Proceedings of CONF-SEML 2026 Symposium: Multimodal Data Acquisition: Applications in Physiological and Behavioral Research
DOI: 10.54254/2755-2721/2026.AD33345

# *Scorecard-Guided Credit Limit Assignment with Thompson Sampling*

**Hanchao Yu**

*College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China*
*15882233313@163.com*

***Abstract.*** Credit scoring remains a standard tool in the banking industry because it provides an interpretable way to estimate default risk. In particular, scorecards based on Weight of Evidence (WOE), Information Value (IV), and Logistic Regression (LR) are widely adopted due to their transparency and practical usefulness. However, traditional scorecards are mainly used to assess customers' risk. They do not directly optimize subsequent actions, such as credit limit allocation. This paper proposes a two-stage framework for consumer credit analysis. In the first stage, it constructs a scorecard model using an anonymized bank dataset. The workflow includes data cleaning, decision-tree-based binning, WOE/IV transformation, correlation filtering, stepwise logistic regression, and scorecard construction. The resulting model achieves moderate predictive power, with a KS value of 0.386 and an AUC of 0.739. In the second stage, the scorecard is extended into a sequential decision-making framework. Predicted risk levels from the scorecard are used to define feasible credit-limit actions, and Thompson Sampling is applied to learn adaptive limit-allocation policies under uncertainty. In the simulation, three credit-limit levels are considered, and the reward is designed as a simplified risk-adjusted return. The final cumulative reward reaches 276,550, and the learned policy exhibits intuitive risk-consistent behavior: high-risk customers have their credit limits limited to low levels, medium-risk customers have their credit limits mainly at medium levels with a small portion at low levels, while low-risk customers are more likely to have higher credit limits. Overall, this study suggests how a traditional scorecard can be extended from pure risk prediction to scorecard-guided credit limit decision-making. The result also remains interpretable while introducing an adaptive bandit-based decision layer.

***Keywords:*** Credit Scorecard, Weight of Evidence (WOE), Logistic Regression, Thompson Sampling, Credit Limit Assignment

## 1. Introduction

Credit risk management is crucial to the stability and profitability of banking institutions. When banks issue loans or installment-based credit products, they face uncertainty regarding whether customers will repay on time or default. A common method for controlling this uncertainty is to predict default risk through credit scoring models. Among many alternatives, scorecards based on WOE/IV transformation and logistic regression are widely used in practice. They have three

DOI: 10.54254/2755-2721/2026.AD33345

advantages: interpretability, consistency, and regulatory acceptability. A scorecard not only provides a predicted probability of default, but also an additive point-based representation that is easy for analysts and practitioners to understand.

However, traditional credit scoring cards primarily address the prediction problem; lenders still need to make subsequent decisions. Credit limit allocation is a prime example. Even after a customer's credit score has been calculated, the lender still needs to decide whether to allocate a low, medium, or high credit limit. Different credit limits imply different trade-offs between expected gains and potential losses. This creates a natural sequential decision-making problem under uncertainty.

Therefore, this study extends a standard credit scorecard into a scorecard-guided bandit framework. The scorecard is first used to estimate default risk and group customers by predicted risk. Then, a Thompson Sampling policy selects a credit limit from the feasible action set defined by the scorecard. This study helps not only the construction of an interpretable scorecard, but also the design of a two-stage framework that connects risk estimation with adaptive credit limit assignment.

## 2. Related work

### 2.1. Credit scoring and scorecards

Credit scoring has been extensively studied in both academia and industry. Previous research has shown that Weight of Evidence (WOE)-based transformations can improve the interpretability of credit models while maintaining good predictive performance [1]. Logistic regression scorecards, which balance predictive power and transparency, remain commonly used benchmark models in credit risk modeling [2-4]

Research on the construction of personal credit scores typically follows a standard workflow, including variable binning, WOE calculation, information value-based variable selection, and logistic regression estimation [2,5]. Other studies on bank credit scorecards emphasize the importance of scorecard calibration, variable selection, and model validation in real-world lending environments [3,4,6]. These studies motivate the first-stage methodology used in this paper.

### 2.2. Bandits and adaptive decision-making

Multi-armed bandits provide a framework for sequential decision-making under uncertainty. In bandit problems, decision-makers repeatedly choose from multiple actions and observe stochastic feedback, with the goal of maximizing cumulative reward over time. Bandit methods are widely used in recommendation systems, advertising, dynamic pricing, and other fields that require a balance between exploration and exploitation [7,8].

Among classical bandit algorithms, Thompson Sampling provides a probabilistic mechanism for balancing exploration and exploitation [7-9]. Rather than always choosing the empirically best action, it samples from posterior beliefs and therefore maintains a natural degree of exploration [8-10].

This study does not replace scorecards with a bandit model. Instead, it uses the scorecard as an interpretable risk-estimation layer and adds a bandit-based decision layer for credit limit assignment. In this sense, the present work is motivated by both the traditional credit-scoring literature and the bandit literature.

DOI: 10.54254/2755-2721/2026.AD33345

## 3. Methodology

The proposed framework comprises two interconnected components, including a traditional scorecard model for interpretable risk assessment, and Thompson's sampling simulation for scorecard-guided adaptive credit limit allocation. The complete workflow can be summarized as follows: Customer Data → Scorecard → Risk Estimate/Risk Grouping → Bandits Layer → Credit Limit Assignment

### 3.1. Data preparation

The datasets used in this study contains anonymized bank loan and installment-payment records. Each row corresponds to one customer application, and some variables contain missing values. The last column lists the target variables for each customer, indicating whether the customer becomes overdue.

$$\mathrm{OVDU} = \begin{cases} 1, & \text{default / overdue,} \\ 0, & \text{non-overdue.} \end{cases} \tag{1}$$

The dataset contains demographic, behavioral, financial, and external-credit variables. Representative predictors include age, gender, marital status, loan amount, installment payment term, income, identity verification status, and external platform credit scores.

Data preprocessing includes handling missing values, separating predictor variables from target labels, and standardizing the representations of categorical and continuous variables. Variables with high cardinality or continuous support are prepared for discretization, while low-cardinality variables are retained in categorical form..

### 3.2. Decision-tree-based binning

To improve stability and interpretability, continuous variables or high cardinality variables are discretized using a decision tree-based method. This method identifies candidate split points by recursively partitioning the variable space according to the target label.

Compared to arbitrary manual binning, decision tree-based binning methods can better reflect the empirical relationship between predictors and default behavior, while maintaining interpretable segmentation.

### 3.3. WOE and IV calculation

After binning, each variable is transformed using WOE. For bin $i$,

$$\mathrm{WOE_i} = \ln\left(\frac{\mathrm{Bad_i/TotalBad}}{\mathrm{Good_i/TotalGood}}\right). \tag{2}$$

WOE converts raw bins into values that are more suitable for linear modeling and easier to interpret in a scorecard setting.

The predictive strength of each variable is measured by IV:

$$\mathrm{IV} = \sum_{i=1}^{k} \mathrm{WOE_i}\left(\frac{\mathrm{Bad_i}}{\mathrm{TotalBad}} - \frac{\mathrm{Good_i}}{\mathrm{TotalGood}}\right). \tag{3}$$

DOI: 10.54254/2755-2721/2026.AD33345

Variables with stronger separation ability between good and bad customers tend to have larger IV values.

### 3.4. Correlation filtering and logistic regression modeling

After WOE transformation and IV screening, highly correlated variables are first filtered using a correlation matrix, which is designed to reduce redundancy and improve model stability. When two candidate predictors are highly correlated, one is removed according to the filtering rule implemented in the code.

Then, logistic regression based on the remaining WOE-transformed variables is used to estimate the scorecard model. Let p represent the predicted probability of default. The model takes the following form:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \mathrm{WOE}_1 + \beta_2 \mathrm{WOE}_2 + \cdots + \beta_n \mathrm{WOE}_n. \tag{4}$$

To further improve the model, stepwise selection method is applied to determine the final set of predictor variables. The resulting coefficients provide an interpretable estimate of how each transformed variable contributes to default risk.

### 3.5. Scorecard construction

Once the logistic regression model is obtained, it is converted into a scorecard. Each variable bins will receive a partial score, and the total score is calculated by summing the scores across all variables.

This step is crucial for practice in that it transforms the logistic-regression output into a scoring system that is easy for analysts to understand. In general, the higher the total score, the lower the risk of default.

### 3.6. Scorecard-guided bandit framework

After a customer has been scored, the lender still needs to decide which credit limit to assign. In this paper, credit limit assignment is modeled as a simplified multi-armed bandit problem. Each arm corresponds to a limit level:

$$\mathscr{A} = \{2000,\ 5000,\ 10000\}. \tag{5}$$

The goal is not merely to estimate risk, but to learn which action generates better cumulative reward over time.

The scorecard output is used to classify customers into three risk groups based on predicted default probability. Specifically, in the low-risk group, all limits are feasible; in the medium-risk group, only the low and medium limits are feasible; and in the high-risk group, only the low limit is feasible. Thus, the scorecard does not directly choose the final action, but it constrains the feasible action set in a risk-consistent way.

To keep the simulation interpretable, reward is defined as a simplified risk-adjusted return:

$$R = \begin{cases} 0.05 \times \text{limit}, & \text{if the customer is non-default,} \\ -0.25 \times \text{limit}, & \text{if the customer defaults.} \end{cases} \tag{6}$$

DOI: 10.54254/2755-2721/2026.AD33345

This reward is not intended to replicate a production-level banking profit model. Instead, it captures the basic trade-off that higher limits can generate larger upside but also larger downside.

In each feasible action set, this study applies Thompson sampling to select the credit limit. Thompson sampling can sense the uncertainty of arm mass and sample based on these beliefs, thus balancing exploration and exploitation. In this study, it is used as a simple adaptive decision rule rather than as a full Bayesian business deployment system. Its role is to learn a profitable decision pattern inside the constraints imposed by the scorecard.

### 3.7. Evaluation metrics

The scorecard is evaluated using KS and AUC. The KS statistic measures the maximum separation between the cumulative distributions of good and bad customers. In other words, it reflects how well the model distinguishes between the two groups, and a larger KS value generally indicates stronger discriminatory power. The AUC measures the overall ranking ability of the model across thresholds. A larger AUC means that the model is better at ranking risky customers above non-risky customers. For the bandit stage, the main outputs are cumulative reward, final cumulative reward, and credit-limit selection patterns by risk group.

## 4. Results and discussion

### 4.1. Scorecard performance and interpretability

The scorecard was successfully executed on the anonymized dataset. The final model produces both a predicted probability of default and an interpretable total score for each customer. The key performance results are shown in Figure1 and Figure 2. The KS value is 0.386 and the AUC is 0.739. These values indicate moderate predictive power.

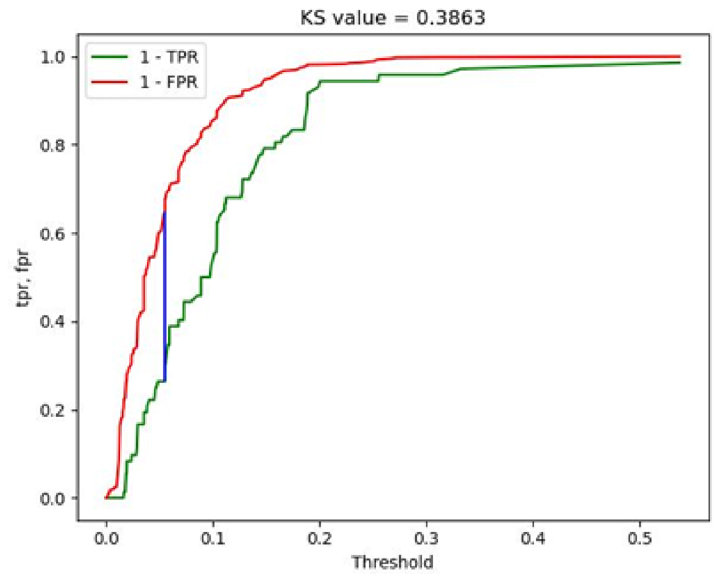


Figure 1. KS value curve

DOI: 10.54254/2755-2721/2026.AD33345

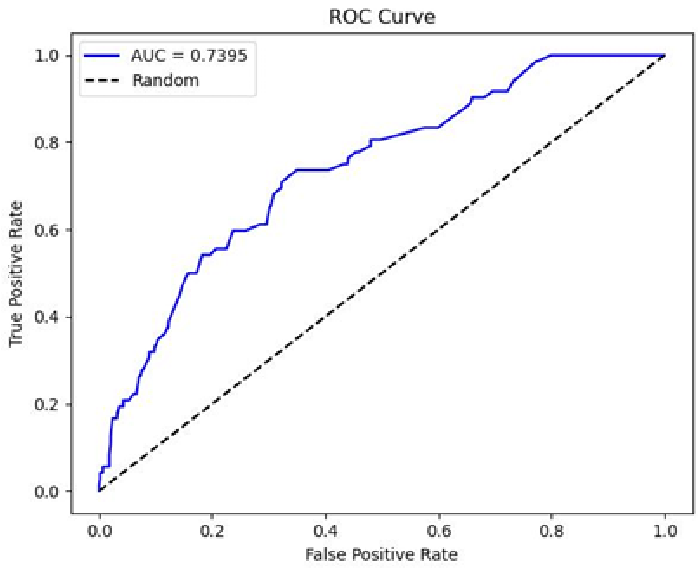


Figure 2. ROC curve

The scorecard representation also remains interpretable. Each variable bin contributes a partial score, and the total score provides a compact summary of customer risk.

## 4.2. Scorecard-guided bandit results

The second-stage simulation applies Thompson Sampling under scorecard-based risk constraints. The cumulative reward increases steadily over time, and the final cumulative reward reaches: 276550 , as shown in Figure 3:

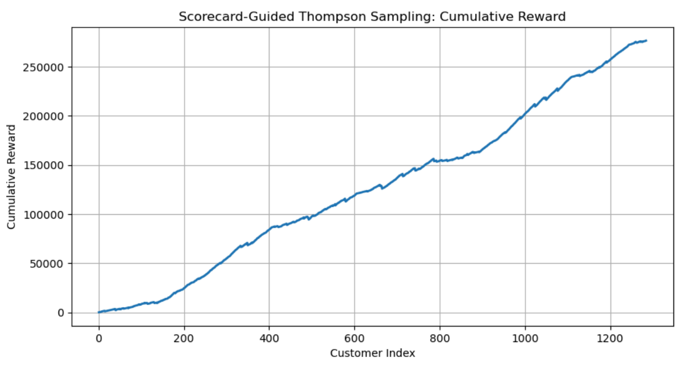


Figure 3. Cumulative reward curve

More importantly, the credit-limit allocation pattern is consistent with practical risk-control logic: high-risk customers only receive the low limit; medium-risk customers mainly receive low or medium limits; low-risk customers are much more likely to receive the high limit.

DOI: 10.54254/2755-2721/2026.AD33345

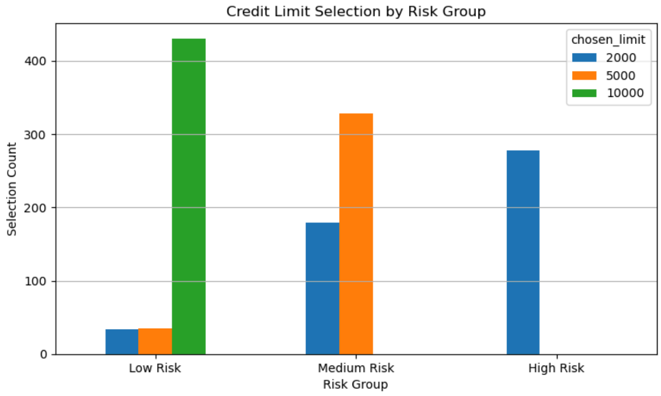


Figure 4. Credit limit selection by risk group

Figure 4 illustrates the major result of this study. The scorecard is no longer used only as a standalone predictive model; instead, it can actively guide the feasible action set of the bandit layer. In this way, the framework creates a meaningful transition from risk prediction to adaptive decision-making.

## 4.3. Limitations

However, there's still exists several limitations.

First, the reward design is simplified and does not represent a full banking profit-and-loss model. In practice, reward would depend on more factors such as utilization, interest income, recovery rates, and regulatory capital constraints.

Second, current bandit framework is rule-based in its use of risk group. The action constraints are imposed by predicted risk thresholds, rather than learned through a contextual policy.

Third, the scorecard itself shows only moderate predictive power. This is sufficient to support the demonstration, but stronger data or richer features might improve both the prediction stage and the downstream decision stage.

## 5. Conclusion

This study finds that the proposed framework has two key advantages. First, it retains the interpretability of traditional scorecards. The first phase remains easy for analysts to understand, as it relies on binning, WOE/IV, logistic regression, and additive scoring. Second, it extends the system beyond static prediction. A simple scorecard can estimate risk, but it does not directly optimize subsequent actions. By adding a Thompson sampling layer, the framework introduces adaptive decision-making while still adhering to a scorecard-based risk structure. Overall, this study demonstrates a transition from pure risk prediction to adaptive credit-limit decisions. The bandit layer does not replace the scorecard but builts a multi-armed bandit layer on top of it.

Future work may expand this framework in several directions. One natural direction is a contextual multi-armed bandit layer, where richer customer attributes can directly influence action choices. Another direction is a more practical reward model based on business profit and loss. Additional external validation and online experiments are required before actual deployment.

DOI: 10.54254/2755-2721/2026.AD33345

## References

[1] Seitshiroa, S., & Govender, N. (2024). Credit risk prediction models with and without weights of evidence using quantitative learning models.

[2] Proceedings of the 4th Asia-Pacific AI & Big Data Forum. (2024) Research on the bank loan access model based on the improved logistic regression scorecard.

[3] Kil, K., Ciukaj, R., & Chrzanowska, J. (2021). Scoring models and credit risk: The case of cooperative banks in Poland.

[4] Jiang et al. (2024), An Evaluation of Interpretable algorithms for Credit Scoring Model

[5] Liang, S., & Tan, L. (2022). Research on the construction of personal credit score model based on WOE analysis and logistics model.

[6] Chen, Y., Calabrese, R., & Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(1), 357–372.

[7] Slivkins, A. (2019). Introduction to Multi-Armed Bandits.

[8] Lattimore, T., & Szepesvári, C. (2020). Bandit Algorithms.

[9] Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika.

[10] Hong, J., Kveton, B., Zaheer, M., Ghavamzadeh, M., & Boutilier, C. (2022). Thompson Sampling with a Mixture Prior. In Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (PMLR 151, pp. 7565–7586).