Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 2, 2026

Research on a Dynamic Behavior Analysis Framework Based on Random Forest and Markov Chain

Xiyao Liu *, Xiangrui Li, Yuxuan Qiu

Dalian University of Technology, Dalian, China

* **Corresponding author**: (Email: lxy4419@126.com)

Abstract: This paper constructs a comprehensive analytical framework integrating constrained optimization, machine learning and dynamic transition modeling based on multi-season competition data. First, statistical feature analysis and correlation tests are adopted to model the stability of the scoring system. On this basis, a constrained convex optimization model is established to inversely estimate unobservable audience voting ratios, and the Spearman correlation coefficient is used for result verification to enhance the reliability and consistency of estimations. Next, the random forest regression model combined with the SHAP interpretation method is applied to quantitatively analyze key factors affecting scoring and voting results, realizing the importance assessment of multi-dimensional features and interpretable modeling. Furthermore, a Markov chain-based dynamic transition model is developed. With the Softmax probability allocation mechanism and utility function incorporated, this model simulates the transition of audience support across different stages and characterizes the dynamic evolution of popularity in competitive scenarios. Experimental results demonstrate that the proposed model can accurately recover hidden voting distributions and effectively capture result discrepancies and dynamic variations under different rules. Featuring good universality, stability and scalability, this method can serve as a reference for complex behavior prediction and dynamic decision analysis.

Keywords: Random Forest, Markov Chain, Dynamic Migration Modeling.

1. Introduction

Against the increasingly intensive interaction of multi-source information in complex data environments, traditional analytical methods relying on single statistical indicators or static rules can hardly accurately depict the evolution of dynamic behaviors. Especially in scenarios involving scoring, ranking, preference transition and phased elimination mechanisms, distinct nonlinear correlations and temporal coupling exist among various factors. Simple regression or empirical rules tend to result in poor interpretability and fail to effectively recover latent behavioral variables. Accordingly, developing an integrated algorithm framework with robust prediction performance, interpretability and dynamic modeling capability has become a key research direction for complex behavior analysis.

Most existing studies focus on optimizing individual models, such as applying conventional machine learning for outcome prediction or adopting static probabilistic models to analyze behavioral distributions. However, they are deficient in the unified modeling of dynamic state transition, latent variable estimation and coupled effects of multiple mechanisms [1, 2]. In addition, some methods lack stable constraints when handling phased data evolution, which may generate results inconsistent with actual rules.

To address the above limitations, this paper proposes a multi-level algorithm framework that integrates constrained optimization, random forest, SHAP interpretation model and Markov chain-based dynamic transition mechanism. Convex optimization is adopted to realize the inversion of latent variables. Feature importance analysis is introduced to enhance model interpretability, and the state transition mechanism is combined to characterize the evolution of dynamic support behaviors. This framework unifies static feature analysis and dynamic process simulation. Validated on competition scoring and audience support behaviors, the proposed model proves stable and adaptable under complex rules. It provides a general analytical approach for multi-stage behavior prediction and dynamic decision modeling.

2. Data Preprocessing Framework Based on Statistical Features and Correlation Analysis

Before conducting vote inversion and audience behavior analysis, basic statistical processing was first applied to the competition data, mainly focusing on two questions: whether scores changed as the competition progressed, and whether scoring remained consistent across different judges.

2.1. Data Overview

The cleaned dataset contains 34 seasons, 421 contestants, and 66 variables, covering contestant attributes, weekly performance, judges' scores, final rankings, and related information. Some variables with substantial missing values or obvious anomalies had already been removed beforehand.

2.2. Temporal Changes in Competition Scores

After calculating the valid average scores for each week, it can be observed that contestants' average scores generally increased as the competition progressed, as shown in Figure 1.

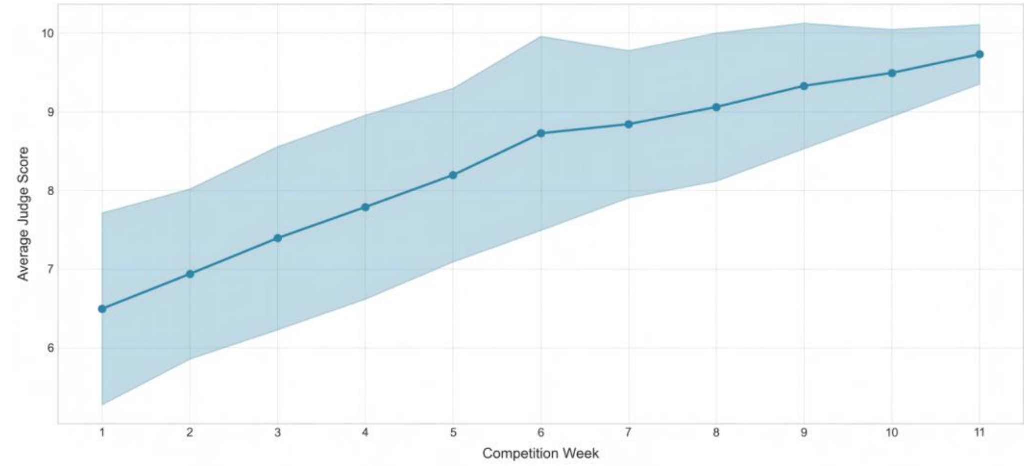

Figure 1. Weekly average judge score trend with standard deviation

The average score was around 6.2 in the early stages and reached nearly 9.1 in the later stages. A clear positive correlation exists between competition week and average score, with a Spearman correlation coefficient of $\rho = 0.872$, satisfying $p < 0.001$.

This pattern is fairly consistent with the nature of the competition process itself. Contestants with lower scores are gradually eliminated, while those who remain usually have higher overall ability. In addition, with increased training time, performance quality and stability tend to improve compared with earlier weeks. The scoring scale in later stages may also differ slightly from that in the beginning, which could have some influence.

The standard deviation for each week remained mostly between 0.8 and 1.2, with relatively limited fluctuations. In other words, although performance differences between contestants do exist, the overall scoring system remained reasonably stable, without particularly obvious abnormal distributions.

2.3. Analysis of Judge Scoring Consistency

To evaluate the stability of the scoring system, Pearson correlation analysis was conducted on the judges' scoring results, with outcomes shown in Figure 2.

Figure 2. Judge score correlation heatmap

The correlations between different judges were generally high, with an average correlation coefficient of 0.78. The highest correlation reached 0.85, while the lowest remained above 0.69. Although some differences in scoring preferences existed among judges, their overall evaluations of contestant performance were broadly consistent.

Therefore, using the "weekly average score" as a comprehensive indicator of professional evaluation is reasonable for subsequent analysis, and it can also reduce redundant effects introduced by multiple judges' scores.

2.4. Summary

The statistical results indicate that competition scores show clear temporal evolution patterns, while judges' scores maintain relatively high consistency overall. Subsequent vote inversion and audience migration analyses can be directly built upon this foundation.

3. Hidden Vote Inversion and Credibility Modeling Based on Constrained Optimization

Since audience voting data are not publicly available, this study estimated audience vote proportions by combining competition rules with scoring outcomes. Overall, a constrained convex optimization model was constructed using CVXPY, and the estimated results were validated through Spearman correlation analysis. Based on this, further analysis was conducted on how the credibility of the estimates changes across different competition stages.

3.1. Vote Proportion Inversion and Result Validation

Audience voting data in this type of competition are usually inaccessible, making it difficult to recover actual voting patterns using only public rankings. Completely unconstrained estimation may also produce results that clearly violate competition rules, such as unusually concentrated votes for individual contestants. Therefore, competition rules were directly incorporated into the optimization process, with constraints imposed on the feasible range of vote proportions.

After data cleaning, 34 seasons, 412 valid weeks, and 1480 contestant–week samples were retained. Since scoring mechanisms differed across seasons, corresponding constraint models were built separately. The differences between the two main competition formats are shown in Table 1.

During model optimization, all vote proportions satisfied the basic conditions of $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 100\%$, while elimination rules were incorporated into dynamic week-by-week simulations. The main purpose was to avoid estimation results with obvious distortions. Compared with methods that only fit rankings, constrained optimization produces results that align more closely with actual competition logic.

Table 1. Comparison Of Competition Scoring Rules

Voting Mechanism	Core Aggregation Principle	Behavioral Bias
Ranking-based	Relative ranking aggregation between judge scores and fan votes	Stronger fan-driven bias
Percentage-based	Weighted aggregation based on vote proportions	More balanced professional influence

To evaluate the reliability of the inversion results, the Spearman correlation coefficient was used to test ranking consistency between estimated results and actual advancement weeks. The calculation is given by:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{1}$$

The results are shown in Figure 3. Overall correlation remained at a relatively high level, suggesting that the method can recover the relative distribution of audience votes with reasonable stability.

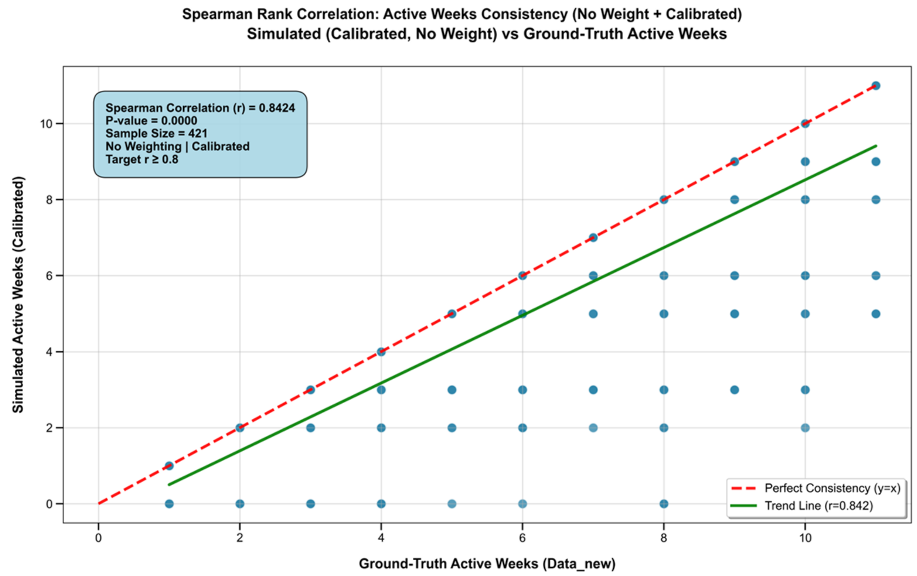

Figure 3. Spearman's test

3.2. Credibility Analysis of Vote Estimation

After obtaining vote estimates, further analysis was performed on estimation credibility. Here, credibility mainly reflects uncertainty in the model results, with sources including sample size, constraint satisfaction, and differences in original proportion distributions.

To address this, a two-level credibility evaluation framework at the individual and group levels was constructed. The normalized residual is defined as:

$$RE_{i,s,w} = \frac{|p_{i,s,w} - \hat{p}_{i,s,w}|}{\hat{p}_{i,s,w}} \times 100\% \tag{2}$$

Furthermore, the individual credibility indicator is defined as:

$$CI_{i,s,w} = 0.7 \times (100 - RE_{i,s,w}) + 0.3 \times SW_{i,s,w} \tag{3}$$

Where SW represents the sample size weight.

The analysis shows that contestants with higher levels of controversy generally have noticeably lower credibility than typical contestants, and this difference becomes more pronounced during the final stages. At the same time, credibility in regular weeks is generally higher than during later elimination stages, indicating that as the number of remaining contestants decreases, small-sample effects substantially increase estimation uncertainty.

Overall, the proposed method maintains relatively stable inversion performance across most competition stages, while also reflecting the influence of different competition formats and sample sizes on result reliability.

4. Dynamic Decision Impact Analysis Under Multi-Mechanism Scoring Systems

After completing audience vote inversion, further comparisons were made regarding the effects of two different voting mechanisms on competition outcomes. The analysis mainly focuses on two questions: how much elimination results differ under different rules, and to what extent audience voting dominates the final outcomes.

4.1. Outcome Differences Under Different Voting Mechanisms

To quantify differences between the two mechanisms, the elimination disagreement rate DR was defined as:

$$DR = \frac{\sum_{s=1}^{34}\sum_{w=1}^{W_s} I(E_{s,w,rank} \neq E_{s,w,pct})}{\sum_{s=1}^{34} W_s} \times 100\% \tag{4}$$

Where $I(\cdot)$ is an indicator function, taking the value 1 when the eliminated contestants differ under the two mechanisms, and 0 otherwise.

Statistical results indicate that approximately 28% of competition weeks produce different elimination outcomes under the two mechanisms. This suggests that scoring rules themselves can significantly affect final advancement results, rather than merely changing ranking order. Further significance testing yielded $Z = -9.87$, satisfying $p < 0.001$, indicating that the observed differences are unlikely to be caused by random fluctuations.

To further examine the influence of audience voting under different mechanisms, the fan contribution ratio FCR was introduced:

$$FCR_{i,s,w,type} = \frac{S_{i,s,w,fan,type}}{S_{i,s,w,total,type}} \times 100\% \tag{5}$$

Where S_{fan} denotes audience scores and S_{total} represents overall scores.

The overall results show that the average audience contribution ratio reached 32.7% under the ranking-based system, compared with only 21.4% under the percentage-based system. Meanwhile, the Spearman correlation coefficient between audience scores and final overall rankings reached $\rho = 0.75$ under the ranking-based system, noticeably higher than $\rho = 0.58$ under the percentage-based system. This indicates that the ranking-based mechanism amplifies audience preferences more strongly and tends to favor highly popular contestants.

Further examination of more controversial contestants reveals that this difference becomes even larger. Under the ranking-based system, these contestants tend to achieve higher weekly rankings, whereas their overall rankings drop noticeably under the percentage-based system. In other words, when audience support conflicts with professional evaluations, the ranking-based system is more likely to favor highly popular contestants.

4.2. Effects of the Judge Intervention Mechanism

Based on the original rules, an additional mechanism was introduced in which judges decide the eliminated contestant from the bottom two candidates. The effects of this mechanism on competition outcomes were then analyzed. For this purpose, the elimination replacement rate SR was defined as:

$$SR = \frac{\sum_{s=1}^{34}\sum_{w=1}^{W_s} I(E_{s,w,no-mech} \neq E_{s,w,mech})}{\sum_{s=1}^{34} W_s} \times 100\% \tag{6}$$

The results show that after introducing this mechanism, elimination outcomes changed in around 28% of competition weeks. This suggests that judge intervention is not merely a procedural adjustment but can genuinely alter competition results.

Next, consistency between overall rankings and judge score rankings was compared. Without the mechanism, the Spearman correlation coefficient between the two was $\rho = 0.62$; after introducing the mechanism, it increased to $\rho = 0.78$, with $p < 0.001$.

Overall, this mechanism reduces biases caused by relying too heavily on popularity to some extent, making final outcomes more aligned with professional evaluations. However, it does not completely remove the influence of audience voting, allowing a certain degree of audience interaction and entertainment value to remain.

5. A Dynamic Behavior Migration Framework Integrating Random Forest and Markov Chain

To analyze how different contestant characteristics affect professional scores and audience votes, a feature interpretation framework was constructed using random forests and SHAP methods [3]. Based on this, a Markov chain was further applied to simulate dynamic changes in audience support throughout the competition [4]. Compared with purely static analysis, this approach makes it easier to observe how audience preferences gradually shift over time.

5.1. Feature Impact Analysis

Since sample sizes vary across different types of contestants, the data were first standardized and processed using nearest-neighbor matching to reduce the influence of imbalanced sample distributions as much as possible. Variables including age, occupation type, and partner characteristics were then used as inputs to build separate random forest regression models for professional scores and audience votes:

$$\hat{y} = \frac{1}{K}\sum_{k=1}^{K} f_k(X) \tag{7}$$

Where K denotes the number of decision trees, and $f_k(X)$ represents the prediction result of the k-th tree.

After model training, SHAP values were further used to analyze the contribution of different features to prediction outcomes. Figure 4 presents the importance ranking of the main features.

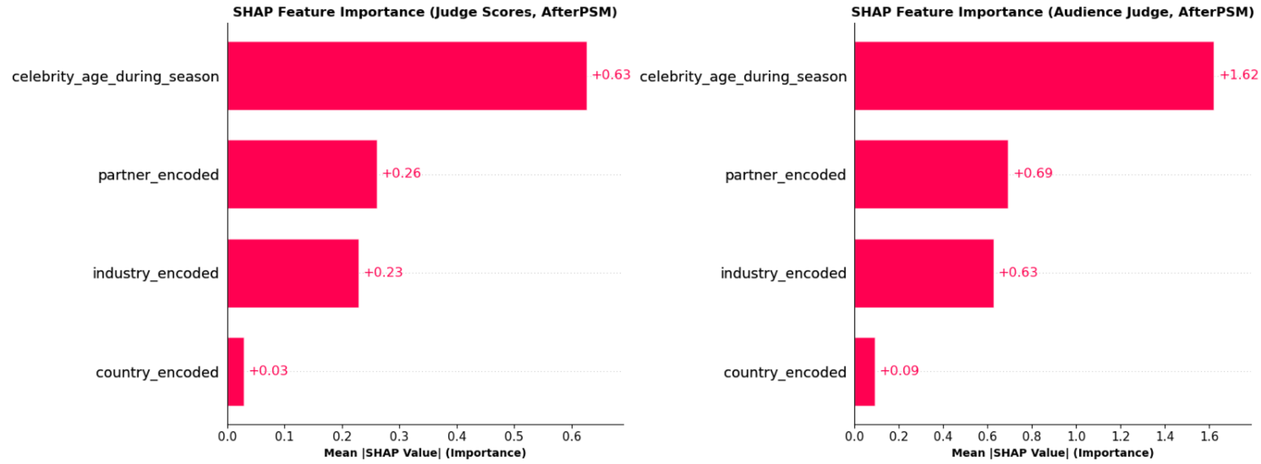

Figure 4. SHAP feature importance for judge scores and audience votes (after PSM)

The results indicate that occupational background, age, and partner ability all influence final outcomes, although professional scores and audience votes do not focus on exactly the same factors. Judge evaluations tend to place greater emphasis on contestants' performance and consistency, whereas audience voting is more easily influenced by factors such as occupational exposure and public recognition.

To visualize the effects of feature variation more directly, SHAP scatter plots were also used for analysis, with results shown in Figure 5.

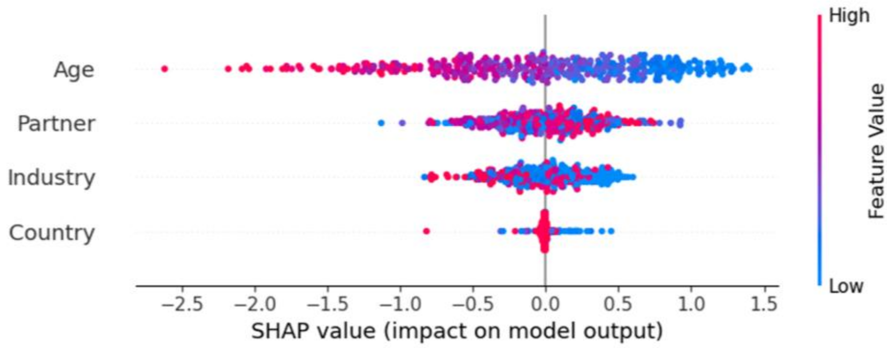

Figure 5. SHAP summary scatter plot for judge scores

The influence of some features does not follow a simple linear pattern. For example, occupations with higher public exposure often attract stronger audience support but do not provide clear advantages in professional scoring. In contrast, some contestants with stronger technical ability may not receive the highest audience attention, yet are more likely to gain recognition from judges.

Overall, professional evaluations and audience preferences are not fully aligned. There is some overlap between the two, but also relatively clear differences.

5.2. Dynamic Modeling of Audience Migration

After completing the static feature analysis, the dynamic evolution of audience support was further modeled. A Markov chain was used here to describe audience migration behavior throughout the competition, making it possible to simulate how fan support is redistributed after contestant eliminations [5].

The model assumes that audience migration mainly occurs within core fan groups, and that migration targets are associated with contestant feature similarity and winning probability. At the same time, the audience distribution in the current week is assumed to depend only on the previous week's state, satisfying the Markov property.

To quantify audience preferences when selecting new contestants to support, a utility function was introduced:

$$U(c)=\sum_{i=1}^{k} w_i \cdot Sim(x_{i,elim}, x_{i,c}) + w_{win} \cdot P_{win}(c) \tag{8}$$

Where $Sim(\cdot)$ denotes feature similarity between contestants, and $P_{win}(c)$ represents the current winning probability of the target contestant. Feature weights w_i were derived from the importance results of the random forest model described earlier [6].

At the same time, considering that some audience members stop participating after their preferred contestant is eliminated, the fan attrition rate was further defined as:

$$L(C_{elim}) = L_0[1+\alpha \cdot Uni(C_{elim}) - \beta \cdot AvgSim(C_{elim})] \tag{9}$$

Based on this, audience migration probabilities were calculated using a Softmax formulation:

$$P(F_{elim} \to c) = \frac{\exp(U(c))}{\sum_{j \in Survivors} \exp(U(j))} \tag{10}$$

The resulting increase in support is then given by:

$$\Delta V_c = V_{elim} \times (1 - L(C_{elim})) \times P(F_{elim} \to c) \quad (11)$$

The corresponding state transition process is illustrated in Figure 6.

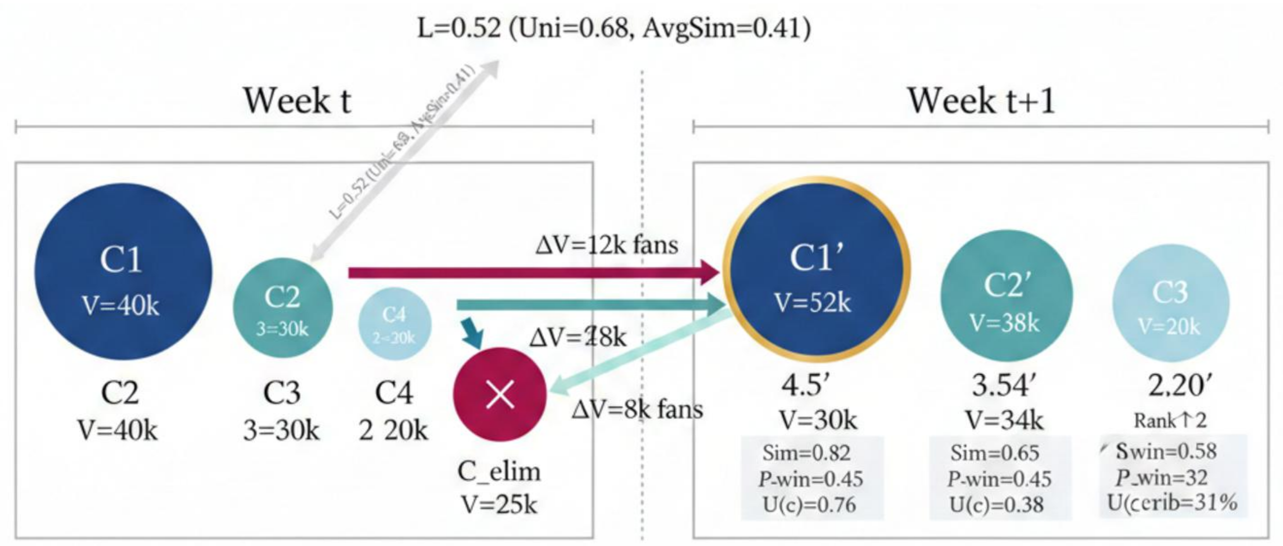

Figure 6. Markov Chain simulation of fan migration dynamics (week t to week $t+1$)

Simulation results suggest that audience migration does not occur randomly. Instead, support tends to shift toward contestants with similar characteristics and stronger current performance. This dynamic redistribution mechanism helps explain the rapid ranking changes often observed in later competition stages, and also indicates that audience behavior itself continues to influence subsequent outcomes.

6. Conclusions

This paper establishes an integrated algorithm model combining constrained optimization, random forest and Markov chain to systematically analyze competition data of 421 contestants across 34 seasons. The results indicate that ranking-based and percentage-based mechanisms lead to divergent elimination outcomes in approximately 28% of competition weeks. The contribution ratio of audience votes reaches 32.7% under the ranking mechanism, notably higher than 21.4% under the percentage mechanism, which demonstrates the substantial impact of institutional design on behavioral outcomes.

Furthermore, SHAP feature analysis reveals that age, occupational background and partners' capabilities exert differentiated effects on scoring and voting, indicating inconsistencies between professional evaluation and public preference. Overall, the proposed model effectively fuses multi-source information to realize the inversion of latent behavioral variables and simulation of dynamic evolution, serving as a quantitative tool for social behavior research. Future work will further explore the model's adaptability for cross-domain and multi-stage behavior prediction, as well as its expandability under complex rule scenarios.

References

[1] Liu, J., Luo, H., Yu, P., et al. (2023). Development and evaluation of a machine learning-based model for predicting hospital-acquired infections in patients with acute ischemic stroke. Chinese Journal of Infection Control, 22(2), 129–135.

[2] Li, J., Wei, Y., & Xue, H. (2024). Analysis of risk factors for influenza-like illness using a random forest model and the SHAP algorithm. Information Technology and Informatization, (2), 3–6.

[3] Luo, Y., Wang, C., & Ye, W. (2022). An explainable prediction model for acute kidney injury based on XGBoost and SHAP. Journal of Electronics and Information Technology, 44(1), 27–38.

[4] Wang, X., Wang, L., & Shi, P. (2025). Markov chain-based prediction of regional trends in COVID-19 prevalence. Journal of Shaanxi University of Science and Technology, 43(1), 211–218. https://doi.org/10.19481/j.cnki.issn2096-398x.2025.01.027

[5] Xie, Y., & Tian, Q. (2020). Reliability analysis of CTCS wireless communication based on interactive Markov chains. Journal of Railway Engineering, 42(5), 84–90.

[6] Chen, M., Chen, S., Wang, Y., et al. (2023). Opportunity-constrained optimization scheduling of an integrated electricity-hydrogen energy system considering hydrogen green certificates. Electric Power Automation Equipment, 43(12), 206–213. https://doi.org/10.16081/j.epae.202309024