

From Default Risk to Priced Intervention: A Cost-Benefit Approach to Credit Card Risk Management

Aryan Patre

Graduate Assistant, Texas Tech University, Lubbock, United States
aryanpatre77@gmail.com

Abstract:

Credit card default prediction is a mature application of machine learning in consumer finance, yet published studies typically stop at discrimination metrics without connecting a model's output to the economics of a specific intervention a bank could actually deploy. This paper builds an interpretable default-risk model using the public UCI Default of Credit Card Clients dataset (Yeh & Lien, 2009) and couples it to a cost-benefit framework built from cited, real-world figures. A logistic regression baseline achieves test ROC-AUC of 0.741, and a random forest comparison model achieves 0.780, on a cohort of 30,000 Taiwanese credit card clients with a 22.12% next-month default rate. Rather than a single flat dollar loss, this paper estimates avoided loss per client from the client's own most recent bill amount multiplied by a published average loss-given-default rate for credit cards (61.54%; Bellotti & Crook, 2012). Against this, two realistic intervention channels are costed: an SMS reminder (\$0.03) and a live-agent call (\$2.70, reused from a companion study), with effectiveness drawn from a randomized field experiment on credit card debt repayment (a 4.4-percentage-point reduction in delinquency; Bursztn, Fiorin, Gottlieb, & Kanz, 2019). Under these figures, the break-even cost per intervention for untargeted, blanket outreach is \$457.45, far above both realistic channel costs, so model-based targeting adds negligible net value over contacting every client at either cost. This paper argues the finding is not a property of the models but of the ratio between typical credit card exposure sizes and low-cost communication channels, and discusses what would need to change for precision targeting to matter in this setting.

Keywords: credit card default; credit risk; cost-benefit analysis; early-warning modeling; logistic regression; random forest; loss given default

I. INTRODUCTION

Consumer credit card debt in the United States exceeded \$1.1 trillion by the end of 2023, and credit card charge-offs, the point at which a bank writes off an account as an uncollectible loss, have risen substantially in recent years, reaching a multi-year high in 2024. Predicting which cardholders are likely to default on their next payment is a long-standing application of statistical and machine learning methods in consumer finance, but a bank's ultimate decision is not merely which customers are riskiest; it is whether contacting a given customer with a proactive intervention, a reminder, an outreach call, a restructured payment plan, is worth its cost relative to the loss it might avoid.

A. Motivation and Gaps

The academic literature on credit card default prediction is extensive, and the UCI Default of Credit Card Clients dataset used in this study has itself been analyzed in dozens of published studies, almost always reporting accuracy, AUC, precision, and recall as the outcomes of interest. A separate, largely disjoint literature in behavioral economics and credit risk management studies the loss given default of credit card accounts and the effectiveness of specific low-cost interventions, reminders and nudges, in reducing delinquency. These two literatures are rarely brought together: predictive-modeling papers rarely connect their probability outputs to a real intervention's cost and effectiveness, and intervention-effectiveness papers rarely discuss how a predictive model should be used to decide whom to contact.

This paper addresses that gap directly, coupling an interpretable default-risk model to a cost-benefit framework built entirely from cited, real-world figures, a published loss-given-default rate applied to each client's own observed

exposure, and a randomized field experiment's effectiveness estimate for a real, low-cost intervention. Rather than assuming a single intervention channel, this paper also compares two realistic channels of very different cost, an SMS reminder and a live-agent call, to test whether the resulting targeting recommendation is sensitive to that choice.

B. Contributions

- An interpretable logistic regression baseline and a random forest comparison model for next-month credit card default, built on a documented multicollinearity remediation process addressing several redundant engineered features.
- A cost-benefit framework that estimates avoided loss per client from the client's own observed exposure (most recent bill amount) and a published loss-given-default rate, rather than a single flat dollar figure applied uniformly, a more precise treatment than the companion studies in this series.
- A direct two-channel cost comparison (SMS versus live-agent call) that identifies the exact break-even cost per intervention at which model-based targeting would begin to add value, situating this paper's finding within a broader pattern observed across companion early-warning studies in other domains.

II. LITERATURE REVIEW

A. The Default of Credit Card Clients Dataset

The dataset used in this study was introduced by Yeh and Lien [1], who compared six data mining methods for predicting the probability of default among 30,000 credit card clients in Taiwan, observed from April to September 2005, with the outcome being default on the October 2005 payment. The dataset has since become a standard public benchmark for credit default prediction, offering genuinely interpretable

features, demographics, credit limit, six months of repayment status, and six months of bill and payment amounts, in contrast to anonymized transaction-level fraud datasets more commonly used for point-in-time fraud detection.

B. Loss Given Default for Credit Cards

A separate literature in credit risk management studies loss given default (LGD), the fraction of a defaulted balance a lender ultimately fails to recover. Bellotti and Crook [2] develop and validate several LGD models on UK retail credit card data, reporting a long-run default-weighted average loss rate of 61.54%, meaning roughly 38% of a defaulted balance is typically recovered through collections or debt sale. This paper applies that published rate directly to each client's own observed exposure rather than assuming a single dollar figure applies uniformly across very different account sizes.

C. Behavioral Interventions in Debt Repayment

A growing behavioral economics literature has tested low-cost reminder and nudge interventions for reducing consumer debt delinquency using randomized field experiments. Bursztyn, Fiorin, Gottlieb, and Kanz [3], working with a large Islamic bank in Indonesia, find that a simple moral-appeal text message reduces the share of delinquent credit card customers by 4.4 percentage points from a 66% baseline, and that the effect concentrates among customers with the highest ex-ante credit risk. Campbell, Grant, and Thorp [4], studying a broadcast repayment reminder for already-delinquent Australian credit card debtors, find a 2.7-percentage-point increase in the likelihood of repaying all arrears, though the effect is concentrated among debtors with stronger credit records. Both studies support this paper's central design choice: use a real, causally estimated effectiveness figure for a specific low-cost intervention rather than an assumed or illustrative one.

D. The Gap This Paper Addresses

Across the predictive-modeling literature on this dataset, evaluation is framed almost entirely in terms of discrimination metrics, with no connection to the financial consequence of acting on a prediction. Across the LGD and behavioral-intervention literatures, loss severity and intervention effectiveness are well studied in isolation but rarely combined into a single framework that a lender could use to decide, given a specific model and a specific intervention channel, whom to contact and at what threshold. This paper addresses that gap directly.

III. DATA AND METHODS

A. Dataset

This study uses the public UCI Default of Credit Card Clients dataset [1], comprising 30,000 client records and 24 attributes: demographics (sex, education, marriage, age), credit limit, six months of repayment status codes, six months of bill amounts, and six months of payment amounts, with the outcome indicating default on the following month's payment. The row count and column structure were verified against the dataset's own documentation before analysis. Two categorical fields contain undocumented codes, EDUCATION (values 0, 5, 6 are not defined in the official documentation) and

MARRIAGE (value 0 is not defined), a known data-quality property of this release; both were recoded into the documented "other/unknown" category rather than dropped.

B. Feature Engineering

Beyond the raw fields, several derived features were constructed: a utilization ratio (average bill amount relative to credit limit), a payment ratio (average payment relative to average bill, capturing minimum-payment or partial-payment behavior), a count of months with delayed payment and the maximum delay observed across the six-month window, an over-limit flag, and a count of zero-payment months. Unlike the companion studies in this series, this dataset provides no genuine per-client chronological ordering beyond the six repayment-status columns themselves, since every client is observed over the identical six-month window; those six PAY_* columns are used directly as features rather than collapsed into a single trend summary.

An initial feature set exhibited severe multicollinearity: several engineered features (a total-months-delayed count, a most-recent-delay indicator, and a bill-amount trend) were exact or near-exact linear combinations of raw features already in the model, producing undefined variance inflation factors, and the six monthly bill amounts were highly correlated with one another, a well-documented property of this dataset. These redundant features were removed in favor of the most recent bill amount and the six-month average bill amount, reducing the maximum VIF in the final feature set to approximately 12.

C. Train/Test Split

Because every client in this dataset is observed over the identical six-month window with no per-client calendar date, no genuine temporal train/test split could be constructed. A random, stratified 75/25 split was used: 22,500 training and 7,500 test clients, each preserving the cohort's 22.12% default rate.

D. Models

A logistic regression model was fit on standardized features as an interpretable baseline. A random forest classifier (300 trees, maximum depth 10, minimum 20 samples per leaf) was fit as a comparison model expected to capture non-linear interactions among repayment-history and utilization features. Missing values in the engineered payment-ratio feature (clients with a zero or negative average bill amount, for whom the ratio is undefined) were imputed with the training-set median.

E. Cost-Benefit Framework

At a given classification threshold, every client whose predicted probability meets or exceeds it is flagged for a proactive intervention. A cost is incurred for every flagged client; avoided loss is recovered only for true defaulters who are diverted by the intervention, at the assumed effectiveness rate, applied to that client's own exposure:

$$Net\ Value(t) = \sum EAD \times LGD \times E_{ff} - N \times Cost \quad (1)$$

where the summation runs over true defaulters flagged at threshold t , EAD (exposure at default) is each client's own most recent bill amount, LGD is the published loss-given-default rate

(61.54%), Eff is intervention effectiveness (6.67%, a relative risk reduction), N is the total number of clients flagged, and Cost is the per-intervention cost of the channel used.

IV. RESULTS

A. Predictive Performance

Table I summarizes test-set performance for both models.

TABLE I. PREDICTIVE PERFORMANCE (TEST SET)

Metric	Logistic	Random Forest
ROC-AUC	0.741	0.780
Accuracy	0.815	0.818
Precision	0.650	0.663
Recall	0.357	0.364

Both models substantially exceed chance-level discrimination and are broadly consistent with figures reported elsewhere for this dataset and prediction task. The random forest model outperforms the logistic baseline on every metric, though the gap is more modest than in some of the companion early-warning studies in this series, suggesting the underlying repayment-history features carry most of the available signal regardless of model family.

B. Feature Importance and Interpretation

For both models, the single most important predictor was the most recent repayment status (PAY_1), followed by the maximum delay observed across the six-month window, the engineered utilization ratio, and the engineered payment ratio. This ordering is intuitive from a credit-risk standpoint: how a client is currently behaving on their account, whether they are delayed right now and how severely, dominates over slower-moving signals like credit limit or demographics, consistent with the general finding in the credit scoring literature that recent behavioral variables outperform static application-level variables for short-horizon default prediction.

V. COST-BENEFIT INTERVENTION-TARGETING FRAMEWORK

A. Sourced Figures

- Loss given default: 61.54%, the long-run default-weighted average loss rate for UK retail credit cards (Bellotti & Crook [2]), applied to each client's own most recent bill amount as exposure at default rather than a single flat dollar figure.
- Cost per intervention: \$0.03 for an SMS reminder (conservative point within the industry-standard \$0.01–\$0.05 per-message range across several 2026 SMS platforms), compared against \$2.70 for a live-agent call (reused from a companion study's call-center industry benchmark).
- Effectiveness: 4.4 percentage points absolute reduction in delinquency from a 66% baseline, a 6.67% relative risk reduction, from a randomized field experiment on credit card debt repayment (Bursztyn et al. [3]).

B. Results

Fig. 1 plots net value against classification threshold for both models under the SMS channel; net value is positive across almost the entire threshold range and is maximized at low thresholds: \$3,405,149 at $t = 0.02$ for logistic regression and \$3,430,655 at $t = 0.04$ for random forest, on a test set of 7,500 clients. A naive policy of flagging every client, with no model at all, yields \$3,430,654, effectively identical to the random forest's model-optimized value. Switching to the costlier call channel barely changes this picture: the naive "everyone" policy still yields \$3,410,629, only about 0.6% lower than the SMS scenario, because the average avoided loss per client, \$31,814, dwarfs even a \$2.70 contact cost.

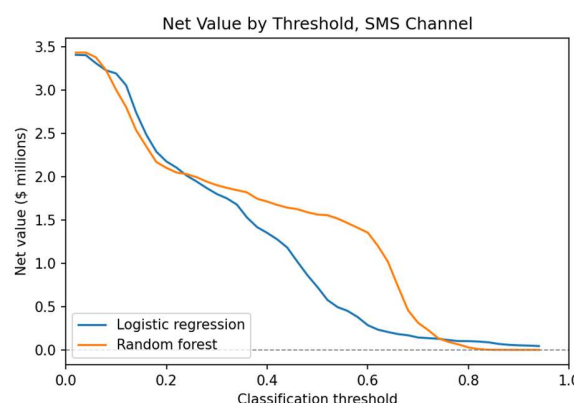


Fig. 1. Net value by classification threshold, both models, SMS channel, in millions of dollars.

To make this concrete, this paper computes the exact break-even cost per intervention at which an untargeted, blanket-outreach policy stops being net-positive: \$457.45. Fig. 2 places both realistic channel costs against this threshold on a logarithmic scale. The SMS channel sits roughly 15,000 times below break-even and the call channel roughly 170 times below break-even, explaining why neither channel's targeting recommendation is sensitive to model choice or threshold: the intervention would need to become dramatically more expensive, closer to the scale of a specialized financial counseling session than a reminder message or a routine collections call, before precision targeting would meaningfully change the economics.

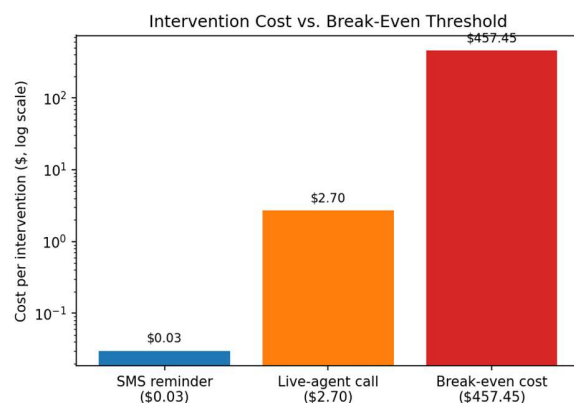


Fig. 2. Intervention cost per channel against the \$457.45 break-even threshold, log scale.

C. Interpretation

This finding generalizes a pattern also observed in a companion healthcare readmission study using the same framework: when a real, low-cost intervention channel is compared against a real, large avoided loss, the economics favor broad rather than narrow targeting, and the predictive model's main contribution shifts from precisely ranking individual risk to providing a defensible, auditable basis for reaching every client who plausibly needs contact. The two-channel comparison in this paper adds a further layer: because the break-even threshold is quantified explicitly, a bank can immediately see how much room exists before a costlier, higher-touch intervention, a phone-based financial counseling session rather than a call center script, would need to be justified by materially higher effectiveness than the SMS or basic-call figures used here.

VI. DISCUSSION

A. Why Credit Card Losses Favor Broad Targeting

The scale of a typical credit card exposure, several tens of thousands of NT dollars' worth of bill amount in this dataset's population, is large relative to the cost of almost any communication-based intervention channel. This is a structural feature of consumer credit risk rather than a property of the models tested here: as long as a bank's chosen intervention costs meaningfully less than the loss-given-default-weighted exposure of a typical account, precision targeting will add only marginal value over broad outreach, and the break-even calculation in Section V-B gives a bank a concrete number against which to check any specific channel it is considering.

B. Contrast with Companion Domains

This finding echoes, but is not identical to, a companion study's conclusion in a hospital readmission setting, where a real evidence-based intervention was similarly cheap relative to its avoided cost. It also contrasts with a companion subscription-churn study, where a costlier live-agent outreach channel failed to clear its break-even bar at all under conservative revenue assumptions. Read together, these three findings support a general principle: the value of predictive targeting in an early-warning system is not fixed but depends entirely on where a specific domain's intervention cost sits relative to its avoided-loss scale, a ratio that must be checked empirically rather than assumed.

C. Interpretability Versus Predictive Power

As in the companion studies, the logistic regression and random forest models offer complementary insight. The logistic regression's coefficients give a credit risk team directly interpretable, signed relationships useful for explaining a flagged account; the random forest's feature importances indicate which behavioral signals the model relies on most, with the same caution as in the companion studies regarding potential bias toward high-cardinality or continuous features relative to sparse categorical ones [5].

D. Practical Implications for Issuer Deployment

For a card issuer considering a system like this one, the practical implication is that channel selection, not model selection, is the first-order decision. Given the break-even figure computed here, an issuer using SMS or routine outbound calling should expect broad, low-friction outreach to capture nearly all of the available value, and should reserve investment in more sophisticated targeting for higher-cost interventions, dedicated relationship-manager outreach or in-person financial counseling, where the economics genuinely depend on precision.

VII. LIMITATIONS AND FUTURE WORK

Historical, non-US data. The dataset reflects Taiwanese credit card clients observed in 2005; default patterns, typical exposure sizes, and regulatory context have changed since, and the loss-given-default figure used here is drawn from UK rather than Taiwanese data, a genuine cross-market approximation.

Uniform effectiveness assumption. A single relative risk reduction is applied to every flagged true defaulter regardless of account size or risk profile; the source RCT itself found effects concentrated among higher-risk customers, suggesting the true economics could be even more favorable to targeting than this paper's uniform assumption implies.

No temporal validation. Because every client is observed over the same six-month window, this paper cannot evaluate whether the model generalizes to a genuinely future period, unlike designs with true sequential cohorts.

Cross-domain LGD estimate. The 61.54% loss-given-default figure is drawn from UK credit card data rather than this dataset's own Taiwanese population, since no equivalent published LGD estimate for Taiwan was located; this is flagged as an estimate rather than a precise match.

Future work could extend this framework by validating the loss-given-default and effectiveness figures against a more recent, market-matched dataset; modeling heterogeneous intervention effectiveness by risk segment rather than a uniform rate; and testing whether the break-even-cost approach introduced here generalizes to other consumer credit products with different typical exposure sizes, such as personal loans or auto loans.

VIII. CONCLUSION

This paper builds an interpretable credit card default model and couples it to a cost-benefit framework built entirely from cited, real-world figures, a published loss-given-default rate applied to each client's own exposure, and a randomized field experiment's effectiveness estimate for a real intervention. Both models achieve solid discrimination consistent with published benchmarks. The resulting economic analysis computes an exact break-even cost per intervention, \$457.45, and shows that two realistic communication channels, an SMS reminder and a live-agent call, both sit far enough below this threshold that model-based targeting adds negligible value over blanket outreach at either cost.

More broadly, alongside its companion studies in education, subscription retention, and healthcare, this paper demonstrates that the value of predictive early-warning modeling is not a fixed property of the model or the domain, but a direct function of the ratio between a chosen intervention's cost and the loss it can help avoid. Quantifying that ratio explicitly, as this paper's break-even calculation does, gives a business a concrete, auditable answer to the question that matters most before deploying any predictive system: not how accurate is the model, but at what cost does acting on it start to pay for itself.

REFERENCES

- [1] I.-C. Yeh and C.-H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, 2009.
- [2] T. Bellotti and J. Crook, "Loss given default models incorporating macroeconomic variables for credit cards," *Int. J. Forecast.*, vol. 28, no. 1, pp. 171–182, 2012.
- [3] L. Bursztyn, S. Fiorin, D. Gottlieb, and M. Kanz, "Moral incentives in credit card debt repayment: Evidence from a field experiment," *J. Polit. Econ.*, vol. 127, no. 4, pp. 1641–1683, 2019.
- [4] D. Campbell, A. R. Grant, and S. Thorp, "Reducing credit card delinquency using repayment reminders," *J. Bank. Finance*, vol. 142, art. 106549, 2022.
- [5] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, art. 25, 2007.