Big Data and Information Analytics

https://www.aimspress.com/journal/BDIA

BDIA, 10: 350–373.
DOI:10.3934/bdia.2026018
Received: 22 December 2025
Revised: 26 August 2026
Accepted: 28 August 2026
Published: 06 November 2026

Research article

Crowdfunding success in game development: an explainable machine learning study of Kickstarter projects

Han Yu[1,*], Felix Junior Appiah Kubi[1] and Hong Huang[2]

[1] Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO, USA

[2] School of Information, University of South Florida, Tampa, FL, USA

* **Correspondence:** Email: han.yu@unco.edu.

Abstract: Crowdfunding platforms function as virtual communities in which creators, backers, and platform signals jointly shape the project visibility, credibility, and funding outcomes. This study examines the prediction and interpretation of Kickstarter game-campaign outcomes. First, it provides descriptive context using a combined record of 10,935 Kickstarter game campaigns launched between January 1, 2018, and April 12, 2026. Comparable records for both tabletop games and video games are only available from 2022 onward. Accordingly, the primary predictive and interpretive analyses used an outcome-complete contemporary analytical sample of 9627 campaigns launched between January 1, 2022, and April 12, 2026. Using five pre-launch predictors, eight machine-learning classifiers were evaluated using nested cross-validation, and their temporal generalizability was assessed on campaigns launched in 2025. The soft-voting ensemble achieved the highest nested cross-validation receiver operating characteristic (ROC) area under the curve (AUC) (AUC = 0.896; accuracy = 90.08%), while XGBoost produced nearly identical results (AUC = 0.895; accuracy = 90.02%) and was used for model interpretation. A Shapley additive explanations (SHAP) analysis characterized the relative contributions of the funding goal, exact category, preparation time, campaign duration, and blurb length to the predicted campaign success. Within the contemporary analytical sample, the observed success rate was substantially higher for Tabletop Games than for Video Games. Overall, the findings show that a parsimonious set of pre-launch predictors retains substantial out-of-sample and out-of-time predictive discrimination for Kickstarter game-campaign outcomes while supporting a transparent interpretation of model behavior.

Keywords: crowdfunding; Kickstarter; game crowdfunding; explainable machine learning; temporal validation; signaling theory

1. Introduction

Crowdfunding has become an important financing mechanism for creative and technology ventures, allowing entrepreneurs to directly raise capital from geographically dispersed online audiences. In game development, platforms such as Kickstarter have played a particularly visible role by connecting independent creators with backers who may simultaneously serve as funders, early adopters, and community advocates. Therefore, game crowdfunding is shaped not only by financing needs, but also by trust, signaling, and collective attention within a platform-mediated community.

These dynamics make project evaluation especially difficult under conditions of substantial information asymmetry. Before committing funds, backers cannot directly observe the underlying quality of a proposed game or fully assess a creator's ability to complete and deliver it. Instead, they must rely on observable pre-launch and platform-mediated signals when forming judgments about project credibility and feasibility. Therefore, signals related to the project scope, category, preparation, campaign design, and presentation may be informative about the campaign credibility, feasibility, and eventual funding outcomes [1].

Prior crowdfunding research has identified a broad range of factors associated with the campaign success, including project-quality signals, personal networks, geography, social capital, campaign presentation, and funding goals [2–4]. However, much of the empirical literature either relies on data from the platform's earlier years, does not distinguish between tabletop games and video games, or emphasizes the predictive accuracy without fully articulating the validation framework and explainability strategy. More recent studies have begun to incorporate richer textual representations, interpretable machine-learning methods, and greater attention to contextual heterogeneity [5, 6], yet relatively little work combines rigorous nested validation, out-of-time evaluation, and post hoc model explanation within a recent game-crowdfunding cohort.

This study addresses these gaps through a two-component empirical framework. The combined descriptive records of available campaigns launched from 2018 through April 2026 provide a contextual background and broad descriptive information, whereas the outcome-complete contemporary analytical component covering January 2022 through April 2026 supports predictive modeling, model evaluation, temporal validation, and explainability analyses. The objective is not merely to predict campaign success, but also to determine how observable pre-launch signals contribute to model predictions in the contemporary game-crowdfunding environment.

The study makes three contributions. First, it situates the contemporary analysis within the combined available Kickstarter game-crowdfunding record from 2018 through April 2026, complementing earlier research based largely on data from the platform's earlier years. Second, it evaluates whether a parsimonious set of readily observable pre-launch predictors can support accurate predictions across distinct game categories. The contribution does not lie in treating these individual characteristics as previously unknown predictors. Rather, the primary analysis deliberately restricts prediction to a structured subset of characteristics available at or before campaign launch and asks how much predictive information can be retained under this constrained pre-launch representation. Richer pre-launch representations and information that dynamically emerges after launch represent broader prediction settings beyond the primary specification.

Third, it integrates comparative model benchmarking, nested validation, out-of-time evaluation, cross-model permutation importance, and probability-scale SHAP interpretation within a unified contemporary

game-crowdfunding framework. These contributions are organized around three questions: how accurately can campaign outcomes be predicted using a parsimonious pre-launch predictor set; how well does predictive performance generalize to campaigns launched in a later calendar year; and which observable signals contribute most strongly to model predictions?

2. Related work

2.1. Crowdfunding success, signaling, and virtual communities

A substantial body of crowdfunding research treats information asymmetry as a central feature of campaign decision making, because prospective backers must evaluate projects under uncertainty and with limited verifiable information. Evidence from early studies of Kickstarter indicates that project-quality signals, personal networks, and geography are associated with funding outcomes [2], while multidimensional social capital also shapes backer decisions [3]. In the specific context of video game crowdfunding, Cha [4] found that human capital, geography, media choice, and media-use intensity were associated with funding outcomes.

These findings are consistent with the signaling theory [1], which provides a framework for understanding how creators communicate otherwise difficult-to-observe aspects of project quality, feasibility, and credibility. Within crowdfunding, observable campaign and creator characteristics may function as signals that help prospective backers evaluate uncertain projects. However, crowdfunding platforms are not solely transactional marketplaces; they also operate as virtual communities in which such signals are interpreted within a social environment shaped by interaction, visibility, endorsement, and information diffusion. Therefore, a backer's judgment may depend not only on the presence of individual signals, but also on how those signals circulate and acquire meaning among platform participants. These dynamics may be especially relevant in game crowdfunding, where the experiential nature of games and the interactive affordances of Kickstarter increase the salience of creator communication and community engagement [7].

2.2. Machine learning and explainable prediction in crowdfunding

Crowdfunding prediction research has evolved from regression-based and early dynamic-forecasting approaches to flexible machine-learning models and richer textual representations [5, 8]. A systematic review of 122 reward-based crowdfunding studies identified four broad classes of success factors—campaign characteristics, campaign presentation, entrepreneur characteristics, and entrepreneur networks—while documenting substantial inconsistencies and context dependence across findings [6]. These patterns indicate that predictive relationships may vary across domains, feature representations, and analytical designs.

Recent research further suggests that crowdfunding prediction is moving beyond conventional tabular modeling toward richer representational frameworks. Transformer-based and embedding-based approaches have shown promise for predicting campaign outcomes from textual signals [5]. Likewise, broader reviews of crowdfunding for innovation highlight multimodal content, narrative quality, and visual presentation as potentially important correlates of crowdfunding outcomes [9]. These developments motivate greater attention not only to the predictive performance, but also to the transparent attribution of feature contributions in settings where backers must evaluate projects

under uncertainty.

Despite this progress, relatively few studies provide a contemporary and interpretable comparison of machine-learning classifiers specifically for game crowdfunding while also separating the model selection from the performance estimation. Reusing the same cross-validation procedure for both purposes can produce optimistically biased performance estimates [10]. Additionally, variables realized during or after a campaign may introduce temporal or look-ahead leakage, while permutation-importance values computed under different scoring metrics are not directly comparable across models. Therefore, the present study combines nested cross-validation, out-of-time temporal validation, cross-model permutation importance measured on a common receiver operating characteristic-area under the curve (ROC-AUC) scale, and probability-scale Shapley additive explanations (SHAP) interpretation [11].

3. Materials and methods

3.1. Data collection and temporal scope

The study uses two temporally nonoverlapping data components with distinct analytical roles. The historical descriptive component provides contextual background and broad descriptive information, whereas the contemporary analytical component supports the study's quantitative analyses. The historical descriptive component was obtained from a previously constructed and validated campaign-level file and includes launches from January 1, 2018, through December 31, 2021. The contemporary analytical component was obtained from the June 11, 2026 Web Robots Kickstarter archive, timestamped at 03 : 20 : 11 UTC, and includes campaigns launched on or after January 1, 2022 [12]. No project ID appeared in both data sources.

Within the contemporary analytical component, records were only retained when the exact category slug was `games/tabletop games` or `games/video games`. Duplicate records were resolved by project ID. Outcomes were restricted to the terminal states *successful* and *failed*; campaigns classified as live, canceled, suspended, or in another nonterminal state were excluded. The historical descriptive component contains validated pre-2022 video games records but no validated pre-2022 tabletop games records. This difference reflects the source coverage and should not be interpreted as evidence that Tabletop Games campaigns were absent from Kickstarter before 2022. Accordingly, direct comparisons between Tabletop Games and Video Games were restricted to the contemporary analytical component.

Kickstarter permits project durations of up to 60 days [13]. To avoid incomplete outcomes while conservatively accommodating minor timestamp discrepancies, the data pipeline applied a 60.1-day follow-up buffer. Relative to the June 11, 2026 snapshot, this produced an inclusive launch-date endpoint of April 12, 2026 at 00 : 56 : 11 UTC. The combined study record contains 10,935 campaigns, consisting of 1308 campaigns in the historical descriptive component and 9627 outcome-complete campaigns in the contemporary analytical component. The full record was used for contextual and descriptive reporting, whereas all formal statistical, predictive, temporal-validation, and explainability analyses used the contemporary analytical component.

3.2. Feature engineering and variable selection

The funding goal in U.S. dollars was computed by multiplying the native-currency goal by the archived static USD exchange rate. The preparation time was defined as the interval between project

creation and launch, whereas the campaign duration was defined as the interval between launch and deadline. The funding goals were required to be positive, and all time intervals were required to be internally consistent.

Across the two source periods, 37 records in the historical descriptive component and 68 records in the contemporary analytical component had campaign durations between 60 and 60.1 days, consistent with a one-hour timestamp shift, and were normalized to 60 days. One historical record with a duration of 64.995 days and two contemporary records with durations of 66.958 days were excluded as material exceptions. Therefore, the maximum retained campaign duration was 60 days. Supplementary Tables S1 and S2 document these decisions.

The binary outcome was coded as successful = 1 and failed = 0. Table 1 defines each retained field and its role in the study.

Table 1. Variables and their roles in the analysis.

Variable	Operationalization	Analysis role
Funding goal	Natural log of one plus the funding goal in USD	Primary pre-launch predictor representing project scope and financial feasibility.
Campaign duration	Number of days from launch to deadline	Primary pre-launch predictor; values between 60 and 60.1 days were normalized to 60 days.
Preparation time	Natural log of one plus the number of days from project creation to launch	Primary pre-launch predictor representing the pre-launch preparation interval.
Exact category	Tabletop Games = 1; Video Games = 0	Primary pre-launch predictor based only on exact Kickstarter category slugs.
Blurb length	Character count of the short project blurb	Primary pre-launch predictor representing a parsimonious measure of project presentation.
Staff Pick	Kickstarter platform-endorsement indicator	Excluded from the primary pre-launch model and included only in a separate Staff Pick sensitivity analysis.
Backer count	Number of observed backers	Used for descriptive reporting only; excluded from all predictive models because it is realized during the campaign.
Creator experience	No validated creator-history field was available in the contemporary source	Excluded from models based on the contemporary analytical component and from all variable-importance analyses.
Analysis period	Historical descriptive component: 2018–2021; contemporary analytical component: January 2022–April 2026	The combined 2018–April 2026 record was used for contextual and descriptive reporting, whereas formal statistical, predictive, temporal-validation, and explainability analyses used only the contemporary analytical component.

The primary pre-launch model uses five predictors available at or before campaign launch: $x_i = [\log(1+\text{goal}_{i,\text{USD}}),\ \text{duration}_i,\ \log(1+\text{preparation days}_i),\ \mathbb{1}\{\text{Tabletop}_i\},\ \text{blurb length}_i]^\top$. The launch year was reserved for temporal splitting and was not used as a predictor. The backer count is realized during the campaign and was therefore only used for descriptive reporting. Because the archived Staff Pick field

could not be established as fixed at launch, it was excluded from the primary pre-launch specification and was only included in a clearly labeled staff pick sensitivity analysis. The available source schema did not contain a validated creator-history field; accordingly, no creator-experience variable was imputed, included in the predictive models, or reported in the variable-importance analyses. Therefore, the primary specification represents a deliberately restricted, structured, and leakage-controlled subset of the information available at or before campaign launch. This restriction defines the informational scope of the prediction task rather than implying that omitted characteristics are unimportant. Richer characteristics that may also be available at or before launch, including images, videos, full campaign text, reward structure, team composition, and verified creator history, could expand the pre-launch representation. In contrast, campaign updates, comments, backer activity, and evolving engagement signals become available during the campaign and therefore correspond to a different, dynamically updated prediction setting. Consequently, the predictive performance should be interpreted relative both to the information available at the stated prediction time and to the representation of that information used by the model.

3.3. Predictive models

Let $Y_i \in \{0, 1\}$ denote the binary campaign outcome, with $p_i = \Pr(Y_i = 1 \mid \mathbf{x}_i)$. Eight model classes were selected to span complementary modeling paradigms, including linear, regularized, tree-based, kernel-based, neural-network, gradient-boosting, and ensemble approaches. This selection permits comparison across substantially different functional forms and levels of model complexity under a common evaluation protocol. Standard model formulations follow Hastie et al. [14], unless otherwise noted.

L2-regularized and lasso-logistic regression (M1–M2). The logistic-regression model is as follows:

$$\operatorname{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}. \tag{1}$$

For M1, an unpenalized intercept and ℓ_2-penalized slope coefficients were estimated. The lasso-logistic variant estimates an unpenalized intercept and ℓ_1-penalized slope coefficients as follows [15]:

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \arg\min_{\beta_0, \boldsymbol{\beta}} \left[-\frac{1}{N} \sum_{i=1}^{N} \{ y_i \log p_i + (1 - y_i) \log(1 - p_i) \} + \lambda_{\text{lasso}} \|\boldsymbol{\beta}\|_1 \right], \tag{2}$$

where $p_i = \{1 + \exp[-(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta})]\}^{-1}$.

Decision tree (M3). The classification tree recursively partitions the predictor space to reduce class impurity. The maximum tree depth and minimum terminal-node size were selected through inner-fold tuning.

Support vector machines (M4–M5). Linear and radial-basis-function (RBF) support vector machines were evaluated [16]. For the linear support vector machine (SVM), the primal soft-margin objective is as follows:

$$\min_{w, b, \xi} \ \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{N} \xi_i, \qquad y_i^*(w^\top x_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \tag{3}$$

where $y_i^* \in \{-1, 1\}$. The radial basis function (RBF)-SVM replaces the linear inner product with the following:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left\{-\gamma_{\text{RBF}}\|\mathbf{x} - \mathbf{x}'\|_2^2\right\}. \tag{4}$$

Probability estimates required for probability-based evaluation and soft voting were obtained through probability calibration conducted within the corresponding training data.

Multi-layer perceptron (M6). To include a neural-network benchmark within the structured tabular comparison, the multi-layer perceptron (MLP) minimizes the binary cross-entropy as follows:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right]. \tag{5}$$

The hidden-layer architecture and ℓ_2 regularization strength were selected within the inner folds.

XGBoost (M7). For binary logistic XGBoost, additive trees determine the raw prediction score η_i, with $p_i = \{1 + \exp(-\eta_i)\}^{-1}$. At the boosting iteration t, the regularized objective is as follows [17]:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{N} \ell_{\text{logistic}}\left(y_i, \hat{\eta}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t), \tag{6}$$

where

$$\Omega(f_t) = \gamma_{\text{XGB}} T + \frac{\lambda_{\text{XGB}}}{2}\|\mathbf{w}\|_2^2. \tag{7}$$

Here, T is the number of terminal leaves and $\mathbf{w}$ is the vector of leaf weights.

Soft-voting ensemble (M8). The soft-voting ensemble combines predicted probabilities from the RBF-SVM, MLP, and XGBoost models:

$$\hat{p}_{i,\text{ens}} = \frac{\sum_{m\in\{\text{RBF,MLP,XGB}\}} w_m \hat{p}_{im}}{\sum_{m\in\{\text{RBF,MLP,XGB}\}} w_m}. \tag{8}$$

Within each outer-training fold, the weighting scheme was selected through inner-fold validation from two candidate configurations: equal weights, (1, 1, 1), and XGBoost-emphasized weights, (1, 1, 2).

3.4. Evaluation protocol

The primary evaluation framework used stratified five-fold outer cross-validation [10]. Within each outer-training fold, stratified three-fold inner cross-validation selected hyperparameters by ROC AUC. Then, the model with the selected hyperparameter configuration was refitted on the complete outer-training fold and used to generate predicted probabilities for the untouched outer-test fold. Pooling predictions across the five outer folds yielded one out-of-fold predicted probability for every campaign.

The model performance was evaluated using multiple complementary measures. Discrimination was assessed using ROC AUC, reported with a 95% confidence interval, and average precision. At a fixed classification threshold of 0.5, the threshold-dependent performance was summarized using accuracy,

precision, recall (sensitivity), specificity, F1 score, and Cohen's kappa. The probability quality was assessed using the Brier score [18],

$$\text{Brier} = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2, \tag{9}$$

and binary log loss,

$$\text{LogLoss} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right]. \tag{10}$$

The Brier score and log loss were used to evaluate the quality of predicted binary-event probabilities. The root mean square error (RMSE) was not reported as a separate performance measure because, for binary outcomes and predicted probabilities, it is the square root of the Brier score and therefore provides no additional ranking information. Percentile 95% confidence intervals for nested-cross-validation ROC AUC were computed from 1000 bootstrap resamples of the pooled out-of-fold predictions.

Temporal generalizability was evaluated by training the models on campaigns launched from 2022 through 2024 ($N = 3605$) and testing them once on campaigns launched in 2025 ($N = 4506$). Hyperparameters were selected through a stratified three-fold grid search using only the temporal training sample. Campaigns launched in 2026 were retained for descriptive reporting but were excluded from the temporal test set because the 2026 observation period was incomplete.

3.5. Model interpretation and diagnostic analyses

Model interpretation was conducted at both the cross-model and model-specific levels. Permutation importance was used to compare the predictive relevance of the five pre-launch predictors across model classes, whereas SHAP was used to characterize the prediction behavior of the temporal XGBoost model. SHAP provides locally additive feature attributions grounded in Shapley-value principles [11]. The resulting explanations were interpreted as predictive associations rather than causal effects.

The permutation importance was evaluated on the common temporal test set. For the feature j,

$$I_j^{\text{perm}} = \text{AUC}(f, X, y) - \frac{1}{R}\sum_{r=1}^{R}\text{AUC}\left(f, X_{\pi_j^{(r)}}, y\right), \tag{11}$$

where $X_{\pi_j^{(r)}}$ denotes the test data after the values of the feature j were permuted in repetition r, and $R = 10$. Using the same temporal test set and scoring rule provided a common basis for a descriptive comparison of feature-importance rankings across model classes.

Probability-scale SHAP values were computed for the temporal XGBoost model using a model-agnostic permutation explainer [11]. For an explained observation,

$$\hat{p}(x) = \phi_0 + \sum_{j=1}^{J}\phi_j, \tag{12}$$

where ϕ_0 is the baseline predicted probability, and ϕ_j represents the contribution of the feature j to the predicted probability of campaign success. Thus, the SHAP values are expressed on the probability

scale rather than the log-odds or raw-margin scale. Global SHAP importance was summarized as

$$I_j^{\mathrm{SHAP}} = \frac{1}{n_{\mathrm{SHAP}}} \sum_{i=1}^{n_{\mathrm{SHAP}}} |\phi_{ij}|, \tag{13}$$

using 500 observations from the temporal test set and a 100-observation background sample. Partial-dependence and grouped predicted-probability plots were used as descriptive model diagnostics and were not interpreted as causal estimands.

4. Results

4.1. Descriptive summary of the combined descriptive record

The combined descriptive record contained 10,935 campaigns launched between January 1, 2018, and April 12, 2026. Of these, 8733 were successful and 2202 failed, corresponding to an overall observed success rate of 79.86%. These full-period summaries only provide descriptive context. Because category coverage before 2022 was inseparable from the source period, direct comparisons between tabletop games and video games were restricted to the contemporary analytical component comprising 9627 campaigns launched from January 1, 2022, through April 12, 2026. All model-based analyses reported below used this contemporary analytical component.

Table 2. Descriptive summary of the combined 2018–April 2026 record (N = 10,935).

Variable	Mean	SD	Median	IQR	Min	Max
Funding goal (USD)	10,836.13	91,365.72	984.21	5659.92	1.00	6,666,350.00
Campaign duration (days)	27.11	12.52	29.96	11.37	1.56	60.00
Preparation time (days)	78.95	192.16	22.03	63.06	0.01	4765.46
Staff Pick	0.086	0.281	0.000	0.000	0.000	1.000
Tabletop Games	0.722	0.448	1.000	1.000	0.000	1.000
Blurb length (characters)	99.94	30.97	108.00	50.00	3.00	135.00

Note: Means for staff pick and tabletop games are proportions.

Table 3. Descriptive statistics for the combined descriptive record (January 2018–April 12, 2026; N = 10,935).

Variable	Mean	SD	1	2	3	4
1. Funding goal	10,836.13	91,365.72	–			
2. Campaign duration	27.11	12.52	0.12 [0.10, 0.14]	–		
3. Preparation time	78.95	192.16	0.02 [0.01, 0.04]	0.08 [0.06, 0.10]	–	
4. Backer count	300.37	1,325.73	0.04 [0.02, 0.06]	0.02 [0.00, 0.04]	0.08 [0.06, 0.09]	–

Note: Bracketed values are 95% Fisher-z confidence intervals for the Pearson correlations.

Bivariate correlations among the funding goal, campaign duration, preparation time, and backer count were small in the combined descriptive record (Table 3). The backer count was included for descriptive purposes but was not used as a predictor in any machine-learning model.

4.2. Descriptive outcomes by launch period and game category

The historical descriptive component contained 1308 available Video Games campaigns launched from 2018 through 2021, of which 498 were successful, corresponding to an observed success rate of 38.07%. The contemporary analytical component contained 9627 campaigns launched from January 1, 2022, through April 12, 2026, of which 8235 were successful, corresponding to an observed success rate of 85.54%. Within the contemporary analytical component, the observed success rate was 93.40% for Tabletop Games and 49.62% for Video Games. Supplementary Table S3 presents the historical, contemporary, and combined descriptive summaries side by side.

Differences between the historical and contemporary periods should not be interpreted as a causal time trend. The historical source only contributes validated Video Games records, whereas the contemporary source contains comparable records for both exact game categories. Accordingly, the observed period difference reflects changes in the source coverage and category composition, in addition to any underlying temporal variation.

Table 4. Campaign outcomes by launch year and exact game category.

Year	Exact category	*N*	Success (%)	95% CI
2018	Video games	273	34.80	[29.39, 40.62]
2019	Video games	318	29.56	[24.81, 34.80]
2020	Video games	380	42.37	[37.50, 47.39]
2021	Video games	337	43.92	[38.72, 49.25]
2022	Tabletop games	83	87.95	[79.22, 93.32]
2022	Video games	17	17.65	[6.19, 41.03]
2023	Tabletop games	170	91.18	[85.95, 94.58]
2023	Video games	53	30.19	[19.52, 43.54]
2024	Tabletop games	2,706	93.16	[92.15, 94.05]
2024	Video games	576	44.27	[40.27, 48.35]
2025	Tabletop games	3,667	92.91	[92.03, 93.70]
2025	Video games	839	53.16	[49.78, 56.51]
2026	Tabletop games	1,272	95.99	[94.77, 96.94]
2026	Video games	244	56.56	[50.28, 62.63]

Note: Confidence intervals are Wilson 95% intervals. The 2026 rows include campaigns launched through April 12, 2026 and therefore represent a partial calendar year.

Within each year of the contemporary analytical period, the observed success rate was higher for tabletop games than for video games (Table 4). However, the precision of the annual estimates substantially varied with sample size, as reflected in the wider confidence intervals for the relatively small 2022 and 2023 samples. The 2026 rows include launches only through April 12 and should not be interpreted as full-year counts or rates. Differences in the numbers of campaigns observed across years reflect the coverage of the available data sources and should not be interpreted as estimates of platform-wide campaign incidence.

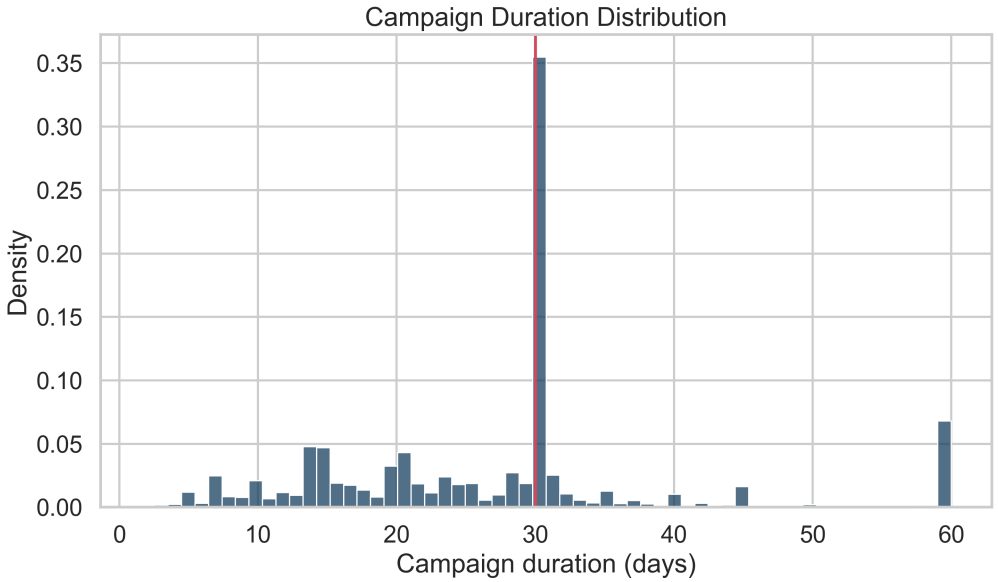


Figure 1. Distribution of retained campaign durations. The observed concentration near 30 days is consistent with Kickstarter's recommendation that campaigns run for 30 days or less [13]. The maximum retained duration was 60 days.

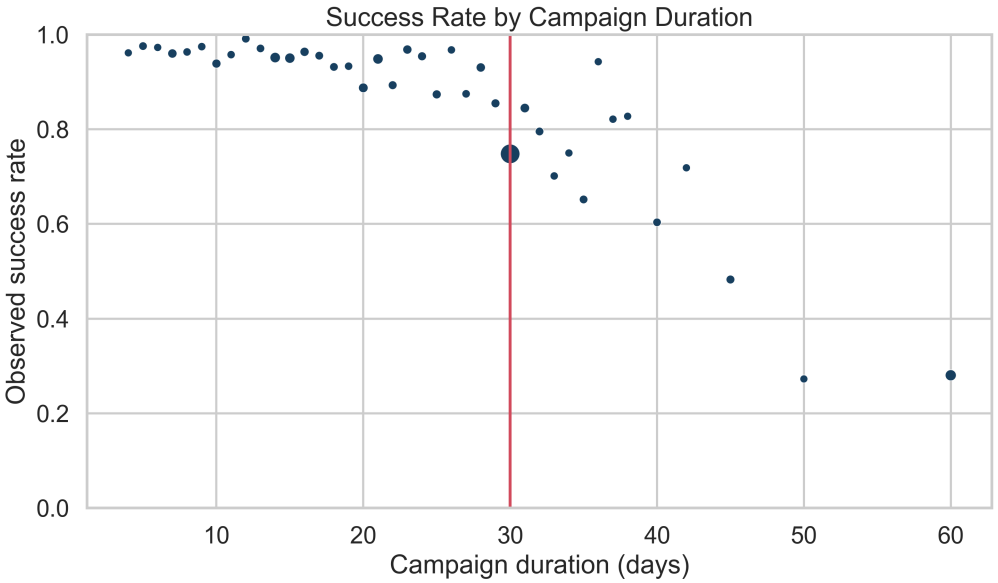


Figure 2. Observed success rate by campaign-duration bin. The plotted relationship is unadjusted and descriptive.

4.3. Comparative performance of predictive models

The contemporary analytical component used for predictive modeling and model evaluation contained 9627 campaigns, including 8235 successful campaigns and 1392 failed campaigns. The

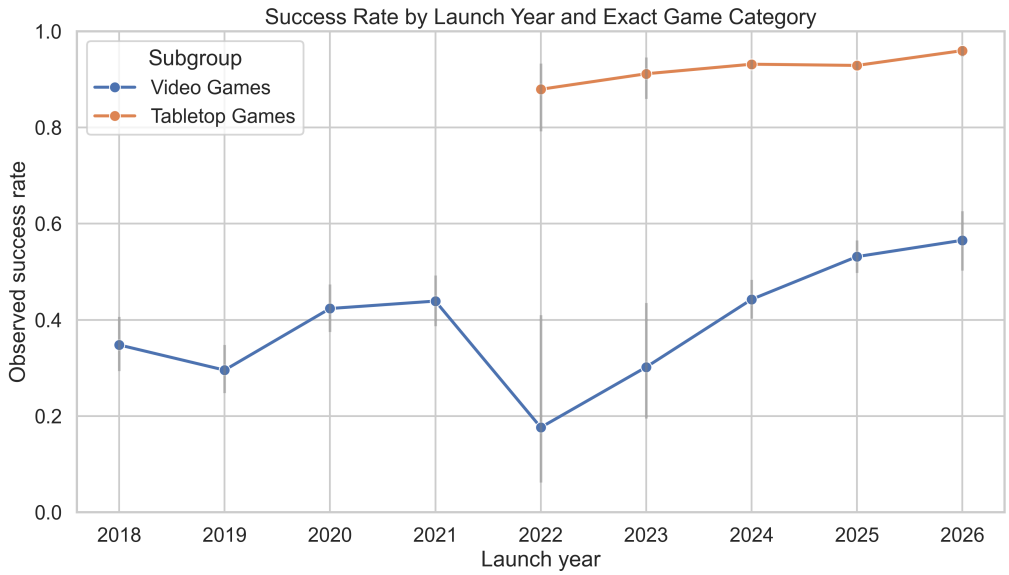


Figure 3. Observed success rates by launch year and exact game category. Error bars represent 95% Wilson confidence intervals [19]. The 2026 observation window ends on April 12 and therefore represents a partial calendar year.

observed success prevalence was 85.54%. Given this class imbalance, specificity and Cohen's kappa provide important complements to the accuracy and recall. The combination of high recall and moderate specificity indicates that the leading models identified successful campaigns more readily than failed campaigns.

Table 5 reports the complete set of performance metrics computed from the pooled out-of-fold predictions obtained during nested cross-validation. The soft-voting ensemble achieved the highest nested-cross-validation ROC AUC (0.896, 95% CI [0.887, 0.905]) and accuracy (90.08%; Table 5). XGBoost yielded nearly identical discrimination (ROC AUC = 0.895, 95% CI [0.886, 0.904]) and accuracy (90.02%). The substantial overlap between the bootstrap confidence intervals does not support interpreting the small AUC difference as evidence of a decisive performance advantage. At the prespecified classification threshold of 0.5, XGBoost had the highest Cohen's kappa (0.531) and a higher specificity than the ensemble (0.486 versus 0.456). Additionally, it produced the lowest log loss (0.251) and a slightly lower Brier score than the ensemble (0.0736 versus 0.0738). These results supported the selection of XGBoost as the model-specific interpretive reference, while recognizing the ensemble's marginally higher nested-CV AUC.

Models trained on campaigns launched from 2022 through 2024 maintained broadly similar discrimination when evaluated on the 2025 temporal holdout (Table 6). The soft-voting ensemble achieved the highest temporal ROC AUC (0.880), closely followed by XGBoost and the neural network (both 0.879). The neural network achieved the highest temporal accuracy (89.59%) and the lowest Brier score (0.0785), although performance differences among the leading models were small.

Table 5. Nested-cross-validation performance in the contemporary analytical component (N = 9627).

Panel A. Threshold-dependent classification metrics

Model	Accuracy (%)	Precision	Recall	Specificity	F1	Kappa
M8: Soft Voting Ensemble	90.08	0.914	0.976	0.456	0.944	0.519
M7: XGBoost	90.02	0.918	0.970	0.486	0.943	0.531
M6: Neural Net	89.98	0.917	0.971	0.481	0.943	0.527
M1: L2-Logistic	89.77	0.915	0.970	0.468	0.942	0.514
M2: Lasso-Logistic	89.69	0.912	0.974	0.441	0.942	0.499
M4: SVM (Linear)	89.79	0.913	0.973	0.453	0.942	0.508
M3: Decision Tree	89.63	0.914	0.970	0.460	0.941	0.506
M5: SVM (RBF)	89.99	0.908	0.982	0.415	0.944	0.495

Panel B. Discrimination and probability-quality metrics

Model	AUC [95% CI]	PR AUC	Brier	Log loss
M8: Soft Voting Ensemble	0.896 [0.887, 0.905]	0.977	0.074	0.254
M7: XGBoost	0.895 [0.886, 0.904]	0.976	0.074	0.251
M6: Neural Net	0.891 [0.882, 0.901]	0.975	0.075	0.255
M1: L2-Logistic	0.889 [0.879, 0.899]	0.974	0.076	0.258
M2: Lasso-Logistic	0.887 [0.878, 0.897]	0.973	0.076	0.262
M4: SVM (Linear)	0.883 [0.873, 0.893]	0.971	0.076	0.261
M3: Decision Tree	0.878 [0.868, 0.889]	0.970	0.079	0.270
M5: SVM (RBF)	0.801 [0.784, 0.817]	0.928	0.081	0.292

Note. Estimates are based on pooled out-of-fold predictions from nested cross-validation with five outer folds and three inner folds. ROC AUC confidence intervals are percentile 95% bootstrap intervals based on 1000 resamples. Threshold-dependent metrics use a cutoff of 0.5.

Table 6. Temporal holdout performance: training on 2022–2024 and testing on 2025.

Model	Acc(%)	Prec	Rec	Spec	F1	AUC	PR-AUC	Brier	LogLoss	Kappa
M8: Soft Voting Ensemble	89.48	0.914	0.969	0.459	0.940	0.880	0.971	0.081	0.285	0.502
M7: XGBoost	89.08	0.916	0.960	0.484	0.938	0.879	0.972	0.079	0.268	0.502
M6: Neural Net	89.59	0.913	0.971	0.453	0.941	0.879	0.972	0.079	0.268	0.503
M1: L2-Logistic	88.99	0.915	0.961	0.472	0.937	0.878	0.971	0.080	0.270	0.493
M2: Lasso-Logistic	89.01	0.915	0.961	0.475	0.937	0.878	0.971	0.080	0.271	0.495
M4: SVM (Linear)	89.28	0.914	0.965	0.464	0.939	0.877	0.970	0.079	0.269	0.498
M3: Decision Tree	87.84	0.916	0.944	0.492	0.930	0.858	0.962	0.086	0.294	0.470
M5: SVM (RBF)	89.17	0.906	0.974	0.406	0.939	0.826	0.944	0.086	0.301	0.465

Note. Models were trained on 3605 campaigns launched during 2022–2024 and evaluated on 4506 campaigns launched in 2025. Hyperparameters were selected only using the temporal training sample. Threshold-dependent metrics use a cutoff of 0.5. Abbreviations: Acc = Accuracy, Prec = Precision, Rec = Recall, Spec = Specificity.

4.4. Model predictor importance and interpretation

Using a common temporal-holdout ROC-AUC permutation-importance metric, all eight model classes ranked the funding goal as the most important predictor (Table 7). For XGBoost, the remaining predictors were ranked as the exact game category, preparation time, campaign duration, and blurb length. This ordering was identical to the global probability-scale SHAP ranking for XGBoost, although permutation importance and SHAP quantify distinct aspects of predictive relevance.

Table 7. Predictor rankings based on temporal-holdout ROC-AUC permutation importance.

Model	1st	2nd	3rd	4th	5th
M1: L2-Logistic	FG(log USD)	PT(log days)	EC(Tabletop)	CD(days)	BL
M2: Lasso-Logistic	FG(log USD)	PT(log days)	EC(Tabletop)	CD(days)	BL
M3: Decision Tree	FG(log USD)	EC(Tabletop)	CD(days)	PT(log days)	BL
M4: SVM (Linear)	FG(log USD)	PT(log days)	EC(Tabletop)	CD(days)	BL
M5: SVM (RBF)	FG(log USD)	EC(Tabletop)	PT(log days)	CD(days)	BL
M6: Neural Net	FG(log USD)	EC(Tabletop)	PT(log days)	CD(days)	BL
M7: XGBoost	FG(log USD)	EC(Tabletop)	PT(log days)	CD(days)	BL
M8: Soft Voting Ensemble	FG(log USD)	EC(Tabletop)	CD(days)	PT(log days)	BL

Note. Permutation importance was defined as the decrease in temporal-holdout ROC AUC after permuting each predictor, averaged over 10 repetitions. Abbreviations: FG = Funding Goal; PT = Preparation Time; EC = Exact Category; CD = Campaign Duration; BL = Blurb Length.

Table 8. Global predictor importance for the temporal XGBoost model based on probability-scale SHAP values.

Rank	Predictor	Mean $\lvert\phi_j\rvert$
1	Funding Goal (log USD)	0.1005
2	Exact Category (Tabletop)	0.0578
3	Preparation Time (log days)	0.0522
4	Campaign Duration (days)	0.0377
5	Blurb Length	0.0159

Note. ϕ_j denotes the contribution of predictor j to the predicted probability of campaign success.

The funding goal had the largest mean absolute SHAP contribution (0.1005 on the probability scale), followed by the exact game category (0.0578), preparation time (0.0522), campaign duration (0.0377), and blurb length (0.0159). In the explained observations, higher funding goals generally negatively contributed to the predicted success probability, whereas the tabletop games classification and longer preparation times generally contributed positively. Campaign-duration contributions displayed a nonmonotonic pattern. Therefore, the mean absolute SHAP importance reflects contribution magnitude, not the direction.

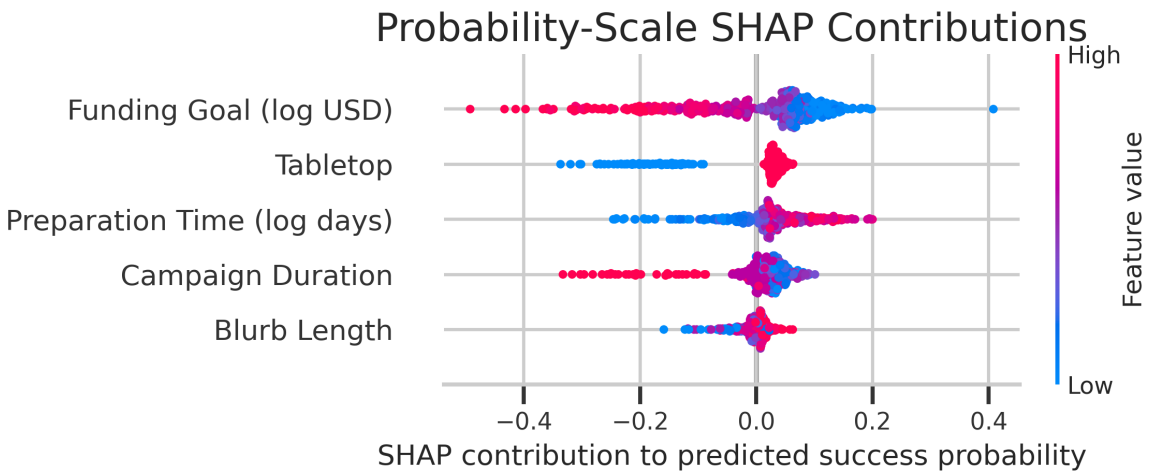


Figure 4. Probability-scale SHAP contributions for the temporal XGBoost model. Horizontal position represents each predictor's contribution to the predicted probability of campaign success, and color represents the observed predictor value.

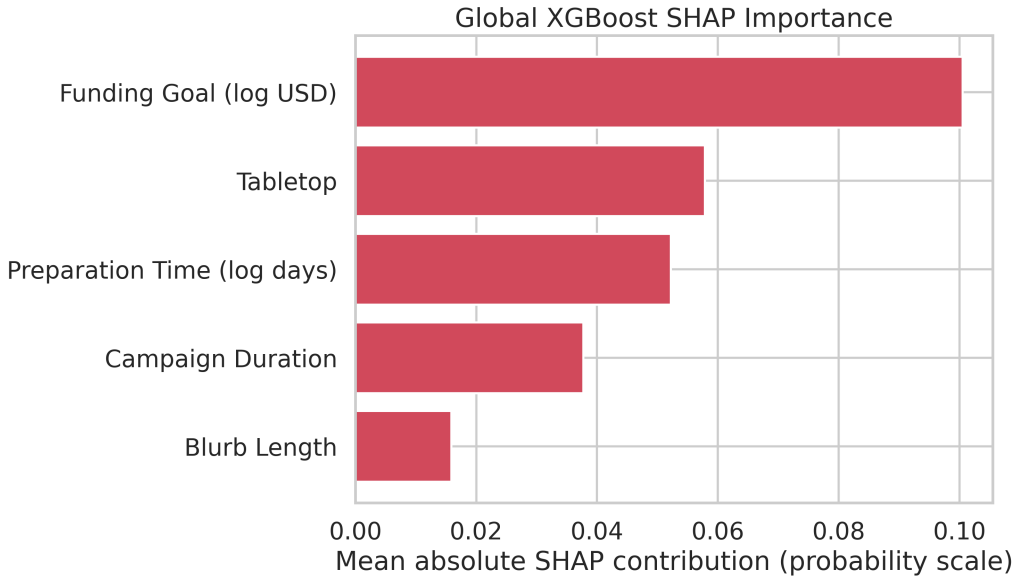


Figure 5. Global predictor importance for the temporal XGBoost model, measured by mean absolute probability-scale SHAP values. Directional patterns are shown in Figure 4.

4.5. *Model diagnostics, calibration, and sensitivity analysis*

The category-by-goal diagnostic uses numerically ordered funding-goal deciles (Figure 6). For Tabletop Games campaigns, the mean predicted success probability monotonically decreased as the funding-goal decile increased. The predicted probabilities for video games campaigns were lower in the highest-goal decile than in the sparsely populated lowest-goal deciles, although the pattern across intermediate deciles was not monotonic. A visible separation between the two categories remained across the plotted goal range. This diagnostic summarizes patterns in model predictions and should not be interpreted as an estimate of a causal category effect.

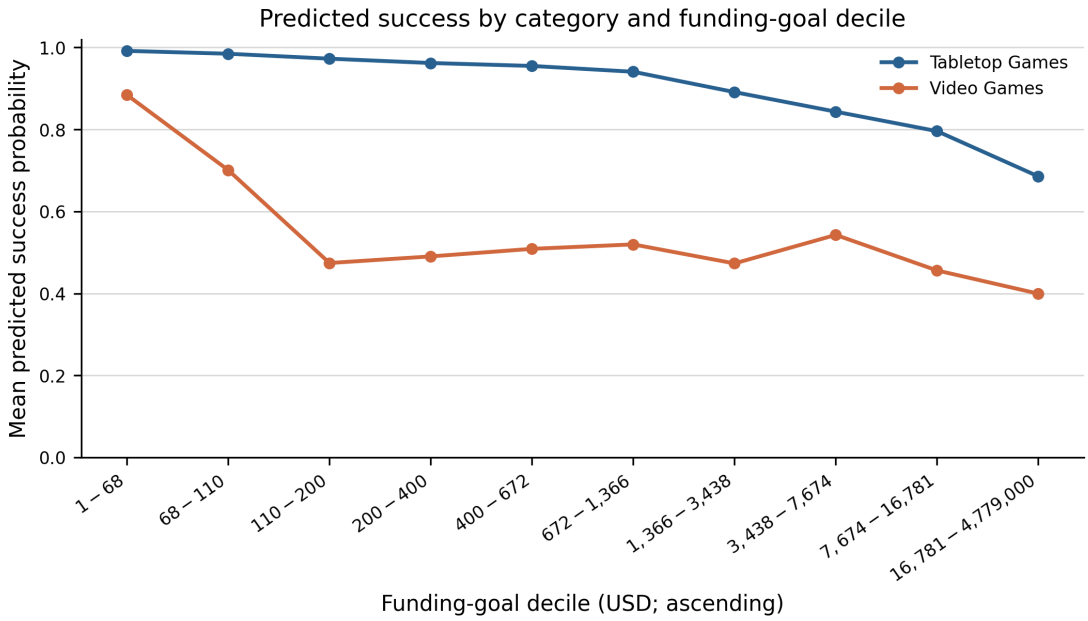


Figure 6. Mean predicted probability of campaign success from the temporal XGBoost model, stratified by the exact game category and numerically ordered funding-goal decile.

Adding Staff Pick to the XGBoost sensitivity specification increased the temporal-holdout ROC AUC from 0.879 to 0.899 and reduced the Brier score from 0.079 to 0.073 (Table 9). This improvement should be interpreted as incremental predictive information rather than evidence that platform endorsement causes campaign success. Staff pick may not be assigned or observable before launch and may reflect unmeasured campaign quality, platform selection, or information emerging after launch.

Table 9. Temporal XGBoost performance and Staff Pick sensitivity analysis.

Specification	Accuracy	AUC	PR AUC	Brier	Log loss	Kappa
Primary pre-launch XGBoost	0.891	0.879	0.972	0.079	0.268	0.502
XGBoost with Staff Pick	0.899	0.899	0.977	0.073	0.247	0.549

Note. Staff Pick was treated as an additional platform signal and was not included in the primary pre-launch model because its archived status could not be established as fixed at launch.

The partial-dependence surface showed generally lower predicted success probabilities at very

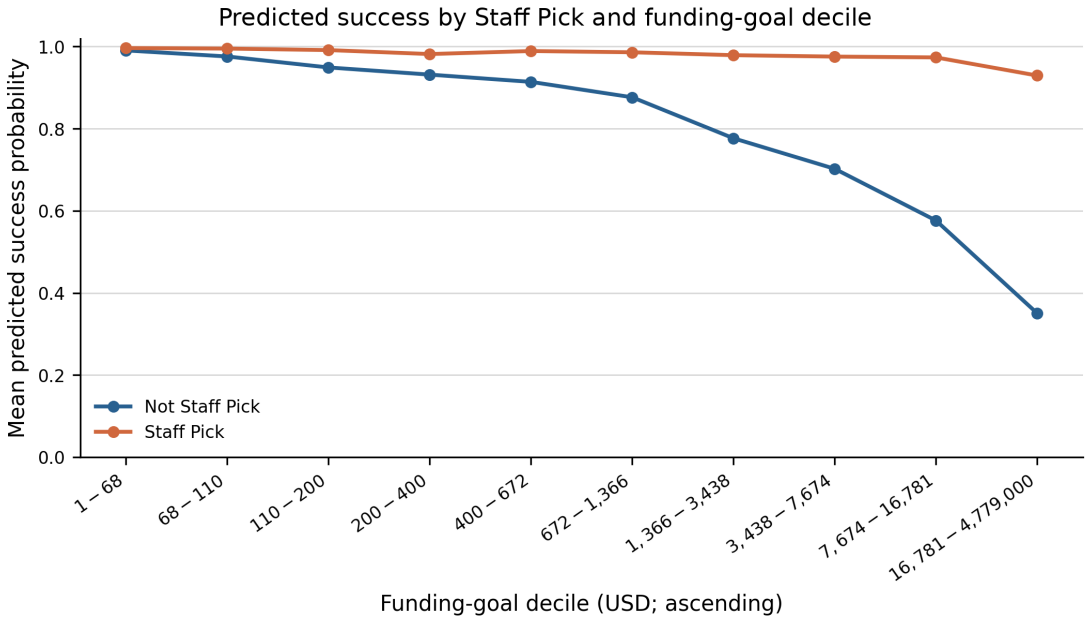


Figure 7. Mean predicted probability of campaign success from the XGBoost with Staff Pick sensitivity model, stratified by Staff Pick status and numerically ordered funding-goal decile. Several Staff Pick cells in the lowest funding-goal deciles were sparsely populated.

long campaign durations and higher predicted probabilities in regions combining moderate campaign durations with longer preparation times (Figure 8). Because partial dependence averages over the joint distribution of the remaining predictors, its interpretation may be affected by the predictor dependence and by combinations of preparation time and campaign duration that are sparsely represented in the data. Therefore, the surface should be treated as a model diagnostic rather than as prescriptive or causal evidence.

The primary pre-launch XGBoost model had a calibration slope of 0.942 and a calibration intercept of 0.263 on the 2025 temporal holdout (Figure 9).

5. Discussion

The combined descriptive record and the contemporary analytical component address distinct but complementary questions. The combined 2018–April 2026 record provides broader historical and descriptive context, whereas the 2022–April 2026 contemporary analytical component supports the statistical analyses, predictive modeling, temporal validation, and interpretation. Keeping these roles separate prevents limitations in the historical source from entering model fitting and evaluation while preserving the broader temporal context of the available data.

Across the combined descriptive record, 79.86% of campaigns were successful. The historical descriptive component only contained validated video games campaigns and had an observed success rate of 38.07%, whereas the contemporary analytical component contained both tabletop games and video games campaigns and had an observed success rate of 85.54%. This contrast should not be interpreted as evidence that the campaign success increased by a particular amount over time. The source coverage, category composition, and calendar period are inseparable in the assembled data.

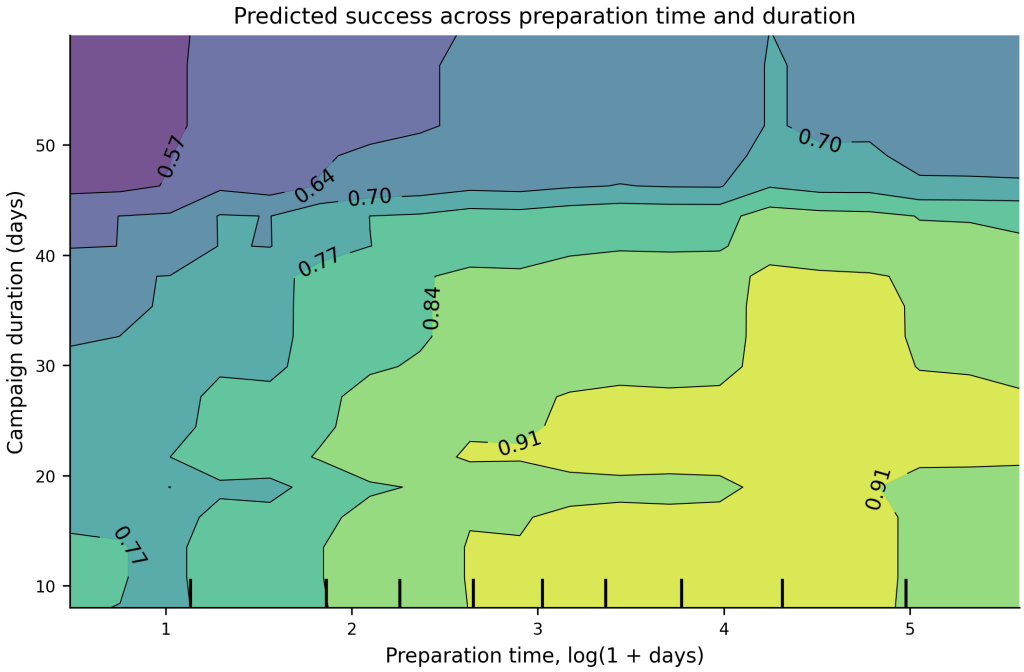


Figure 8. Two-dimensional partial dependence of temporal XGBoost predictions on preparation time and campaign duration. Contours represent predicted probabilities averaged over the remaining predictors in the temporal training sample.

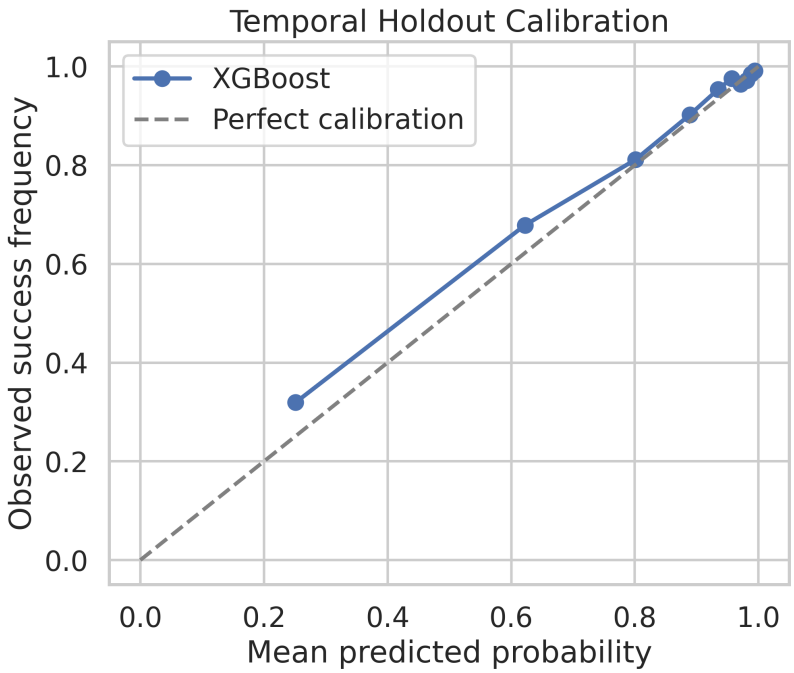


Figure 9. Calibration curve for XGBoost predicted probabilities on the 2025 temporal holdout. The diagonal reference line represents perfect agreement between the predicted and observed success probabilities.

The defensible conclusion is that the available descriptive record is heterogeneous across periods and categories.

Within the contemporary analytical component, the five pre-launch predictors supported useful out-of-sample prediction of campaign outcomes. Nested cross-validation placed the soft-voting ensemble and XGBoost at essentially the same level of discrimination, and evaluation on the 2025 temporal holdout only showed a modest reduction in the ROC AUC. The broad agreement between nested-cross-validation and temporal-holdout performance is encouraging, but it should not be interpreted as evidence of permanent model validity because the campaign composition, creator behavior, and platform conditions may continue to change.

The funding goal was the most consistently important predictor. It ranked first for all eight model classes under the common temporal-holdout permutation-importance metric and had the largest mean absolute probability-scale SHAP value for XGBoost. This pattern is consistent with a feasibility interpretation: larger funding requests may reflect a greater project scope, production complexity, or fulfillment risk. However, this predictive association does not imply that lowering a campaign's funding goal would itself increase its probability of success, because goals are jointly selected with other observed and unobserved project characteristics.

The exact game category was the second-ranked predictor for XGBoost, and the observed success rates substantially differed between Tabletop Games and Video Games within the contemporary analytical component. These results support category-sensitive modeling and interpretation rather than treating game crowdfunding as a homogeneous domain. The available variables do not identify why outcomes differ between categories. Differences in production timelines, prototype visibility, fulfillment requirements, audience structure, and community norms remain hypotheses for future research rather than measured mechanisms in the present study.

The preparation time ranked third for XGBoost and displayed nonlinear predictive contributions. A longer preparation interval may proxy for planning, audience development, or production readiness, but it may also arise from circumstances unrelated to the campaign quality. The campaign duration ranked fourth, and both the unadjusted descriptive results and model-based diagnostics showed less favorable patterns at longer durations. These patterns do not establish that shortening a campaign would improve its outcome. The blurb length contributed less than the other primary predictors, suggesting that the character count only captures a limited dimension of campaign communication.

The Staff Pick sensitivity analysis illustrates why the prediction time must be specified before predictors are selected. Adding the Staff Pick improved the temporal-holdout predictive performance, but the Staff Pick was not treated as a strictly pre-launch predictor because its archived status could not be established as fixed at launch. Additionally, it may encode platform selection, information emerging after launch, and latent campaign quality. Analyzing it separately preserves the temporal interpretation of the primary pre-launch model while quantifying the incremental predictive information available once the Staff Pick status is observed.

This sensitivity analysis provides a concrete empirical illustration of why the predictive performance must be interpreted relative to both the information available at the prediction time and the representation of that information used by the model. Adding the Staff Pick increased the temporal-holdout ROC AUC for XGBoost from 0.879 to 0.899. However, the augmented specification no longer conditions on exactly the same restricted information representation as the primary model and may also incorporate information only observable after launch. Therefore, the improvement should be

interpreted as incremental predictive information available once the additional Staff Pick signal is observed, rather than as evidence that the restricted primary specification was incorrectly defined.

Importantly, substantial predictive discrimination was retained even under the deliberately restricted primary specification. The soft-voting ensemble achieved a nested-cross-validation ROC AUC of 0.896 and a temporal-holdout AUC of 0.880, while XGBoost achieved corresponding AUCs of 0.895 and 0.879. Thus, the empirical contribution is not that the five individual characteristics are previously unknown predictors, but that a parsimonious structured pre-launch representation retains substantial out-of-sample and out-of-time predictive information.

Methodologically, the results caution against declaring a single model the definitive winner on the basis of one performance measure. The soft-voting ensemble achieved the highest nested-cross-validation and temporal-holdout ROC AUCs. XGBoost achieved a higher specificity and Cohen's kappa than the ensemble in nested cross-validation and had a marginally lower nested-CV Brier score. The neural network achieved the highest temporal accuracy and the lowest temporal-holdout Brier score. These differences among the leading models were small. Therefore, the model selection should depend on the intended prediction objective and on the relative importance assigned to discrimination, classification at a prespecified threshold, probability accuracy, interpretability, and computational complexity.

6. Limitations and future research

The combined descriptive record integrates two source vintages at a nonoverlapping boundary of January 1, 2022. The historical descriptive source contributes validated Video Games records but no comparable validated Tabletop Games records before 2022. The Web Robots archive was assembled through web scraping rather than as an administrative census and documents historical limitations in category browsing as well as possible duplication arising from subcategory crawling [12]. Although the present analysis deduplicated the observed records, it cannot recover campaigns that were absent from either source. Consequently, descriptive summaries from the combined record should not be interpreted as a census of Kickstarter campaigns or as unbiased estimates of temporal change.

The April 12, 2026 endpoint provides a conservative 60.1-day observation buffer relative to the June 11, 2026 data snapshot, but the 2026 observation period remains incomplete. The study was also restricted to the exact Tabletop Games and Video Games categories. Therefore, the findings should not be generalized to all projects in Kickstarter's broader Games category or to other reward-based crowdfunding platforms without external validation.

The primary predictor set was intentionally restricted to a structured pre-launch representation and does not exhaust the information potentially relevant to crowdfunding outcomes. Some omitted characteristics, including images, videos, full campaign text, reward structure, team composition, and verified creator history, may belong to the broader set of information available at or before launch but were not included in the restricted structured representation used here. By contrast, campaign updates, comments, backer activity, and evolving engagement or social-network signals arise during the campaign and therefore belong to a later, dynamically updated information setting. Therefore, the backer count was excluded from the primary predictive models because it is realized during the campaign, while the Staff Pick was separately examined because its status could not be established as fixed at launch. These omitted characteristics may contain additional predictive information, and their exclusion

should not be interpreted as evidence that they are irrelevant to the underlying crowdfunding process; rather, they fall outside the structured predictive representation examined in the primary analysis.

The campaign success prevalence was high, particularly for Tabletop Games. Although nested cross-validation was stratified and several class-sensitive performance measures were reported, specificity remained moderate across the leading models. Therefore, the classification threshold of 0.5 should not be assumed to be optimal for operational use. Future applications should select thresholds according to the relative consequences of false reassurance and false warning, as well as the intended decision context.

Finally, SHAP values, permutation importance, partial dependence, and grouped predicted-probability plots characterize the fitted-model behavior; none identifies the causal effect of changing a campaign attribute. Future research should assemble category-complete historical snapshots, incorporate richer text, image, video, reward, creator, and engagement-related predictors, and evaluate models using rolling-origin validation across multiple temporal snapshots. Additional work should also examine external validity on other crowdfunding platforms and use research designs capable of supporting stronger causal identification. As richer textual, visual, sequential, and network representations become available, future studies could also evaluate transformer-based, multimodal, recurrent, or graph-based neural architectures.

7. Conclusions

By separating the combined descriptive record from the contemporary analytical component, this study provides broader historical context without allowing limitations in pre-2022 category coverage to affect the predictive modeling. Within the contemporary analytical component, the five pre-launch predictors supported useful out-of-sample prediction of campaign outcomes.

The funding goal emerged as the most consistently important predictor, and the leading models achieved broadly similar out-of-sample performances across complementary evaluation metrics. Taken together, the findings support category-sensitive predictive modeling of game crowdfunding, but they should not be interpreted as evidence of causal effects.

Use of AI tools declaration

The authors declare that AI-assisted tools were used solely to improve linguistic clarity and readability in the final draft of the manuscript.

Acknowledgments

The authors would like to thank anonymous referees for useful comments and suggestions.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Spence M, (1973) Job market signaling. *Q J Econ* 87: 355–374. https://doi.org/10.2307/1882010
2. Mollick E, (2014) The dynamics of crowdfunding: An exploratory study. *J Bus Venturing* 29: 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005
3. Zheng H, Li D, Wu J, Xu Y, (2014) The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Inf Manage* 51: 488–496. https://doi.org/10.1016/j.im.2014.03.003
4. Cha J, (2017) Crowdfunding for video games: Factors that influence the success of and capital pledged for campaigns. *Int J Media Manage* 19: 240–259. https://doi.org/10.1080/14241277.2017.1331236
5. Gunduz H, (2024) Comparative analysis of BERT and FastText representations on crowdfunding campaign success prediction. *PeerJ Comput Sci* 10: e2316. https://doi.org/10.7717/peerj-cs.2316
6. Hohen S, Hüning C, Schweizer L, (2026) Reward-based crowdfunding—A systematic literature review. *Manage Rev Q* 76, 2997–3044. https://doi.org/10.1007/s11301-025-00544-2
7. Jónasdóttir H, (2019) Innovation in game development: IT enablement and affordances on Kickstarter. *Comput Games J* 8: 65–81. https://doi.org/10.1007/s40869-018-0075-6
8. Etter V, Grossglauser M, Thiran P, (2013) Launch hard or go home! Predicting the success of Kickstarter campaigns. In: *Proceedings of the First ACM Conference on Online Social Networks*, 177–182. https://doi.org/10.1145/2512938.2512957
9. Hoque MM, (2024) Crowdfunding for innovation: A comprehensive empirical review. *Future Bus J* 10: 102. https://doi.org/10.1186/s43093-024-00387-5
10. Varma S, Simon R, (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinf* 7: 91. https://doi.org/10.1186/1471-2105-7-91
11. Lundberg SM, Lee SI, (2017) A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30*, 4765–4774.
12. Web Robots, (2026) *Kickstarter Datasets: June 11, 2026 Snapshot and Archive Documentation*, Available from: https://webrobots.io/kickstarter-datasets/.
13. Kickstarter (n.d.), *What is the Maximum Project Duration? Kickstarter Help Center*, Available from: https://help.kickstarter.com/hc/en-us/articles/115005128434-What-is-the-maximum-project-duration.
14. Hastie T, Tibshirani R, Friedman J, (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, Springer. https://doi.org/10.1007/978-0-387-84858-7
15. Tibshirani R, (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 58: 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
16. Cortes C, Vapnik V, (1995) Support-vector networks. *Mach Learn* 20: 273–297. https://doi.org/10.1007/BF00994018
17. Chen T, Guestrin C, (2016) XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

18. Brier GW, (1950) Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 78: 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

19. Wilson EB, (1927) Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 22: 209–212. https://doi.org/10.1080/01621459.1927.10502953

Appendix

A. Supplementary tables

Table S1. Construction of the study datasets.

Stage	Count
Original engineered rows available for historical screening	11,303
Historical rows launched from January 1, 2018 through December 31, 2021	1309
Historical rows excluded for invalid goal, timing, duration, or outcome	1
Final historical descriptive component	1308
Contemporary exact-category rows before deduplication	18,067
Contemporary unique exact-category projects	13,511
Contemporary rows excluded for nonterminal outcomes	3398
Contemporary rows excluded outside January 1, 2022–April 12, 2026	484
Contemporary rows excluded for invalid goal, timing, or material duration	2
Final contemporary analytical component	9627
Final combined descriptive record, 2018–April 2026	10,935

Note. January 1, 2022 defines the nonoverlapping boundary between sources. The engineered historical source contributes pre-2022 descriptive records, whereas the June 11, 2026 Web Robots snapshot contributes all records from 2022 onward. The contemporary outcome-complete cutoff is April 12, 2026 at 00:56:11 UTC. Historical records were only used for descriptive reporting and were not included in model fitting, tuning, validation, or interpretation.

Table S2. Campaign-duration exceptions.

Source component	Project ID	Exact category	Calculated Duration (days)	Outcome
Historical descriptive	1894742905	Video Games	64.995	Failed
Contemporary analytical	1441307505	Video Games	66.958	Failed
Contemporary analytical	1554600022	Video Games	66.958	Successful

Note. Thirty-seven historical records and 68 contemporary records had calculated durations between 60 and 60.1 days, consistent with a one-hour timestamp-boundary shift, and were normalized to 60 days. The three material exceptions shown above were excluded from their respective components. The maximum retained campaign duration was 60 days.

Table S3. Campaign success rates by data component.

Data component	Group	*N*	Successful	Success (%)	95% CI
Historical descriptive: 2018–2021	Video Games	1308	498	38.07	[35.48, 40.74]
Contemporary analytical: 2022–April 2026	All campaigns	9627	8235	85.54	[84.82, 86.23]
	Tabletop Games	7898	7377	93.40	[92.83, 93.93]
	Video Games	1729	858	49.62	[47.27, 51.98]
Combined descriptive record: 2018–April 2026	All available campaigns	10,935	8733	79.86	[79.10, 80.60]
	Tabletop Games	7898	7377	93.40	[92.83, 93.93]
	Video Games	3037	1356	44.65	[42.89, 46.42]

Note. No comparable validated Tabletop Games records were available from the historical source before January 1, 2022. Therefore, the combined descriptive record reflects the category coverage of the available sources and should not be interpreted as a category-complete platform census or as an unbiased basis for estimating temporal change. All formal statistical, predictive, temporal-validation, and model-interpretation analyses only used the contemporary analytical component beginning January 1, 2022.



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)