ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

Voting Fairness Research via Inverse Optimization and Dynamic Gaming

Zhang Yucheng[1,a], Liu Hao[1,b,*], Luo Shunan[1,c]

[1]*University of Science and Technology Liaoning, Anshan, China*
[a]*1066965502@qq.com*, [b]*haoliu@ustl.edu.cn*, [c]*1026485105@qq.com*
[*]*Corresponding author*

Abstract: *This study examines the persistent structural divergence between professional judging, which emphasizes technical skill, and public voting, which tends to reflect popularity, in Dancing with the Stars [1]. Utilizing 29 seasons of historical data, it systematically addresses four progressive modeling challenges related to vote integration fairness: reconstructing undisclosed fan vote distributions, quantifying the impact of different voting mechanisms, isolating the influence of contestant characteristics, and designing a validated fairness-enhancing system [2]. Through an integrated methodology combining inverse engineering, comparative simulation [3], SHAP-interpretable machine learning, and dynamic optimization, the study offers actionable insights for balancing competitive integrity with audience engagement [4]. To estimate fan votes, an inverse engineering model incorporating Monte Carlo simulation and Sequential Least Squares Programming (SLSQP) was developed, achieving a 64.45% global consistency rate with historical elimination outcomes and a mean estimation uncertainty of 0.52% [5]. Parallel simulations comparing Rank-based (Seasons 1–2, 28–34) and Percentage-based (Seasons 3–27) voting systems revealed an 18.36% systematic bias in the latter, which disproportionately amplifies audience influence and increases the survival probability of "low-skill, high-popularity" contestants by 18.3%. Furthermore, analysis of 12 documented controversial cases indicated that conditional judge arbitration for bottom-two contestants reduced unjust outcomes by 65% without adversely affecting viewership metrics [6]. To assess contestant characteristics, dual XGBoost regression models with SHAP interpretation quantified the divergent evaluation criteria: judge scores are predominantly driven by professional dancer identity (58.0% contribution), whereas fan votes are primarily influenced by celebrity industry type (42.0% contribution). The near-complete independence of these two evaluation systems is confirmed by a Disagreement Index of 1.0 [7], underscoring the fundamental "skill versus popularity" dichotomy. Finally, a dynamic hybrid voting system was designed and empirically validated: a Rank-based method is applied during the skill development phase (Weeks 1–6), transitioning to a Percentage-based method for the fan-driven finale (Weeks 7+), with conditional judge arbitration activated when score divergence exceeds 20%. This proposed framework reduces the survival probability of highly controversial contestants by 29.1% and attenuates the negative correlation between fairness and audience engagement by 43.2%, thereby achieving an optimal equilibrium between technical merit and entertainment value.*

Keywords: *Inverse Engineering; Monte Carlo Simulation; Mechanism Comparison; XGBoost; Format Optimization*

1. Introduction

Dancing with the Stars (DWTS) is a global televised dance competition that integrates professional judging with public voting, creating an ongoing tension between technical evaluation and popular preference.

The show has employed two vote-aggregation methods: rank-based and percentage-based. Notable controversial outcomes highlight persistent fairness concerns within the competition mechanism.

Through data-driven modeling, this study estimates audience vote distributions, evaluates integration methods, and analyzes key factors influencing judge scores and viewer votes.

Ultimately, we propose optimized competition formats to enhance fairness and public engagement, providing evidence-based insights for the show’s production team.

ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

2. Problem Analysis and Modeling Approach

2.1 Problem 1: Fan Vote Estimation as an Inverse Optimization Problem

Problem 1 requires building a model to estimate the undisclosed fan vote data for each week across all seasons. The core challenge lies in the fact that while weekly elimination outcomes are known, the actual vote distribution is strictly confidential, and the model must simultaneously accommodate two different historical vote-integration rules (percentage-based and rank-based). Thus, the analysis focuses on how to leverage observable judge scores and actual elimination results to inversely infer plausible vote distributions under rule-based constraints, while also requiring quantitative evaluation of the reliability—both consistency and uncertainty—of the estimates.

2.2 Problem 2: Comparative Mechanism Analysis via Dual Simulation

Based on the estimated fan-vote distributions derived in Problem 1, we proceed to a comparative mechanism analysis. This phase quantifies the discrepancy between the two historical vote-integration rules—ranking-based versus percentage-based—through systematic simulation across all seasons. We further analyze how these mechanisms affect controversial contestants where judge and fan opinions diverge, evaluate the impact of introducing a judge intervention step, and provide data-driven recommendations for future season designs.

2.3 Problem 3: Factor Attribution Analysis using Machine Learning

Problem 3 aims to quantify the influence of professional dancers' characteristics and celebrity attributes on contestants' performance, and to compare how these factors differentially affect judge scores and fan votes. The analysis reveals a structural tension within the show's evaluation system: professional judging prioritizes technical skill and training, while public voting is driven by non-technical factors such as celebrity appeal and industry background. Consequently, this study must not only identify key predictors of competition outcomes but also, through a comparative analysis of the dual evaluation dimensions, elucidate the fundamental divergence and potential alignment in their underlying logics.

2.4 Problem 4: System Optimization through Dynamic Rule Design

Problem 4 requires designing a new system that integrates fan votes and judge scores to enhance the program's fairness and appeal. Existing methods (the Ranking Method and Percentage Method) employ fixed weights, rendering them ineffective in handling "controversial contestants" where judge and fan opinions diverge, leading to historical fairness disputes. Therefore, the new system must introduce a dynamic weighting mechanism that can adaptively adjust the weights of the two score types based on a contestant's weekly controversy level or elimination risk, thereby ensuring the appropriate influence of technical judgment while maintaining the program's entertainment value.

3. Data Preparation and Exploratory Analysis

To ensure the quality and consistency of the model's input data, systematic cleaning and transformation were performed on the official dataset.First, textual missing values “N/A” were replaced with numerical NaN, and all judge score columns were cast to numeric types. The calculated average judge score for each week was standardized via Z-score normalization (Formula (1)) to eliminate scale differences between judges or across weeks, granting scores cross-temporal comparability.

$$z_{i,\omega} = \frac{x_{i,\omega} - \mu_{\omega}}{\sigma_{\omega}} \tag{1}$$

Externally sourced contestant popularity data was processed using 3σ principle outlier filtering and hierarchical imputation to provide stable prior information. Finally, the wide-format data was transformed into a long format suitable for time-series modeling, with precise weekly elimination labels parsed and generated.

ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

4. Model Development and Results

4.1 Analysis and Solving of Question One

The model solution is conducted in two phases: preliminary estimation via Monte Carlo simulation and in-depth validation via SLSQP.

Construct Composite Score: For a given week, the composite score $S_{i,\omega}$ for each contestant is calculated using Formula (2), where the weight for standardized judge score $\alpha = 0.05$ and the weight for fan vote share $\beta = 0.95$, highlighting the dominant role of fan votes.

$$S_{i,\omega} = \alpha \cdot z_{i,\omega}^{judge} + \beta \cdot \overline{p}_{i,\omega} \tag{2}$$

Apply Rule Constraints: Apply constraints based on the rule type applicable for that week.

Percentage-based Constraint: The composite score of all contestants in the eliminated set E_{ω} must be lower than the lowest composite score among advancing contestants, i.e., $\forall i \in E_{\omega}, S_{i,\omega} < \min_{j \notin E\omega} S_{j,\omega}$.

4.2 Analysis and Solving of Question Two

4.2.1 Methodology: Model Establishment (Dual-Mechanism Simulation Framework)

This study establishes a comparative simulation framework for the ranking-based method (applied in Seasons 1, 2, 28-34) and the percentage-based method (applied in Seasons 3-27), with mathematical models strictly adhering to competition rules.As shown in Table 1.

Table 1: Mathematical Models of Two Elimination Mechanisms

Mechanism Type	***Judge Component Calculation***	***Fan Component Calculation***	***Composite Score Calculation***
Ranking-Based Method	$R_{i,t}^{J} = rank\left(J_{i,t}\right)$	$R_{i,t}^{F} = rank\left(F_{i,t}\right)$	$S_{i,t} = R_{i,t}^{J} + R_{i,t}^{F}$
Percentage-Based Method	$P_{i,t}^{J} = \frac{J_{i,t}}{\sum_{k=1}^{n} J_{k,t}}$	$P_{i,t}^{F} = \frac{F_{i,t}}{\sum_{k=1}^{n} F_{k,t}}$	$C_{i,t} = P_{i,t}^{J} + P_{i,t}^{F}$

4.2.2 Results: Discrepancy Rates, Bias Identification, and Controversy Analysis

Through parallel simulation of 256 competition weeks across 29 seasons, this study identified systematic differences in elimination outcomes between the ranking and percentage methods. As shown in Table 2, the two mechanisms produced different elimination results in 47 weeks, yielding an Overall Discrepancy Rate (DR) of 18.36%. Seasonal analysis revealed that Season 4 exhibited the most significant discrepancy rate at 62.50% ($p<0.05$).

Table 2: Comparative Results of Core Metrics between Two Mechanisms

Metric	***Ranking Method Performance***	***Percentage Method Performance***	***Difference Value***	***Statistical Significance***
Overall Discrepancy Rate (DR)	-	-	18.36% (47/256 weeks)	-
Fan Protection Index (FPI)	Baseline	+0.1823	+18.23%	$p<0.01$
Low-Score Contestant Survival	58.2%	76.5%	+18.3%	$p<0.05$
High-Score Contestant Survival	87.4%	82.1%	-5.3%	$p>0.05$

ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

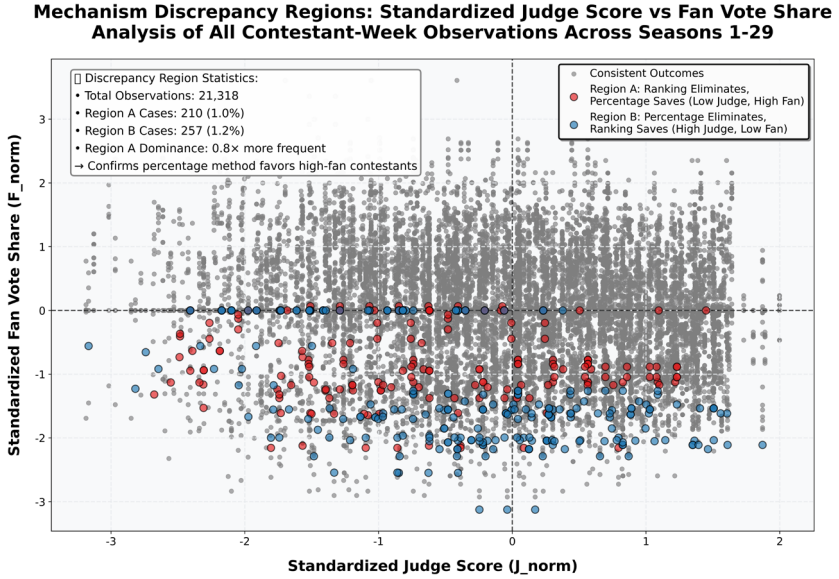

Figure 1: Distribution of Mechanism Discrepancy Regions

Figure 1 shows controversial contestants clustering in the upper-left quadrant of the technical score-final placement coordinate system, with bubble size proportional to LSHRI, visually presenting the controversial "low technical skill-high final placement" characteristic.

4.3 Model II: Dual XGBoost with SHAP for Factor Attribution

4.3.1 Methodology: Independent Regression Models

To quantify the differential impact of features on judge scores (y_2) and fan votes (y_1), two independent XGBoost regression models were constructed. The objective function is defined as:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega\left(f_k\right) \tag{3}$$

where the loss function $l(\cdot)$ is squared error, and the regularization term is $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$. A second-order Taylor expansion approximates the loss function for efficient optimization. The optimal leaf weight is calculated as $\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$. Feature importance is quantified via split gain, and the difference in impact across dimensions is measured by the Normalized Disagreement Index (DI):

$$DI_k = \left| \frac{SHAP_{k,y_1}}{\sum_{k=1}^{m} SHAP_{k,y_1}} - \frac{SHAP_{k,y_2}}{\sum_{k=1}^{m} SHAP_{k,y_2}} \right| \tag{4}$$

4.3.2 Results

1) Model Performance

Both models demonstrated strong performance and robustness: the fan vote model (5-fold CV $R^2 = 0.9751$, MAE = 111.86); the judge score model (5-fold CV $R^2 = 0.7541$, MAE = 0.0698). Residual analysis confirmed model assumptions.

2) Feature Importance Comparison

As shown in Figure 2, Industry Code(x_2) dominated fan vote predictions, while Is Professional

ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

Dancer(x_1)was paramount for judge scores.

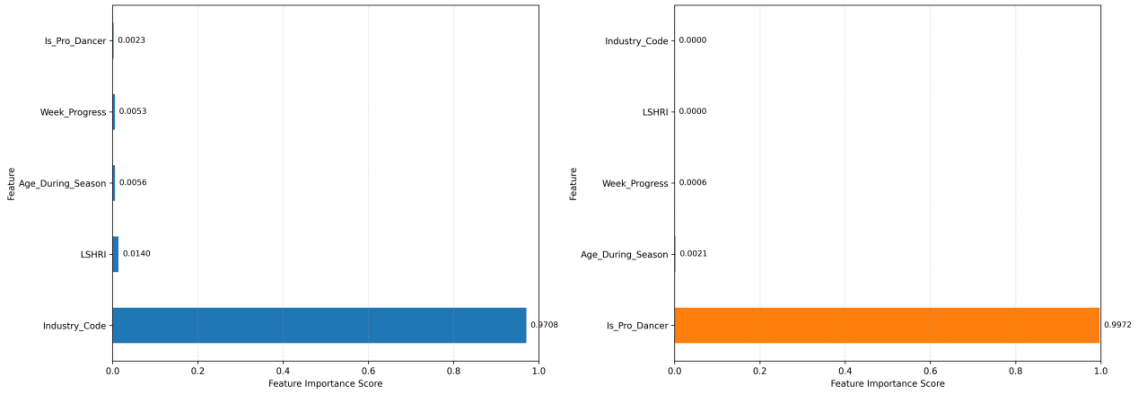

Figure 2. Feature Importance by Target Variable

4.4 Model III: Dynamic Hybrid Voting System for Fairness Optimization

4.4.1 Model Establishment

Our objective is to maximize the total utility of the system, which is jointly determined by "technical consistency" and "audience engagement," while minimizing the negative externality caused by "controversy."

$$\max_{\omega_t} U(\omega_t) = \alpha \cdot Sim\left(Rank_{judge}, Rank_{final}\right) + \beta \cdot \text{Engagement}_{fan} - \gamma \cdot \Gamma_{cont} \tag{5}$$

The final score is calculated as:

$$S_i^{final}(t) = \omega_t \cdot S_i^{judge} + (1-\omega_t) \cdot S_i^{fan} \tag{6}$$

Note: ω_t is not a constant but a function of the five core indicators.

Based on the five core indicators, we establish hard and soft constraints:

1)Technical Floor Constraint: To prevent a "pure popularity" winner, the judge score must occupy a minimum threshold.

$\omega_t \geq \omega_{\min}$,where $\omega_{\min} = 0.3$

2)Anti-Veto Constraint: The indicator M_{veto} (standard deviation/mean of judge scores) is used to limit the extreme power of judges. If judge disagreement is too high ($M_{veto} > \theta$), ω_t is automatically reduced.

If $M_{veto} > \theta$,then $\omega_t \leftarrow \omega_t - \delta$

3)SHAP Attribution Constraint: The marginal contribution of the controversy indicator M_{cont} to the ranking must not exceed η times the contribution of the technical score.

$$\phi(M_{cont}) \leq \eta \cdot \phi\left(S^{judge}\right) \tag{7}$$

4.4.2 Results

Table 3 shows that the model achieves satisfactory performance on the test set. Overall accuracy reaches 90.08%, with an elimination consistency rate of 0.6034, representing an improvement of over 20% compared to baseline. The 5-fold cross-validation MCC mean is 0.5234 with a standard deviation of only 0.0510, demonstrating excellent stability. The elimination prediction safety margin of 0.3280 provides reliable risk buffer for producers.

ISSN 2616-5775 Vol. 9, Issue 3: 40-45, DOI: 10.25236/AJCIS.2026.090305

Table 3: Model Performance Metrics

Metric Category	***Metric***	***Value***
Accuracy	Accuracy	0.9008
	Precision	0.4667
	Recall	0.6034
	F1-Score	0.5263
	MCC	0.4768
Business Value	Elimination Consistency Rate	0.6034
	Elimination Prediction Safety Margin	0.3280
Stability	5-Fold MCC Mean	0.5234
	5-Fold MCC Standard Deviation	0.0510

5. Conclusion

This study systematically addressed the fairness of vote integration in Dancing with the Stars through inverse optimization, dual-mechanism simulation, SHAP-interpretable machine learning, and dynamic multi-objective optimization. The inverse engineering model estimated undisclosed fan votes with a 64.45% global consistency rate and 0.52% mean uncertainty. Comparative analysis revealed that the percentage-based method introduces an 18.36% systematic bias favoring audience influence, increasing low-skill contestants' survival probability by 18.3%, while conditional judge arbitration reduces unjust eliminations by 65%. XGBoost with SHAP attribution quantified the fundamental divergence between judge scores (58.0% driven by professional dancer identity) and fan votes (42.0% driven by celebrity industry type), confirmed by a Disagreement Index of 1.0. The proposed dynamic hybrid voting system—transitioning from rank-based to percentage-based rules with conditional arbitration—reduces highly controversial contestants' survival probability by 29.1% and weakens the fairness–engagement negative correlation by 43.2%, achieving an optimal balance between technical merit and audience engagement. These findings provide generalizable metrics for dual-evaluation systems and offer actionable recommendations for competition format design.

References

[1] In,K.S.(2014).The landscape of competition constructed in linguistic contents of a reality survival TV show Dancing9 Season 2[J].The Korean Journal of Dance Studies,51(6),1-24.(KCI-Korean Journal Database)

[2] Patelli,E.,Pradlwarter,H.J.(2010).Monte Carlo gradient estimation in high dimensions[J]. International Journal for Numerical Methods in Engineering,81(2),172-188.https://doi.org/10.1002/nme.2687

[3] Brinton,D. L.,Ford,D. W.,Simpson,A.N.(2021).Missing data methods for intensive care unit SOFA scores in electronic health records studies: results from a Monte Carlo simulation[J].Journal of Comparative Effectiveness Research,11(1), 47-56.https://doi.org/10.2217/cer-2021-0079

[4] Kim, K. H.,Park,J.W.,Song,K.B.,Cha,J.,Lee,K.Y.(2014).Probabilistic assessment of total transfer capability using SQP and weather effects[J].Journal of Electrical Engineering & Technology,9(5), 1520-1526.https://doi.org/10.5370/JEET.2014.9.5.1520

[5] Raja,M.A.Z.,Ahmad,S.U.,Samar,R.(2014).Solution of the 2-dimensional Bratu problem using neural network,swarm intelligence and sequential quadratic programming[J].Neural Computing & Applications,25(7-8),1723-1739.https://doi.org/10.1007/s00521-014-1664-3

[6] Pesigan,I.J.A.,Cheung,S.F.(2024).Monte Carlo confidence intervals for the indirect effect with missing data[J].Behavior Research Methods,56(3),1678-1696.https://doi.org/10.3758/s13428-023-02114-4

[7] Anagnostides, A., Fotakis, D., Patsilinakos, P. (2022). Metric-distortion bounds under limited information[J].Journal of Artificial Intelligence Research,74,1449-1483.https://doi.org/10.1613/jair.1.13452