Feature Importance Ranking.

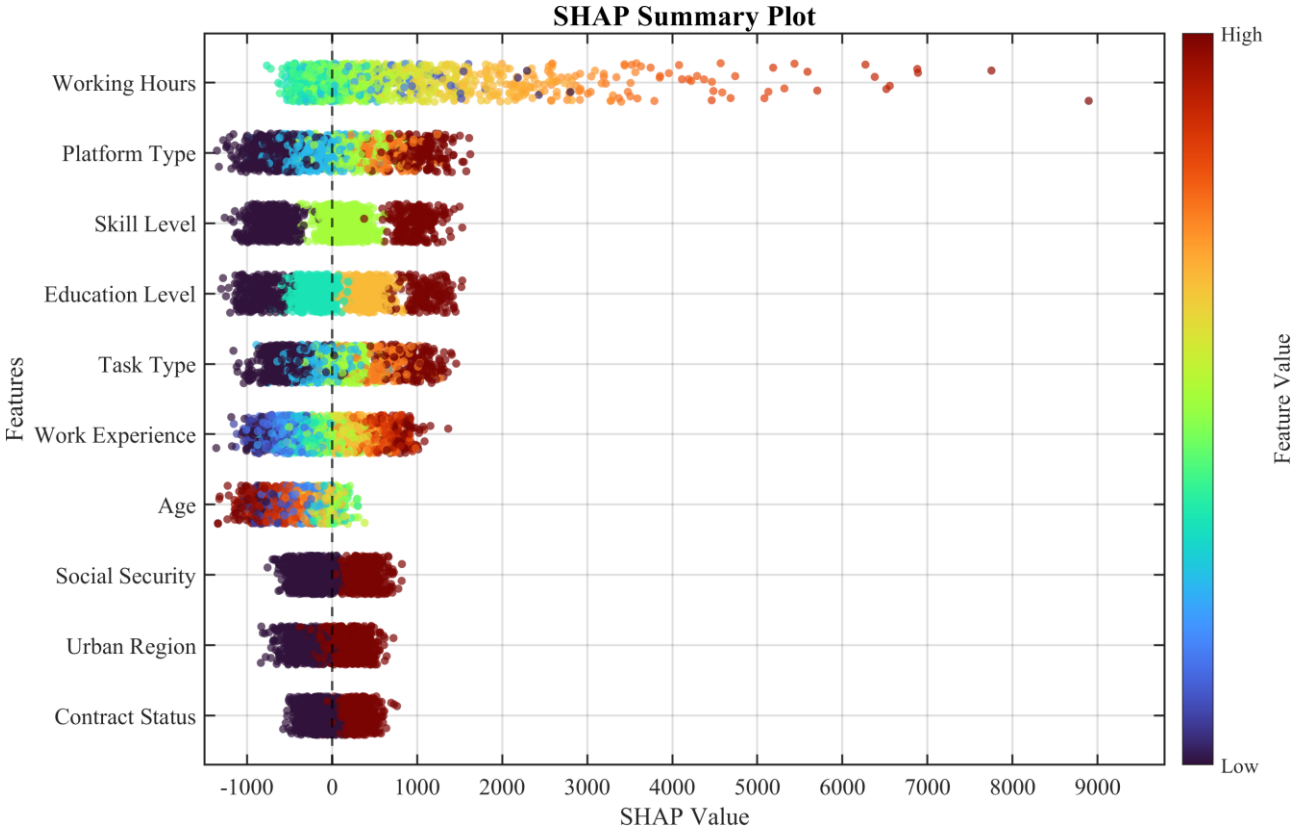

Figure 4: SHAP Summary Plot.

Figure 4 further illustrates the SHAP beehive diagram, used to analyze the direction of influence of different variables on income prediction results. Each point in the figure represents the SHAP value of a sample on a certain variable. The horizontal axis indicates the direction and magnitude of the

variable's contribution to the income prediction result, and the color indicates the level of the variable's value. When the points are mainly distributed on the right side of the horizontal axis, it indicates that the variable tends to increase predicted income; when the points are mainly distributed on the left side of the horizontal axis, it indicates that the variable tends to decrease predicted income. Figure 4 provides a more intuitive way to determine the positive or negative impact of different variables on the income level of flexible workers. For example, longer working hours, higher skill levels, and higher education levels often drive income growth, while a lack of social security, low skill levels, or unstable platform types may be associated with lower income levels.

4. Conclusion

This paper addresses the problem of measuring income inequality and identifying influencing factors among flexible workers, constructing an interpretable machine learning framework based on the XGBoost algorithm and the Shapley additive interpretation method. First, this method uses the Gini coefficient, Theil index, and Lorenz curve to measure the degree of income inequality among flexible workers, characterizing income disparities and concentration within the sample. Then, an XGBoost income prediction model is constructed and compared with multiple linear regression and random forest regression models. Experimental results show that XGBoost outperforms in relevant evaluation indicators, indicating that it can more effectively characterize the nonlinear relationship between the income level of flexible workers and personal attributes, employment characteristics, and social security factors.

Furthermore, this paper introduces the SHAP method to interpret the model's prediction results, thereby identifying key factors influencing income disparities among flexible workers. Results show that variables such as working hours, skill level, education level, platform type, task type, social security participation, and regional characteristics have a significant impact on income prediction results, indicating that income disparities among flexible workers are not determined by a single factor, but rather by the combined effects of multiple factors, including labor input, human capital, platform employment structure, and social security conditions. Overall, the method presented in this paper combines the capabilities of income inequality measurement, income prediction, and model interpretation, providing data support and methodological reference for identifying low-income risks among flexible workers, optimizing platform governance, improving social security, and enhancing vocational skills. Future research could further expand the real-world sample size by incorporating long-term tracking data from different regions and industries to improve the stability and generalizability of the model's conclusions.

Acknowledgements

This research was funded by China University of Labor Relations, grant number “26xjx029" and "The APC was funded by China University of Labor Relations".

References

[1] Altenried M. Mobile workers, contingent labour: Migration, the gig economy and the multiplication of labour[J]. Environment and Planning A: Economy and Space, 2024, 56(4): 1113-1128.
[2] Himani Srihita R. Transformative dynamics of the gig economy: Technological impacts, worker well-being and global research trends[J]. International Journal of Engineering Business Management, 2025, 17: 18479790241310362.
[3] Andabayeva G. Labor market dynamics in developing countries: analysis of employment transformation at the macro-level[J]. Journal of Innovation and Entrepreneurship, 2024, 13(1): 65.
[4] Sankararaman G. Gig economy's impact on workforce dynamics and economic resilience[J]. Available at SSRN 5086559, 2024.

[5] Tariq M U. Balancing flexibility and precarity: Ethical implications and sustainable practices in global gig economies[M]//Sustainability and Adaptability of Gig Economies in Global Business. IGI Global Scientific Publishing, 2025: 323-346.
[6] Au-Yeung T C. The gig economy, platform work, and social policy: food delivery workers' occupational welfare dilemma in Hong Kong[J]. Journal of Social Policy, 2025, 54(2): 673-691.
[7] Alauddin F D A. The influence of digital platforms on gig workers: A systematic literature review[J]. Heliyon, 2025, 11(1).
[8] Griep Y. Sustainable human resource management: The good, the bad, and making it work[J]. Organizational Dynamics, 2025, 54(2): 101112.
[9] Daud S N M. Adapting to the gig economy: Determinants of financial resilience among "Giggers"[J]. Economic Analysis and Policy, 2024, 81: 756-771.
[10] Chrisendo D. Rising income inequality across half of global population and socioecological implications[J]. Nature sustainability, 2025: 1-13.
[11] Karimi A, Lucke C, Palme M. Components of the evolution of income inequality in Sweden, 1990–2021[J]. Fiscal studies, 2024, 45(2): 187-204.
[12] Colak Z. Hybrid machine learning approach in forecasting renewable energy production: Extra trees, catBoost and LightGBM based stacking model[J]. Renewable Energy, 2026, 260: 125167.
[13] LI Y. Application of the adaptive sparrow search algorithm in medical supply engineering[J]. Scientific Reports, 2025, 15: 35775. DOI: 10.1038/s41598-025-11793-2.
[14] FU Z. How compound climate shocks reshape urban resilience and scenario-adaptive pathways: New insights from the Yangtze River Economic Belt[J]. Environmental Impact Assessment Review, 2026, 121: 108563. DOI: 10.1016/j.eiar.2026.108563.
[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