2026 IEEE 6th International Conference on Smart Information Systems and Technologies (SIST)
13-15 May, 2026, Astana, Kazakhstan

2026 IEEE 6th International Conference on Smart Information Systems and Technologies (SIST) | 979-8-3315-8164-0/26/$31.00 ©2026 IEEE | DOI: 10.1109/SIST61674.2026.11596168

Conformal Prediction for Crowdfunding Risk Assessment: A Distribution-Free Approach to Uncertainty Quantification

Sultanbek Jakhanov
School of Information Technology and Engineering
Kazakh-British Technical University
Almaty, Kazakhstan
s_jakhanov@kbtu.kz
https://orcid.org/0009-0003-8946-3051

Zhomart Turarov
Faculty of Information Technologies
L.N. Gumilyov Eurasian National University
Astana, Kazakhstan
turarov_zhm_1@enu.kz
https://orcid.org/0009-0003-7995-3583

Abstract—Crowdfunding platforms have become a significant source of funding for entrepreneurs, yet predicting project success remains challenging. Traditional machine learning models provide point predictions without reliable uncertainty estimates, limiting their practical utility for risk-aware decision-making. This paper presents an integrated framework combining split conformal prediction, a three-tier risk categorization (LOW/HIGH/UNCERTAIN), and per-domain validation, providing statistically guaranteed prediction sets instead of point estimates. Using a dataset of 187,788 Kickstarter projects, we demonstrate that our approach achieves 90% coverage guarantee while improving prediction accuracy by 8.5 percentage points on certain predictions (84.7% vs. 76.2% baseline). We introduce a practical risk categorization framework (LOW RISK, HIGH RISK, UNCERTAIN) that enables backers to make informed investment decisions with quantified uncertainty. Our results show that conformal prediction offers a valuable tool for crowdfunding platforms seeking to provide users with reliable, uncertainty-aware predictions.

Index Terms—conformal prediction, crowdfunding, risk assessment, uncertainty quantification, machine learning

I. Introduction

Crowdfunding has revolutionized how entrepreneurs and creators fund their projects, with platforms like Kickstarter facilitating over $7 billion in pledges since 2009 [1]. This funding model enables creators to raise capital directly from potential customers, bypassing traditional financial intermediaries such as banks and venture capitalists. However, the success rate of crowdfunding campaigns remains around 40%, meaning that a significant portion of backers risk losing their investments in failed projects [2].

The high failure rate presents a significant challenge for potential backers who must decide whether to support a project based on limited information available at launch time. While creators provide project descriptions, funding goals, and timelines, backers lack reliable tools to assess the likelihood of project success and the associated investment risk.

Predicting crowdfunding success has attracted considerable research attention, with machine learning models achieving accuracy rates of 68-93% [3], [4]. However, these models typically provide point predictions (e.g., "87% chance of success") without reliable uncertainty estimates. Such predictions can be overconfident and fail to communicate the inherent uncertainty in the prediction process, potentially leading backers to make poorly informed decisions.

Conformal prediction [5] offers a solution by providing prediction sets with guaranteed coverage probability. Instead of outputting a single class, conformal prediction returns a set of possible outcomes that contains the true label with a user-specified probability (e.g., 90%). This approach provides distribution-free guarantees that hold regardless of the underlying data distribution.

Unlike prior applications of conformal prediction to domains such as medical diagnosis [6] and image classification, crowdfunding presents unique challenges: extreme heterogeneity across project categories, significant class imbalance variation, and the need for interpretable risk communication to non-expert users.

In this paper, we make the following contributions:

- We present an integrated framework that combines conformal prediction, three-tier risk categorization, and per-domain validation, applying it to crowdfunding success prediction as a decision-support layer over an imperfect base classifier.
- We demonstrate that conformal prediction achieves 90% coverage guarantee on real Kickstarter data, validating the theoretical properties of the method.
- We show that by abstaining on uncertain cases, prediction accuracy improves by 8.5 percentage points (from 76.2% to 84.7%).
- We introduce a practical risk categorization framework (LOW RISK, HIGH RISK, UNCERTAIN) that serves as a decision-support system for crowdfunding investment decisions, with per-category validation across 15 project domains.

The remainder of this paper is organized as follows. Section II reviews related work on crowdfunding prediction

and conformal prediction. Section III presents our methodology. Section IV describes our experimental setup and results. Section V discusses the findings and limitations. Section VI concludes the paper.

II. Related Work

A. Crowdfunding Success Prediction

Early work on crowdfunding prediction focused on identifying success factors. Mollick [2] conducted a seminal analysis of Kickstarter data and found that project quality, social capital, and geography significantly influence campaign success. The study revealed that projects with realistic funding goals and shorter durations tend to have higher success rates.

Greenberg et al. [3] developed one of the first machine learning approaches to crowdfunding prediction, using Random Forest classifiers to achieve 68% accuracy. Their work demonstrated that features available at campaign launch time contain predictive information about eventual success.

Mitra and Gilbert [7] demonstrated that linguistic features in campaign descriptions have significant predictive power, accounting for approximately 59% of the variance in funding success. Their analysis of 45,000 Kickstarter projects identified specific phrases associated with successful campaigns.

Recent studies have applied deep learning to this problem. Yu et al. [8] used deep neural networks for crowdfunding prediction, demonstrating improvements over traditional machine learning methods. Wang et al. [4] conducted a comprehensive multimodel comparative study using LSTM networks, achieving up to 93% accuracy. However, these high accuracies often rely on features that exhibit data leakage, such as the number of backers or amount pledged during the campaign, which are not available at launch time.

B. Conformal Prediction

Conformal prediction, introduced by Vovk et al. [5], provides a framework for constructing prediction sets with finite-sample coverage guarantees. The key property is that for any user-specified error rate α, the prediction set contains the true label with probability at least $1 - \alpha$, regardless of the underlying data distribution.

Split conformal prediction [9] is a computationally efficient variant that splits the training data into a proper training set and a calibration set. The calibration set is used to compute conformity scores that determine the prediction set sizes. This approach requires only a single model training step, making it practical for large-scale applications.

Recent applications of conformal prediction span diverse domains. In medical diagnosis, Olsson et al. [6] applied conformal prediction to pathology image analysis, achieving significant reductions in diagnostic errors. Vazquez and Bhavsar [10] provided a comprehensive review of conformal prediction applications in clinical medical sciences.

We are not aware of prior work applying conformal prediction to crowdfunding prediction, though the growing interest in uncertainty quantification for financial applications suggests this is a natural and timely direction.

III. Methodology

A. Problem Formulation

Let $D = \{(x_i, y_i)\}_{i=1}^{n}$ be a dataset of crowdfunding projects, where $x_i \in \mathbb{R}^d$ represents project features and $y_i \in \{0, 1\}$ indicates failure (0) or success (1). Our goal is to construct a prediction set $C(x) \subseteq \{0, 1\}$ such that:

$$P(y \in C(x)) \geq 1 - \alpha \tag{1}$$

where α is the user-specified error rate (e.g., $\alpha = 0.1$ for 90% coverage). This guarantee holds marginally over the joint distribution of (x, y), without assumptions about the data distribution.

B. Split Conformal Prediction

We follow the split conformal prediction framework [9], which consists of the following steps:

Step 1: Data Split. Divide the data into three disjoint sets: training set D_{train} for model fitting, calibration set D_{cal} for computing conformity scores, and test set D_{test} for evaluation.

Step 2: Model Training. Train a base classifier $\hat{f}$ on D_{train} to obtain class probability estimates $\hat{p}(y|x)$. Any probabilistic classifier can be used, including logistic regression, random forests, or gradient boosting.

Step 3: Conformity Scores. For each calibration example $(x_i, y_i) \in D_{cal}$, compute the conformity score as the complement of the predicted probability for the true class:

$$s_i = 1 - \hat{p}(y_i|x_i) \tag{2}$$

Intuitively, a higher score indicates that the model is less confident about the true label, suggesting higher uncertainty.

Step 4: Threshold Computation. Compute the $(1 - \alpha)$ quantile of the calibration scores with a finite-sample correction:

$$\hat{q} = \text{Quantile}\left(\{s_1, \ldots, s_{|D_{cal}|}\}, \frac{\lceil (1 - \alpha)(|D_{cal}| + 1) \rceil}{|D_{cal}|}\right) \tag{3}$$

Step 5: Prediction Set. For a new example x, include class y in the prediction set if its predicted probability exceeds the threshold:

$$C(x) = \{y : \hat{p}(y|x) \geq 1 - \hat{q}\} \tag{4}$$

This construction guarantees that the true label is included in the prediction set with probability at least $1 - \alpha$.

C. Risk Categories

We map the prediction sets to practical risk categories that provide actionable information for potential backers:

- LOW RISK: $C(x) = \{1\}$ (only success predicted). The model is confident that the project will succeed with the specified coverage guarantee.
- HIGH RISK: $C(x) = \{0\}$ (only failure predicted). The model is confident that the project will fail.
- UNCERTAIN: $C(x) = \{0, 1\}$ (both outcomes possible). The model cannot confidently distinguish between success and failure at the specified confidence level.

This categorization provides actionable information for potential backers, allowing them to make informed decisions based on the model's confidence level. Projects in the UNCERTAIN category may require additional due diligence or manual review.

This corresponds operationally to selective classification with a reject option [11], with UNCERTAIN as the reject region and the conformal threshold $\hat{q}$ providing a coverage-guaranteed abstention boundary.

IV. Experiments

A. Dataset

We use the Kickstarter dataset [12] containing 196,298 crowdfunding projects spanning multiple categories and countries. After removing canceled projects and those with missing values, we retain 187,788 projects with a 60.2% success rate. The dataset covers projects launched between 2009 and 2018. While this predates recent shifts in crowdfunding behavior, it remains one of the largest publicly available Kickstarter datasets and has been widely used in prior work [3], [4], enabling direct comparability with existing studies. Our primary contribution is methodological rather than dataset-specific.

We extract 14 features available at campaign launch time (Table I), carefully excluding features that would cause data leakage. Specifically, we exclude: (1) spotlight status, which is set after project success; (2) number of backers, which accumulates during the campaign; and (3) amount pledged, which is the outcome we aim to predict.

The data preparation pipeline involves several steps. First, we remove canceled and suspended projects, retaining only completed campaigns with a definitive success or failure outcome, reducing the dataset from 196,298 to 187,788 projects. Categorical features (project category with 15 classes, creator country with 22 classes) are encoded using ordinal label encoding. While this encoding imposes an arbitrary ordering on nominal features, tree-based models such as Gradient Boosting are invariant to monotonic transformations of individual features, as split decisions depend only on threshold comparisons [13]. Boolean features (staff pick status, creator profile completeness) are cast to integer values. We apply a logarithmic transformation to the funding goal ($\log_{10}(\text{goal}+1)$) to reduce skewness, and derive a goal intensity ratio (USD goal divided by campaign duration in days) as an additional engineered feature.

TABLE I
Feature Description

Feature	Description
Category	Project category (15 categories, encoded)
Country	Creator's country (22 countries, encoded)
Funding Duration	Campaign length in days
Pre-funding Duration	Days from creation to launch
Launch Month	Month when campaign started
Deadline Month	Month when campaign ends
Staff Pick	Binary: selected by Kickstarter staff
Creator Has Slug	Binary: creator profile completeness
Blurb Length	Character count of project blurb
Blurb Word Count	Word count of project blurb
Name Length	Character count of project name
Name Word Count	Word count of project name
Log Goal	Log-transformed USD funding goal
Goal per Day	USD goal divided by campaign duration

B. Experimental Setup

We split the data into training (60%), calibration (20%), and test (20%) sets using stratified sampling to preserve the 60.2% success rate across all splits. We compare three base classifiers:

- Logistic Regression with L2 regularization
- Random Forest with 200 trees and maximum depth 10
- Gradient Boosting with hyperparameters optimized via GridSearchCV

For the Gradient Boosting classifier, we perform hyperparameter optimization using GridSearchCV with 5-fold stratified cross-validation on the training set, optimizing for AUC-ROC. The search space includes: number of estimators $\{200, 300\}$, maximum depth $\{4, 6\}$, learning rate $\{0.05, 0.1\}$, and subsample ratio $\{0.8, 1.0\}$, yielding 16 parameter combinations. The best parameters found are reported in Section IV-B.

We use the MAPIE library [14] for conformal prediction implementation with the score-based method. All experiments are conducted using Python 3.9 with scikit-learn 1.3.

C. Results

1) Baseline Model Comparison: Table II shows the baseline model performance on the test set. The tuned Gradient Boosting classifier achieves the best performance across all metrics, with AUC-ROC of 0.824. The baseline accuracy of 76.2% represents a moderate improvement over the majority-class baseline of 60.2%, reflecting the inherent difficulty of crowdfunding prediction using only pre-launch features. This is consistent with the 68–93% range reported in prior work, noting that higher accuracies often rely on features unavailable at launch time [4]. This moderate baseline performance motivates the conformal

prediction approach: when a model's point predictions are imperfect, it becomes critical to identify which predictions are trustworthy. This model is selected as the base classifier for conformal prediction. The best hyperparameters found by GridSearchCV are: learning rate 0.05, maximum depth 6, 300 estimators, and subsample ratio 0.8 (best CV AUC-ROC: 0.818). Fig. 1 shows the ROC curve for the Gradient Boosting classifier.

TABLE II
Baseline Model Performance

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Regression	0.703	0.721	0.826	0.770	0.758
Random Forest	0.737	0.732	0.889	0.803	0.796
Gradient Boosting	0.762	0.768	0.865	0.814	0.824

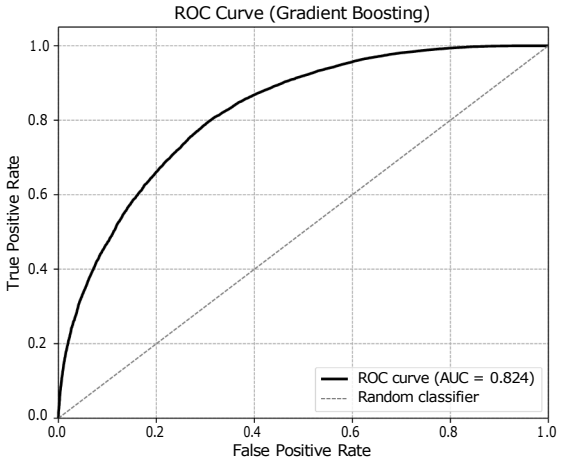

Fig. 1. ROC curve for the tuned Gradient Boosting classifier showing AUC = 0.824, indicating good discriminative performance.

Fig. 2 shows the feature importance analysis for the Gradient Boosting model. The most predictive features are Log Goal (0.272) and Staff Pick (0.178), together accounting for 45.0% of the total importance. This suggests that realistic funding targets are crucial for success, consistent with prior research [2].

2) Conformal Prediction Performance: Table III shows conformal prediction results at different confidence levels. The empirical coverage closely matches the target coverage at all levels, validating the theoretical guarantees of conformal prediction. At 90% confidence, the coverage is 90.3% (target: 90%), with 63.7% of predictions being certain (singleton sets). The accuracy on certain predictions is 84.7%, representing an improvement of 8.5 percentage points over the baseline accuracy of 76.2%. Given the test set size of 37,558 projects, the standard error of the accuracy estimate is approximately 0.2 percentage points, indicating that the observed 8.5pp difference is robust to sampling variability. This improvement is particularly meaningful given the moderate baseline: conformal prediction effectively filters out cases where the model

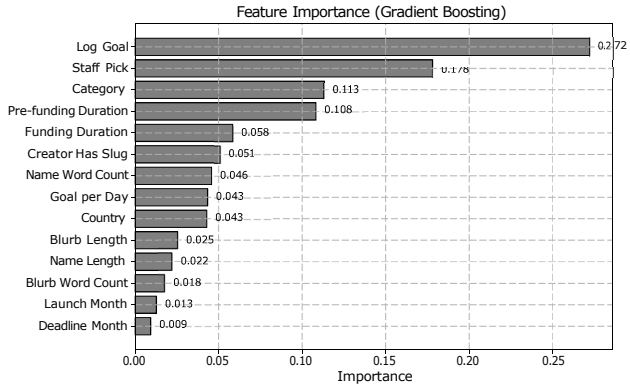

Fig. 2. Feature importance for crowdfunding success prediction. Log Goal and Staff Pick are the most important features.

is unreliable, concentrating accuracy on the 63.7% of predictions where the model has genuine discriminative power.

TABLE III
Conformal Prediction at Different Confidence Levels

Confidence	Coverage	Certain (%)	Accuracy
80%	80.6%	90.5%	78.5%
85%	85.3%	78.6%	81.3%
90%	90.3%	63.7%	84.7%
95%	95.0%	44.9%	88.8%
98%	98.0%	27.9%	92.9%

Fig. 3 illustrates the trade-off between confidence level and prediction certainty. As the confidence level increases from 80% to 98%, the percentage of certain predictions decreases from 90.5% to 27.9%, while the accuracy on those certain predictions increases from 78.5% to 92.9%. This trade-off allows users to choose an operating point that balances coverage guarantees with prediction efficiency.

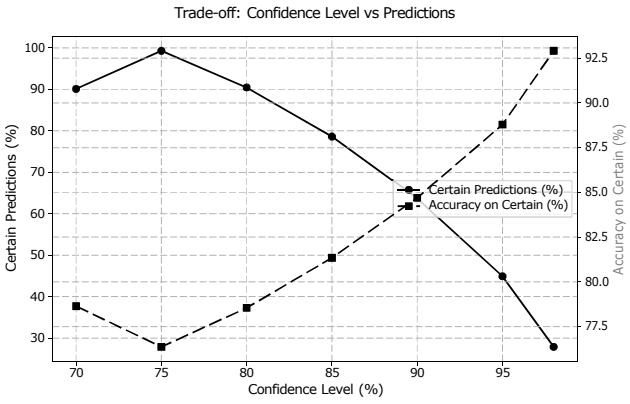

Fig. 3. Trade-off between confidence level, percentage of certain predictions, and accuracy on certain predictions. Higher confidence yields fewer but more accurate certain predictions.

3) Risk Category Analysis: Fig. 4 shows the distribution of risk categories and their corresponding accuracies

at 90% confidence. The test set of 37,558 projects is distributed as follows:

- LOW RISK: 17,217 projects (45.8%) with 83.9% actually successful
- HIGH RISK: 6,704 projects (17.8%) with 86.7% actually failed
- UNCERTAIN: 13,637 projects (36.3%) requiring manual review

The high accuracy within the LOW RISK and HIGH RISK categories (83.9% and 86.7%, respectively) demonstrates the practical utility of the risk categorization framework.

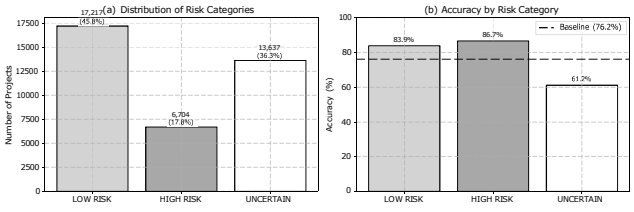

Fig. 4. Risk category distribution (left) and accuracy by category (right) at 90% confidence.

Fig. 5 shows the confusion matrix for the baseline Gradient Boosting classifier. The model correctly classifies the majority of both failed and successful projects, with stronger performance on the successful class due to the higher base rate.

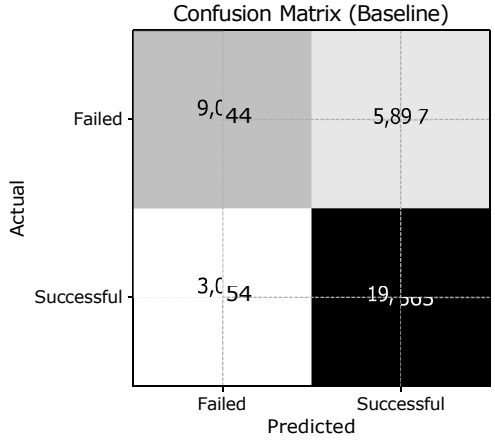

Fig. 5. Confusion matrix for the Gradient Boosting classifier on the test set.

Fig. 6 shows the accuracy on certain predictions broken down by project category at 90% confidence. Performance is consistently high across all 15 categories, demonstrating that the conformal prediction framework generalizes well across diverse project domains.

V. Discussion

Our results demonstrate that conformal prediction provides valuable uncertainty quantification for crowdfunding

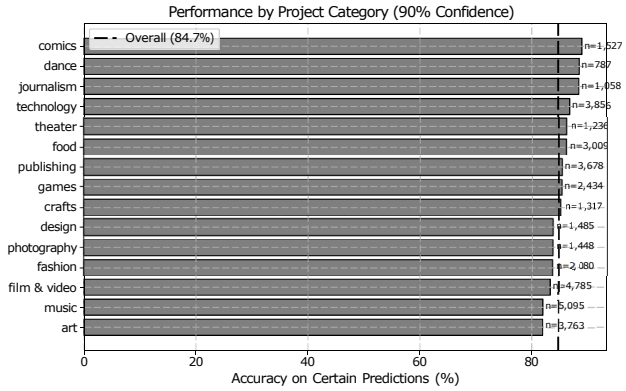

Fig. 6. Accuracy on certain predictions by project category at 90% confidence. Performance is consistent across all categories.

prediction. We discuss the key findings and their implications.

Coverage Guarantee. The empirical coverage closely matches the target across all confidence levels (Table III), confirming the theoretical guarantees of conformal prediction. This is a critical property for real-world deployment where reliable uncertainty estimates are essential for risk-aware decision-making.

Accuracy-Certainty Trade-off. Higher confidence levels yield fewer certain predictions but higher accuracy on those predictions (Fig. 3). At 90% confidence, 63.7% of predictions are certain (singleton sets) with 84.7% accuracy, versus 76.2% when the model is forced to commit on all projects. This 8.5pp gap reflects selective abstention rather than improvement of the underlying model. Crucially, conformal prediction is most valuable precisely when the base model is imperfect—it provides a principled mechanism to separate reliable predictions from unreliable ones, rather than treating all predictions as equally trustworthy.

Practical Utility. The risk categories provide actionable information for backers. LOW RISK projects have 83.9% success rate, providing strong confidence for potential backers. HIGH RISK projects fail 86.7% of the time, serving as a clear warning signal. The UNCERTAIN category (36.3% of projects) identifies cases requiring additional due diligence, such as reviewing project details, creator history, or waiting for early campaign traction.

Feature Insights. The feature importance analysis (Fig. 2) reveals that Log Goal and Staff Pick are the most predictive features, together accounting for 45.0% of total importance. This aligns with prior research showing that realistic goals are associated with success [2] and suggests that backers should pay particular attention to whether funding targets appear achievable.

Deployment Implications. The proposed risk categorization framework can be integrated into crowdfunding platforms as a traffic-light decision-support system: green (LOW RISK), red (HIGH RISK), and yellow (UNCER-

TAIN). Such a system would allow platforms to automatically flag uncertain projects for manual review while providing backers with interpretable risk assessments backed by statistical guarantees.

Limitations and Future Work. Our approach has several limitations that should be considered when interpreting the results.

Feature scope. The model relies exclusively on features available at campaign launch time, excluding dynamic signals such as early backer momentum, social media engagement, and comment sentiment that develop during the campaign. Incorporating such temporal features could improve both base model accuracy and conformal prediction efficiency.

Coverage type. The coverage guarantee is marginal rather than conditional, meaning it holds on average across all projects but may vary for specific subgroups. While Fig. 6 shows consistent performance across project categories, coverage could differ for underrepresented countries or extreme funding goals.

Temporal and platform scope. The dataset spans 2009–2018 (Kickstarter only), predating COVID-19, equity crowdfunding, and recent platform-policy changes. While the methodological contribution is dataset-agnostic, the specific figures may differ on post-2020 data and on other platforms (e.g., Indiegogo, GoFundMe).

Feature encoding. We use ordinal encoding for categorical features. Although tree-based models are theoretically invariant to monotonic feature transformations, one-hot or target encoding might yield different results with other model families.

Statistical testing. We report point estimates; the large sample size ($n = 37{,}558$) yields standard errors below 0.3pp, but bootstrap intervals and formal tests are left to future work.

Future work could extend this in several directions: (1) time-series conformal prediction for dynamic risk assessment as campaigns progress; (2) conditional coverage methods for subgroup-specific guarantees; (3) validation on post-2020 cross-platform data; and (4) integration into platform user interfaces as a real-time decision-support tool.

VI. Conclusion

This paper demonstrated that conformal prediction can transform an imperfect crowdfunding classifier into a reliable decision-support tool by honestly communicating when the model lacks confidence. Rather than forcing a binary prediction on every project, our approach identifies the 36.3% of cases where the model cannot reliably distinguish success from failure—and flags them for human judgment. On the remaining 63.7% of committed predictions, accuracy reaches 84.7% (vs. 76.2% when forced to commit on all), with coverage guarantees that hold regardless of the data distribution.

The practical implication is a deployable traffic-light risk framework: LOW RISK and HIGH RISK categories achieve 83.9% and 86.7% accuracy respectively, while the UNCERTAIN category explicitly communicates the limits of automated prediction. For a domain where billions of dollars flow through platforms annually, providing backers with statistically grounded risk assessments—rather than overconfident point predictions—has the potential to reduce investment losses and improve platform trust.

Several directions remain for future work. Prospective validation on post-2020 crowdfunding data would test the robustness of our findings under shifted platform dynamics. Extending the framework to conformal regression could provide prediction intervals for expected funding amounts, offering richer risk profiles. Finally, comparing reward-based platforms (Kickstarter) with equity-based crowdfunding would assess the generalizability of conformal risk categories across different investment structures.

References

[1] "Kickstarter stats," https://www.kickstarter.com/help/stats, 2024, accessed: 2026-05-04.

[2] E. Mollick, "The dynamics of crowdfunding: An exploratory study," Journal of Business Venturing, vol. 29, no. 1, pp. 1–16, 2014.

[3] M. D. Greenberg, B. Pardo, K. Hariharan, and E. Gerber, "Crowdfunding support tools: predicting success & failure," in CHI'13 Extended Abstracts on Human Factors in Computing Systems. ACM, 2013, pp. 1815–1820.

[4] W. Wang, H. Zheng, and Y. J. Wu, "Prediction of fundraising outcomes for crowdfunding projects based on deep learning: a multimodel comparative study," Soft Computing, vol. 24, no. 11, pp. 8323–8341, 2020.

[5] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic learning in a random world. Springer Science & Business Media, 2005.

[6] H. Olsson, K. Kartasalo, N. Mulliqi et al., "Estimating diagnostic uncertainty in artificial intelligence assisted pathology using conformal prediction," Nature Communications, vol. 13, no. 1, p. 7761, 2022.

[7] T. Mitra and E. Gilbert, "The language that gets people to give: Phrases that predict success on kickstarter," in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, 2014, pp. 49–61.

[8] P.-F. Yu, F.-M. Huang, C. Yang, Y.-H. Liu, Z.-Y. Li, and C.-H. Tsai, "Prediction of crowdfunding project success with deep learning," in 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE). IEEE, 2018, pp. 1–8.

[9] H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman, "Inductive confidence machines for regression," in European Conference on Machine Learning. Springer, 2002, pp. 345–356.

[10] J. Vazquez and H. Bhavsar, "Conformal prediction in clinical medical sciences," Journal of Healthcare Informatics Research, vol. 6, no. 3, pp. 241–264, 2022.

[11] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[12] M. Mouille, "Kickstarter projects dataset," https://www.kaggle.com/kemical/kickstarter-projects, 2018, accessed: 2026-05-04.

[13] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[14] V. Taquet, V. Blot, T. Morzadec, L. Lacombe, and N. Brunel, "Mapie: Model agnostic prediction interval estimator," 2022, version 0.8.x; accessed 2026-05-04. [Online]. Available: https://github.com/scikit-learn-contrib/MAPIE