2025 IEEE 8th Information Technology and Mechatronics Engineering Conference (ITOEC) | 979-8-3315-2948-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ITOEC63606.2025.10968729

Logistic Regression and XGBoost Model of Multiple factors on Trust

Roujia Sun[1,2] , Jiajia Zhu[1,2], Wen Zhang[1,2], Yan Mu[1,2]*
1. CAS Key Laboratory of Behavioral Science, Institute of Psychology Beijing, China
2. Department of Psychology, University of Chinese Academy of Sciences Beijing, China
Corresponding Author: Yan Mu Email: muy@psych.ac.cn

***Abstract*—The trust study was begun in the 1960s. Previous research has been particularly focused on understanding the psychological underpinnings of trust formation and sustenance, with influences from economic interests, social identity, and self-actualization. However, the majority of these studies relied on qualitative approaches, with limited research incorporating covariates and a lack of studies comparing the magnitudes of these variables' effects. To bridge this gap, we conducted an online survey research and collected 268 valid questionnaires. The dependent variable was generalized trust, while the independent variables included a set of personality traits and social preference measures (e.g., self-esteem, self-control, anxiety, cultural Tightness Looseness (TL), and Belief in a Just World (BJW)). Leveraging the advantages of machine learning, this study identified the key factors (i.e., BJW, TL) contributing to Trust. The exploration of these mechanisms has been crucial in advancing our understanding of how trust operates in different contexts and at different levels of human interaction.**

***Keywords*−Trust; BJW; TL; Self-esteem; Binomial logistic regression; XGBoost model**

I. Introduction

Trust constitutes a sophisticated and pivotal element within the fabric of human interactions, exerting a significant influence on personal and social well-being, economic activities, and societal advancement. Academic research into trust continues to be a vibrant and essential field, as it seeks to understand and enhance the conditions under which trust can be built and sustained.

Rotter's seminal work in 1967 and 1971 was instrumental in conceptualizing trust as a stable personal trait. He defined trust as “an individual's general belief in the reliability, cooperation, and goodwill of others”.[1] This definition posits that interpersonal trust transcends specific situations and is characterized as “the trustor's psychological state or attitude towards the trustee, whom they rely upon”.

From a sociological perspective, trust is considered a collective characteristic that applies to various groupings (couples, small groups, large groups), rather than being an isolated phenomenon unique to individuals (Homan, 1958)[2]. Putnam suggests that trust, as a form of social capital, primarily stems from norms of reciprocity and networks of civic engagement. In a society characterized by generalized reciprocity, individuals are more willing to extend trust because they can rely on the expectation that their trust will not be exploited by others.[3]

The integrated model encompasses both conventional variables that influence interpersonal trust and individual difference variables. The conventional variables include initial trust, motive transformation, joint evaluation, attribution, emotion, expectations, and perceived trust and security.[4] Individual difference variables, on the other hand, encompass attachment styles, self-esteem, self-concept differences, and the interactions between these and the conventional variables. [5] The model posits that a comprehensive understanding of the specific circumstances of individuals and their partners across each variable is necessary to explain the development of trust relationships.

Notably, previous studies have primarily relied on qualitative methods. In order to provide a more nuanced understanding of the determinants of trust, we employ Binomial Logistic Regression and XGBoost model to quantitatively evaluate the significance of various factors influencing trust. The results reveals that among the many factors that affect trust, the Belief in a Just World (BJW) has the most significant impact on trust, followed by social norms (daily TL), and then personal traits (Implicit self-esteem (IAT) and anxiety levels). These factors collectively contribute to the formation and development of trust relationships.

II. Research Hypothesis

A. Hypothesis 1

The dependent variable Z is influenced by the independent variable X. Initially, we employed a bi-variate Pearson correlation analysis to investigate whether this hypothesis is held.

B. Hypothesis 2

A multivariate linear relationship can be established between Z and X. We utilized binomial logistic regression to model and quantitatively analyze the relationship between the dependent variable and each factor.

C. Hypothesis 3

A multivariate nonlinear quantitative relationship can be established between Z and X. We employed the

XGBoost classification model, which is a machine learning algorithm, to analyze the quantitative relationship between the dependent and independent variables.

We collected a total of 268 valid survey questionnaires. Mage=23.627 ± 5.098 years. The dependent variable trust was measured using the Generalized Trust Scale (GTS) (Yamagishi and Kosug, 1999). The 9 independent variables are as follows: Explicit self-esteem, implicit self-esteem, contingency of self worth, self control, culture TL, daily TL, anxiety, BJW and involution (Neijuan). Binomial logistic regression and XGBoost model were performed using online SPSSAU as shown in Table 1.

TABLE I. STUDY METHOD

Hypothesis	*Method*	*Result*
H1	Pearson correlation	dependent variable association with multiple factors
H2	Binomial Logistic Regression	Evaluation of independent variable contribution rate and classification prediction effect
H3	XGBoost	Evaluation of Importance Factor Ratio and Classification Prediction Effect

III. DATA ANALYSIS AND MODELING PREPARATION

A. Data preprocessing

Data collation: summarize and clean the data, and directly delete the samples with missing values.

Tagging of dependent variable: the continuous dependent variable Y obtained from the questionnaire is normalized to the tagged dependent variable Z.

B. Treatment of dependent variables

If the trust score for the nth sample is denoted as Yn, the collective trust scores of all samples are expressed as vector Y.

$$Y = [Y1, Y2, \cdots , Yn] \quad (1)$$

The top 27% of higher scores in vector Y were classified as "high trust," while the remaining 73% were categorized as "low trust," forming vector Z.

$$Z = [Z_1, Z_2, \dots , Z_n] \quad (2)$$

Here Zi Represents the transformed Yi, δ represents the 73% quantile of Y. The vector Z is the dependent variable.

$$Z_i = \begin{cases} 1, \; if \; y_i > \delta \\ 0, \; if \; y_i \leq \delta \end{cases} \quad (3)$$

C. Independent Variables

$$X = [X_1, X_2, \cdots, X_n] \quad (4)$$

All variables except trust were aggregated into matrix X, constituting the set of independent variables. Each X_n denotes an individual independent variable.

D. Pearson correlation analysis

We conducted Pearson correlation analysis to examine the relationship between the dependent variable (Y) and the independent variables (X). We observed that except implicit self-esteem, the remaining eight independent variables demonstrate a significant correlation with the dependent variable trust, confirming hypothesis 1. The correlation details are presented in Table 2.

TABLE II. CORRELATION WITH TRUST

Variable	*β*	*Variable*	*β*
X1	.493**	X5	.528**
X2	-.166**	X6	-.478**
X3	.459**	X7	.664**
X4	.434**	X8	-.122*

Here, X1 , X2, ⋯ , X8 represent the independent variables: explicit self-esteem, contingency of self-worth, self control, culture TL, daily TL, anxiety, BJW and involution.

IV. BINOMIAL LOGISTIC REGRESSION

A. Principle

Binomial Logistic Regression is a supervised analysis method that primarily explores the correlation between a dependent variable and multiple independent variables. Similar to the principles of multivariate linear regression, it models the quantitative relationship between the dependent variable and several independent variables as a linear function, examining the contribution of each factor to the dependent variable within the model. However, the key difference lies in the fact that the dependent variable becomes a dichotomous variable.

Here, $\widehat{Z}$ represents the dependent variable, while X_1, X_2, ..., X_n represent the independent variables. β_1, β_2, ..., β_n are the regression coefficients, and ε stands for the residuals.

$$\widehat{Z} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (5)$$

B. Procedure and Results

1) Function fit

After conducting logistic regression analysis on the eight independent variables, we discovered the optimal fitted function obtained. The regression details are presented in Table 3.

TABLE III. BINOMIAL LOGISTIC REGRESSION WITH TRUST

Variable	β	*Variable*	β
X1	-0.745	X5	0.763*
X2	-0.403	X6	-0.983
X3	0.052	X7	2.119**
X4	-0.075	X8	0.395

The variables X1 and X2, ⋯ , X8 represent the independent variables: explicit self-esteem, contingency of self-worth, self control, culture TL, daily TL, anxiety, BJW and involution. Variables with a p-value exceeding 0.05 in the regression model are eliminated to ensure that only those with statistically significant effects are retained in the equation.

$$\widehat{Z} = -12.252 + 0.763X5 + 2.119X7 \tag{6}$$

The successful fit of this model supports Hypothesis 2, indicating the establishment of a multivariate logistic function relationship between Z and X.

2) Model Evaluation

The Fig. 1. ROC (Receiver Operating Characteristic Curve) graph combines sensitivity (TPR) and specificity (FPR) to assess the Binomial Logistic Regression. Ideally, TPR should approach 1 and FPR should approach 0.

$$TPR = TP / (TP + FN) \tag{7}$$

$$FPR = FP / (FP + TN) \tag{8}$$

· TP: True Positive Rate, predicting the positive class as positive.

· TN: True Negative Rate, predicting the negative class as negative.

· FP: False Positive Rate, predicting the negative class as positive (false alarm)

· FN: False Negative Rate, predicting the positive class as positive (underreporting)

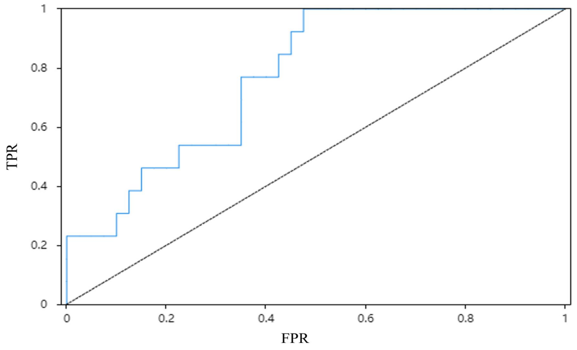

Fig. 1. ROC

By quantifying the performance metrics, the classification effect of the logistic regression model was further evaluated to be good. The model achieved an accuracy of 84.62%, recall of 84.62%, precision of 85.38%, F1-score of 0.83, and an AUC of 0.905. The accuracy was calculated using the formula:

$$Accuracy=TP+TN \;/\; (TP+TN+FN+FP) = 84.62\% \tag{9}$$

$$Precision(PR) = TP / (TP + FP) = 85.38\% \tag{10}$$

$$Recall(RE) = TP / (TP + FN) = 84.62\%. \tag{11}$$

$$F1 = 2 \times PR \times RE / (PR + RE) = 0.83 \tag{12}$$

The training set ratio is set to 0.7, utilizing the L-BFGS optimization algorithm, and employing L2 regularization. The model performance is deemed acceptable.

V. XGBOOST MODEL

A. *Principle of Xgboost Algorithm*

Boosting is a commonly used statistical learning method. In the training process, by changing the weight of the training sample, learning multiple classifiers, and finally obtaining the optimal classifier. After each round of training, reduce the weight of the correctly classified training sample and increase the weight of the incorrectly classified sample. After multiple trainings, some of the incorrectly classified training samples will get more attention, and the correct training sample weight tends to 0. Multiple simple classifiers are obtained, and a final model is obtained by combining these classifiers. Xgboost algorithm [6] is based on traditional Boosting, using CPU multi-threading, introducing regularization items, adding pruning, and controlling the complexity of the model.

B. *Training model*

The prediction problem of trust level is a typical binary classification problem, which belongs to the application category of XGBOOST algorithm. The blood data of this experiment consists of 268 sets, and the dataset is randomly divided into training set and test set with the ratio of 7:3 consistent with logistic regression. . After that, the tuning parameter work is carried out, and the parameters of XGBoost are divided into three kinds, i.e., the generic parameters, the task parameters, and the command line parameters. The task parameters are used to control the resultant metrics and the desired optimisation objective, and the generic parameters are the core of the tuning in this paper. In this paper, the decision tree is chosen as the weak learner, and the parallel thread is the maximum to ensure the running speed. The order of hyperparameter tuning and specific value settings are shown in Table 4.

TABLE IV. Model Parameter Values

Model evaluation effect	Accuracy rate	88.46%
	Precision rate	89.09%
	Recall rate	88.46%
	F1-score	0.886

C. *Model evaluation*

The model evaluation yielded an F1-score of 0.886, along with an accuracy of 88.46%, The model performance is deemed acceptable. The fig.2. represents the proportion of all samples with correct predictions in the total sample of the model. Furthermore, we have confirmed Hypothesis 3.

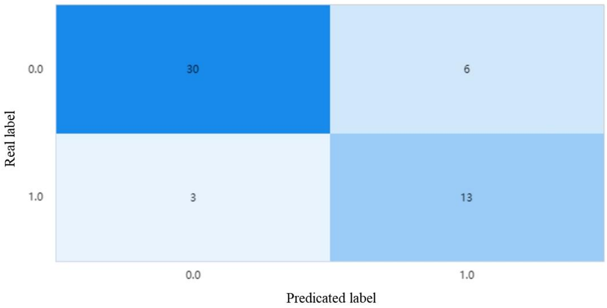

Fig. 2. Confusion Matrix

D. *Output results*

The prediction results were evaluated by the determination coefficient R2 and the root mean square error of the correction set.The Fig.3. represents the proportion of all independent with correct predictions in the model.

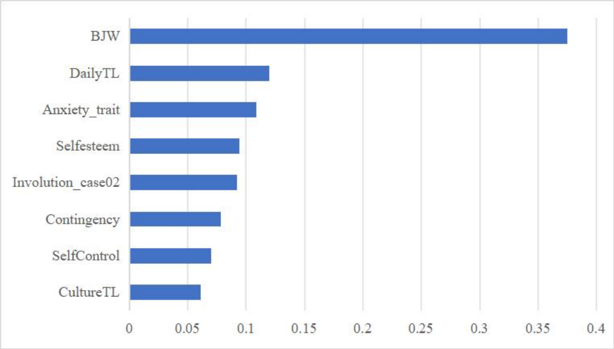

Fig. 3. Feature Importance Plot

E. *Performance comparison between logistic regression and XGBoost model*

In the analysis, we compared the binomial logistic regression and XGBoost models, and found that both methods achieved similar results, with classification accuracy of 80.5% and 88.46%, respectively as Fig 4 Comparing the accuracy of the two models, it can be seen that the prediction performance of the XGBoost model is significantly better than that of the logistic regression model.

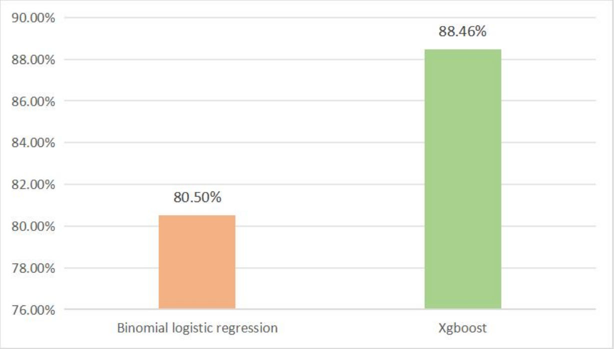

Fig. 4. Classification accuracy

VI. Discuss

In the analysis, we compared the binomial Logistic regression and XGBoost models. Across both modeling techniques, there was a consistent and significant influence on high empathy factors, primarily attributed to BJW and daily TL.

Studies indicate that the belief in a just world (BJW) plays a moderating role between social cognition and prosocial behavior. Individuals with a strong BJW exhibit a stronger positive correlation between social self-efficacy and prosocial behavior. Those with a heightened BJW are more inclined to approach problems with positive thoughts and attitudes, and they engage in more prosocial behavior than others.

Social norms have a secondary impact on trust: Robert D. Putnam, in his book "Making Democracy Work," suggests that trust, as a form of social capital, primarily stems from norms of reciprocity and networks of civic engagement. In a society characterized by generalized reciprocity, individuals are more willing to extend trust because they can rely on the expectation that their trust will not be exploited by others.

VII. Conclusion

The research results show that the Xgboost predictive model can effectively predict the trust level of an individual with BJW, daily TL, anxiety level and other personal traits. Our study explores on the factors and mechanisms that drive empathy, enriching the current understanding and introducing innovative perspectives and methodologies for empathy research. Moreover, these findings have practical significance for fostering trust across different contexts and among diverse groups.

REFERENCES

[1] Rotter, J. B. (1967). A new scale for the measurement of interp ersonal trust. Journal of Personality, 35

[2] Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. (2003). An atlas of interpersonal situations. New York: Cambridge University P

[3] Huesmann, L. R., & Levin ger, G. (1976). Incremental exchange theory: A formal model for progression in dyadic social interaction. Ad- vances in Experimental Social Psychology, 9

[4] Frederik Schwerter and Florian Zimmermann (2020). Determinants of trust: The role of personal experiences. Games and Economic Behavior Volume 122, July 2020, Pages 413-425

[5] Simpson, J. A. (2007). Psychological foundations of trust. Current Directions in Psychological Science, 4

[6] T. Chen, S. Singh, B. Taskar, and C. Guestrin. Efficient second-order gradient boosting for conditional random fields. In Proceeding of 18th Artificial Intelligence and Statistics Conference (AISTATS’ 15), volume 1, 2015. Shandong normal University. 3, 11-14