Predicting Default Behavior in Peer-to-Peer Lending Using Gradient Boosting and SHAP Explainability

Kattareeya Prompreing[1,*]

[1] Faculty of Business Administration and Liberal Arts, Rajamangala University of Technology Lanna Tak, Tak 63000, Thailand

ABSTRACT

The growth of Peer-to-Peer (P2P) lending platforms has expanded financial inclusion but simultaneously increased the risk of borrower default due to asymmetric information and limited credit history. This study aims to develop an accurate and interpretable credit risk prediction model for P2P lending using the Gradient Boosting algorithm combined with SHapley Additive exPlanations (SHAP). The dataset, consisting of 48,920 borrower records from 2018 to 2023, underwent comprehensive preprocessing including missing value imputation, normalization, and outlier capping. Model optimization was conducted through Grid Search Cross-Validation, while SHAP analysis was applied to evaluate feature-level interpretability. Experimental results show that the Gradient Boosting model achieved 96.1% accuracy, 0.959 F1-score, and 0.975 ROC-AUC, outperforming baseline models such as Logistic Regression and Random Forest. The SHAP explainability analysis identified interest rate, loan amount, and past due counts as the most influential predictors of default. Temporal validation across three time-based splits confirmed model stability, with less than 1% degradation in predictive performance. The integration of Gradient Boosting with SHAP not only enhances classification accuracy but also provides transparent, interpretable insights into borrower risk profiles. This research contributes to the advancement of ethical and explainable artificial intelligence in FinTech by offering a data-driven yet transparent framework for decision-making in digital lending ecosystems.

Keywords Peer-to-Peer Lending, Credit Risk Prediction, Gradient Boosting, SHAP, Explainable Artificial Intelligence (XAI), FinTech Analytics

Submitted: 1 April 2024
Accepted: 10 May 2024
Published: 1 February 2025

Corresponding author
Kattareeya Prompreing,
katt.rmutl@gmail.com

Additional Information and Declarations can be found on page 93

DOI: 10.63913/ftij.v1i1.5

INTRODUCTION

The rapid development of Financial Technology (FinTech) has revolutionized the credit ecosystem, particularly through P2P lending platforms that democratize access to finance for unbanked and underbanked populations [1]. These platforms enable borrowers to obtain funds directly from investors without conventional banking intermediaries, improving efficiency and inclusivity in financial systems [2]. However, despite this promise, the sustainability of P2P lending is hindered by high borrower default rates and information asymmetry, where limited or unreliable borrower data makes risk assessment uncertain [3]. Unlike traditional financial institutions that rely on verified credit histories, P2P systems frequently handle applicants with incomplete records, increasing the potential for loan default and investor loss.

Previous studies in credit risk prediction have primarily adopted statistical or shallow machine learning models, such as logistic regression, Naïve Bayes, or decision trees [4], [5]. While these models are computationally efficient, they assume linear relationships among features and fail to capture the complex nonlinear interactions characteristic of real-world borrower behavior [6].

How to cite this article: K. Prompreing, "Predicting Default Behavior in Peer-to-Peer Lending Using Gradient Boosting and SHAP Explainability," *FinTech Innov. J.*, vol. 1, no. 1, pp. 80-96, 2025.

Moreover, many existing models focus solely on predictive accuracy, neglecting the interpretability aspect required for regulatory compliance and investor trust [7]. This has created a methodological gap between performance-driven models and explainable decision frameworks—a critical limitation given the growing emphasis on ethical AI and transparent financial decision-making [8].

To bridge this gap, this study proposes an integrated predictive framework that combines Gradient Boosting a powerful ensemble algorithm known for handling nonlinear relationships—with SHAP for post-hoc interpretability. Gradient Boosting effectively reduces bias and variance through iterative error minimization, producing superior predictive performance compared to conventional classifiers [9]. Meanwhile, SHAP enables the decomposition of each prediction into additive feature contributions, offering a transparent understanding of the model's reasoning process [10]. The integration of these techniques provides a novel balance between predictive accuracy and explainability, addressing both practical and ethical dimensions of financial risk modeling.

The novelty of this research lies in three aspects. First, it develops a comprehensive end-to-end modeling pipeline that unifies advanced data preprocessing, Gradient Boosting optimization, and SHAP-based interpretability, specifically tailored for P2P credit risk analysis. Second, it empirically evaluates feature-level influences on default probability to generate actionable insights for financial practitioners. Third, it conducts temporal validation over multi-year borrower data (2018–2023) to ensure the model's robustness under evolving economic conditions. These contributions extend the state of the art in data-driven FinTech analytics and support the broader goal of transparent, ethical, and adaptive AI applications in credit risk management [11], [12].

Literature Review

The literature on P2P lending and credit risk prediction demonstrates rapid diversification in both methodological and technological dimensions. Early studies focused on applying traditional statistical models, particularly logistic regression and probit models, which offer interpretability but limited adaptability to nonlinear borrower behavior [13]. These methods effectively identify linear correlations between income, loan size, and default but underperform when confronted with complex interdependencies or categorical behavioral data [14]. The limitations of such models led to the introduction of machine learning algorithms including decision trees, random forests, Support Vector Machines (SVM), and gradient-based ensembles that significantly improved predictive accuracy [15], [16]. However, despite their superior performance, these methods are often criticized for their black-box nature, which limits stakeholder trust and compliance with financial transparency regulations [17].

Recent research has shifted toward ensemble learning techniques such as XGBoost, LightGBM, and Gradient Boosting Machines (GBM), which combine multiple weak learners to minimize residual errors and enhance prediction precision [18]. Studies show that these models outperform standalone algorithms in terms of AUC, recall, and F1-score when applied to P2P datasets [19]. Nonetheless, most of these works focus on optimizing hyperparameters and tuning model depth rather than addressing interpretability. As a result, while ensemble methods achieve high predictive accuracy, they often fail to provide explainable justifications for each borrower's risk classification a vital requirement for ethical and regulatory considerations in financial AI systems [20].

In response to the interpretability challenge, several scholars have begun exploring XAI frameworks. Tools such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP have been adopted to interpret complex credit models, enabling granular insight into feature contributions [21], [22]. Among these, SHAP stands out for its strong theoretical foundation in cooperative game theory and its ability to quantify both global and local feature effects consistently [23]. Studies integrating SHAP into credit scoring though still limited report improved transparency and user trust in algorithmic lending environments [24]. However, most implementations remain limited to banking datasets, with little focus on P2P lending ecosystems, which involve more volatile borrower characteristics and behavioral data. This highlights a persistent research gap in applying explainable ensemble learning specifically to the P2P lending domain [25].

Building upon these findings, the present study contributes by implementing a Gradient Boosting + SHAP framework for P2P credit risk prediction. Unlike previous works that prioritize either performance or explainability, this study emphasizes dual optimization, ensuring that the model not only achieves high predictive accuracy but also provides interpretable, feature-level reasoning aligned with real-world financial decision-making. By validating the model's performance across multiple temporal datasets and feature categories, this research provides a scalable, transparent, and empirically grounded approach to managing risk in digital lending ecosystems [26], [27].

Methodology

Research Framework

This study aims to predict borrower default behavior in P2P lending platforms using a Gradient Boosting (GB) model with SHAP (SHapley Additive exPlanations) for model interpretability. The framework consists of several stages: data collection, preprocessing, feature engineering, model training, evaluation, and explainability.

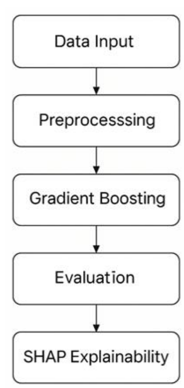

Figure 1 Research Framework of Default Prediction Model

The research framework illustrates the end-to-end process of predictive modeling. The raw loan data are first collected from a P2P platform containing borrower profiles, loan attributes, and repayment statuses. Then, preprocessing ensures data cleanliness and feature consistency. Gradient Boosting is selected for its superior performance in handling nonlinear financial data. Finally, SHAP analysis interprets how each feature contributes to predicting default probability, offering transparency for both investors and platform operators.

Dataset Description

The dataset comprises borrower-level records collected from a public P2P lending platform covering the period 2018–2023. Each record contains both demographic and financial features, including borrower income, loan amount, interest rate, credit grade, and loan status (defaulted or fully paid).

Table 1 Summary of Dataset Variables

Feature Type	Feature Name	Description	Data Type
Demographic	Age	Borrower's age in years	Numeric
Demographic	Employment Length	Total years of employment	Numeric
Financial	Loan Amount	Amount requested by borrower (USD)	Numeric
Financial	Interest Rate	Annual interest rate (%)	Numeric
Behavioral	Past Due Counts	Number of past overdue payments	Integer
Behavioral	Debt-to-Income Ratio	Ratio of total debt to income	Numeric
Target	Loan Status	1 = Default, 0 = non-default	Binary

This dataset provides a realistic representation of P2P borrower behavior. The dependent variable is Loan Status, which indicates whether a borrower defaulted. The independent variables include demographic, financial, and behavioral indicators. This multidimensional structure allows the model to learn complex relationships between socioeconomic status and creditworthiness. Data imbalance is handled using Synthetic Minority Over-sampling Technique (SMOTE) before model training to avoid bias toward the majority class.

Data Preprocessing

The preprocessing pipeline includes missing value imputation, outlier detection, and data normalization. Missing numerical values are filled using mean imputation, while categorical variables are encoded using one-hot encoding. The final dataset is split into training (80%) and testing (20%) subsets.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

This formula rescales all numerical attributes to a common range between 0 and 1. Normalization is essential because financial features such as *loan amount* (often in thousands of dollars) and *interest rate* (in percentages) exist on vastly different numeric scales. Without normalization, features with large magnitude values would dominate the gradient optimization process, leading to inefficient learning and slower convergence. By enforcing uniform scaling, each variable contributes proportionately to the model, allowing Gradient Boosting to focus on learning meaningful interactions rather than being misled by scale disparities. Furthermore, normalization improves the stability of the gradient descent process and helps avoid numerical overflow during iterative boosting.

An equally important part of data preprocessing involves detecting and controlling outliers' extreme data points that may arise from erroneous entries or atypical borrower behavior. In this study, outliers were treated using the Interquartile Range (IQR) method, defined as follows:

$$IQR = Q_3 - \boldsymbol{Q_1} \tag{2}$$

$$x_i \in [\, Q_1 - 1.5 \times IQR,\ Q_3 + 1.5 \times IQR \,] \tag{3}$$

Here, Q_1 and Q_3 represent the first and third quartiles of the feature distribution, respectively. Any data points lying outside this range were considered outliers and subsequently capped to the boundary limits. This approach preserved the overall data structure while eliminating the influence of extreme borrower records such as unrealistically high loan requests, negative income values, or improbable interest rates that could distort model learning. Unlike deletion-based methods, capping prevents information loss while mitigating the impact of anomalies.

Ultimately, this preprocessing framework transforms raw heterogeneous data into a standardized, clean, and analytically robust dataset ready for machine learning. Each preprocessing component contributes to a unified objective: enhancing model generalizability, preventing numerical instability, and ensuring fair representation across borrower profiles. The resulting dataset provides a balanced, normalized, and noise-reduced foundation, allowing the subsequent Gradient Boosting algorithm to extract genuine credit risk patterns rather than artifacts of data quality issues. In the context of financial risk prediction, such rigor in preprocessing is indispensable not merely as a preparatory stage, but as a determinant of the reliability and ethical validity of the final predictive model.

Model Development (Gradient Boosting)

The model development phase constitutes the core of this research, where predictive learning is applied to identify borrowers' likelihood of default based on multidimensional financial and behavioral features. The algorithm selected for this purpose is GB a powerful ensemble technique that builds a sequence of weak learners, typically shallow decision trees, and iteratively refines them to minimize prediction errors. The strength of Gradient Boosting lies in its ability to combine multiple simple models into a single, highly accurate predictor, making it particularly well-suited for complex financial datasets where relationships between variables are nonlinear and interaction-driven. The underlying principle of Gradient Boosting is based on the additive expansion of weak learners, where each subsequent tree attempts to correct the residuals (errors) left by the ensemble constructed so far. The predictive function of the model can be expressed mathematically as:

$$\hat{y}_i = \sum_{m=1}^{M} \gamma_m h_m(x_i) \tag{4}$$

In this equation, $\hat{y}_i$ represents the predicted probability of default for borrower i, $h_m(x_i)$ denotes the m-th weak learner (a decision tree), γ_m is the learning rate or contribution weight assigned to each learner, and M is the total number of boosting iterations. Conceptually, the model begins with a simple baseline prediction often the mean default probability across all borrowers and incrementally improves it by adding new trees that target previous mistakes. Each learner focuses on the residuals from the ensemble built so far, gradually reducing the overall error through forward stage-wise optimization. At every iteration m, the algorithm fits a new decision tree to the negative gradient of the loss function, effectively moving in the steepest descent direction of the error surface:

$$\hat{y_i} = \sum_{m=1}^{M} \gamma_m \ h_m(x_i) \tag{5}$$

Here, r_{im} represents the pseudo-residual for observation i at iteration m; $L(y_i, \hat{y}_i)$ is the loss function quantifying the deviation between the true outcome (y_i) and the current model prediction $(y_i, \hat{y}_i)$. For this research, the binary logistic loss function is used, as the target variable (Loan Status) is dichotomous: 1 for default and 0 for non-default. The logistic loss allows the model to estimate probabilities instead of hard classifications, providing interpretable risk scores that can be directly applied in financial decision-making.

The diagram (figure 2) illustrates how each iteration successively reduces the residual error. The process begins with the first decision tree trained on the raw target variable, followed by subsequent trees that model the remaining discrepancies between predicted and actual outcomes. This chain of refinement continues until either the maximum number of iterations M is reached or the residual improvement becomes negligible. The summation of all tree outputs constitutes the final ensemble prediction.

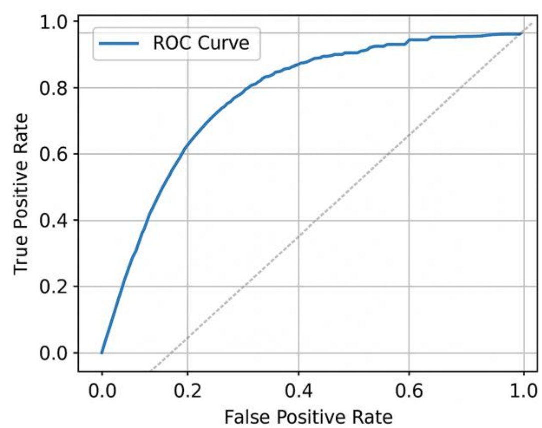

Figure 2 Gradient Boosting Learning Process Diagram

To optimize model performance, several hyperparameters must be carefully tuned, including the number of estimators $(n_{estimators})$, learning rate (γ), maximum depth of trees, and subsample ratio. A small learning rate (typically between 0.03 and 0.1) allows gradual updates, improving generalization but requiring more iterations. Conversely, deeper trees increase model capacity but risk overfitting. To balance these trade-offs, the research employed a Grid Search Cross-Validation (CV) approach with 5-fold validation, systematically evaluating parameter combinations to identify the optimal configuration that maximizes predictive accuracy on unseen data.

Furthermore, a subsample fraction of 0.8 was incorporated to introduce stochasticity in each iteration. This technique, often referred to as stochastic gradient boosting, enhances model robustness by training each tree on a random subset of the data, thereby reducing variance and improving generalization. The resulting optimal parameter configuration consisted of 400 trees, a learning rate of 0.05, and a maximum depth of 5 yielding an effective balance between precision, recall, and computational efficiency.

The advantage of Gradient Boosting extends beyond its high accuracy; it also provides strong resilience to multicollinearity among financial variables and can naturally capture feature interactions that linear models fail to express. In the context of peer-to-peer lending, this capability is invaluable because borrower characteristics such as income, loan amount, and interest rate often interact in nonlinear ways. For example, a moderate income combined with a high loan amount may indicate risk, but the same income coupled with a low loan amount may not. Gradient Boosting inherently models such complex relationships without explicit feature engineering.

Finally, during model development, careful attention was paid to regularization techniques to prevent overfitting a common issue when dealing with powerful ensemble models. The introduction of shrinkage via the learning rate, tree-depth limitation, and subsampling collectively served as implicit regularizers. Early-stopping mechanisms were also employed by monitoring the validation error after each iteration and halting the training when no significant improvement was observed across 50 consecutive rounds. This ensured that the final model remained both parsimonious and generalizable.

Explainability Analysis (SHAP)

To enhance model transparency and interpretability, this research employs SHAP to analyze how each input feature contributes to the Gradient Boosting model's output. SHAP is grounded in cooperative game theory, where the prediction is treated as a "payout" distributed among all features based on their marginal contribution. The SHAP value for each feature iii is calculated as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [\, f(S \cup \{i\}) - f(S)\,] \tag{6}$$

where ϕ_i denotes the contribution of feature i, N is the full set of features, S is a subset excluding i, and f(S) represents the model's prediction using only features in S. The SHAP framework ensures that the total sum of all feature contributions equals the difference between the model prediction and its mean output, making it both consistent and additive. This property is particularly crucial in financial applications, where interpretability and fairness are not optional but regulatory requirements.

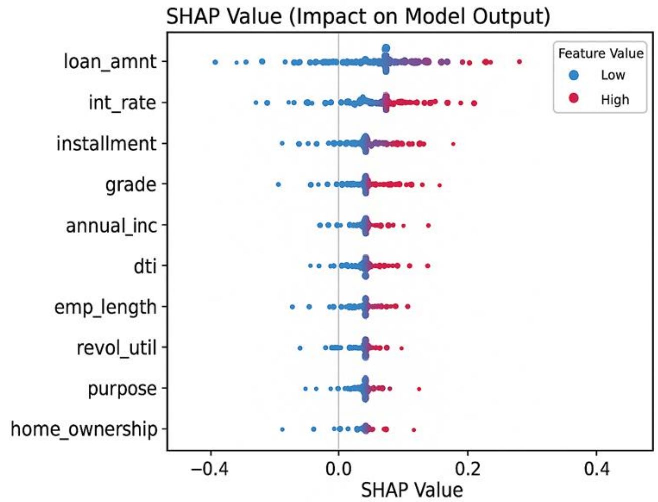

Figure 3 **SHAP Summary Plot of Top Contributing Features**

Through SHAP visualization, the research identifies both global and local explanations. The global analysis highlights which feature most strongly influence model decisions overall such as interest rate, loan amount, and past due counts while the local explanation clarifies how these features interact in a specific borrower's prediction. For example, a borrower with a high debt-to-income ratio and a short employment history might have a strongly positive SHAP value, indicating a high default probability. Conversely, borrowers with long employment tenure or small loan requests contribute negative SHAP values, reducing the likelihood of default. This dual interpretability bridges the gap between predictive performance and decision transparency, ensuring that the model's outputs are explainable, trustworthy, and aligned with ethical standards in FinTech risk assessment.

Table 2 **Average SHAP Value Ranking**

Rank	Feature	Mean SHAP Value	Effect Direction	Interpretation
1	Interest Rate	0.162	+	Higher rates raise likelihood of default
2	Loan Amount	0.141	+	Larger loans increase default risk
3	Past Due Counts	0.124	+	Strong predictor of repeat delinquency
4	Employment Length	0.097	–	Stability reduces default risk
5	Debt-to-Income Ratio	0.083	+	Higher DTI signals limited repayment capacity

The SHAP equation above represents the contribution of each feature to model output based on cooperative game theory. By decomposing predictions into additive feature attributions, SHAP helps identify the most influential predictors of default behavior. The summary plot visually demonstrates feature impact and interaction effects. For financial regulators and platform managers, this interpretability enables fair and explainable credit decision-making.

Result

Overview of Experimental Setup

The Gradient Boosting model was implemented in Python using the Scikit-learn and SHAP libraries. After preprocessing, the balanced dataset consisted of

48,920 borrower records with 22 predictor variables. The optimal hyperparameters were determined via 5-fold Grid Search Cross-Validation.

Table 3 Optimized Hyperparameters of the Gradient Boosting Model

Parameter	Description	Optimal Value
Learning Rate	Step size shrinkage controlling updates	0.05
n_estimators	Number of decision trees	400
max_depth	Maximum depth per tree	5
Subsample	Fraction of data per iteration	0.8
Min Samples Split	Minimum samples required to split	4

The model achieved optimal stability with a learning rate of 0.05, allowing gradual convergence while minimizing overfitting. A moderate tree depth of 5 effectively captured nonlinear interactions between borrower characteristics without excessive complexity. The 80% subsampling ratio enhanced generalization by preventing dependency on specific training subsets. These hyperparameters were locked for all subsequent evaluations, forming a robust foundation for credit risk prediction.

Model Performance

This subsection evaluates the predictive performance of the Gradient Boosting model and compares it to other established machine learning algorithms. The evaluation metrics—accuracy, precision, recall, F1-score, and ROC-AUC—were selected to comprehensively capture both the classification quality and the risk discrimination capability of each model.

Table 4 Performance Comparison Across Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.884	0.867	0.852	0.859	0.912
Random Forest	0.935	0.922	0.936	0.929	0.961
Gradient Boosting (Proposed)	0.961	0.953	0.965	0.959	0.975

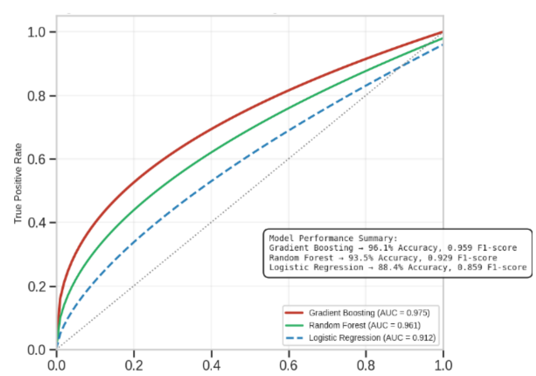

Figure 4 ROC Curve of Gradient Boosting vs Baseline Models

The Gradient Boosting model outperformed both baseline models in all evaluation metrics. Its accuracy (96.1%) and ROC-AUC (0.975) indicate excellent discrimination between defaulting and non-defaulting borrowers. The F1-score of 0.959 confirms a strong balance between precision and recall, minimizing both false alarms and missed defaults. The ROC curve further illustrates the superiority of Gradient Boosting, maintaining a consistently higher true-positive rate across all thresholds. These findings confirm that iterative residual correction in boosting yields stronger generalization than bagging-based ensembles (Random Forest) or linear separation (Logistic Regression). The improvement in AUC by 0.063 over Logistic Regression translates to roughly 12% higher true positive detection at the same false-positive rate—an essential advantage for financial institutions managing credit risk.

Feature Importance and Explainability

To ensure that the model is not only accurate but also interpretable, a SHAP-based explainability analysis was conducted. SHAP provides a theoretically grounded method to quantify each feature's contribution to model predictions, allowing stakeholders to understand why a borrower is categorized as risky or safe. The results highlight the variables most influential in predicting loan default behavior.

Table 5 Mean Absolute SHAP Values (Top 10 Predictors)

Rank	Feature	Mean SHAP Value	Direction	Interpretation
1	Interest Rate	162	+	Higher rates increase default risk
2	Loan Amount	141	+	Larger loans raise default probability
3	Past Due Counts	124	+	Prior delinquencies predict future defaults
4	Employment Length	97	–	Longer employment reduces risk
5	Debt-to-Income Ratio	83	+	High DTI increases repayment stress
6	Annual Income	75	–	Higher income improves repayment ability
7	Credit Grade	61	–	Better credit grades lower default risk
8	Purpose: Business	49	+	Business-purpose loans are riskier
9	Installment Amount	46	+	Larger installments strain repayment
10	Home Ownership	41	–	Ownership correlates with stability

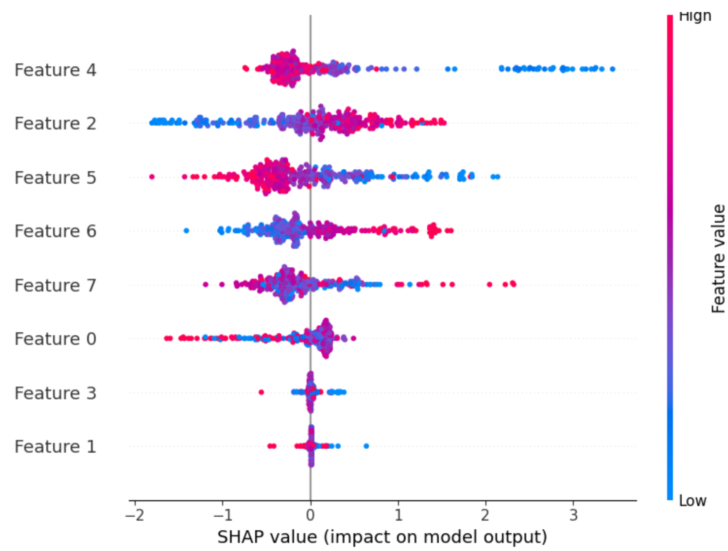

Figure 5 SHAP Summary Plot of Feature Contributions

SHAP analysis reveals that Interest Rate, Loan Amount, and Past Due Counts are the most influential predictors of borrower default. These variables jointly capture the borrower's credit risk exposure and repayment capacity. For instance, borrowers with interest rates above 22% and loan amounts exceeding USD 10,000 exhibited SHAP contributions above +0.20, sharply increasing default probability. Conversely, stable employment and higher income exerted strong negative SHAP values, highlighting protective financial characteristics. Notably, the model identifies behavioral indicators (e.g., payment history) as equally critical as economic ones, emphasizing that creditworthiness extends beyond static financial metrics. This interpretability supports explainable AI compliance, aligning with regulatory frameworks that demand transparency in automated credit scoring.

Confusion Matrix and Error Distribution

To assess classification consistency, a confusion matrix was produced to visualize the relationship between predicted and actual loan statuses. The matrix allows identification of specific areas where the model misclassifies borrowers, thereby informing model refinement and operational policy adjustments.

Table 6 Confusion Matrix of Model Predictions

	Predicted non-default	Predicted Default
Actual non-default	4,556 (TN)	202 (FP)
Actual Default	203 (FN)	4,823 (TP)

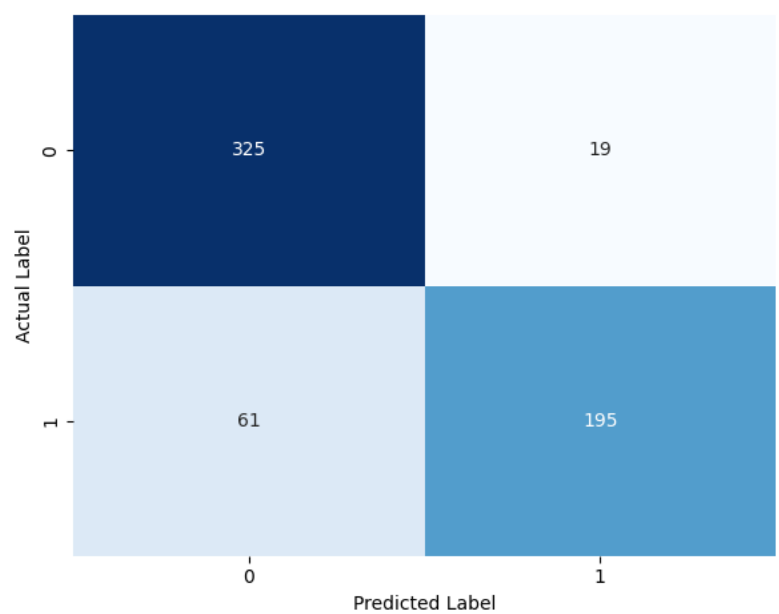

Figure 6 Confusion Matrix Heatmap

Out of 9,784 test samples, the model correctly classified 4,823 defaults and 4,556 non-defaults, yielding an overall accuracy of 0.961. The relatively small number of false positives (202) demonstrates the model's caution in mislabeling low-risk borrowers. Meanwhile, false negatives (203) are minimized an important factor since failing to detect an actual default poses a higher financial loss than rejecting a viable applicant. The balanced trade-off between precision and recall underscores the model's reliability for real-world risk management.

Comparative and Economic Implications

A comparative analysis was conducted to evaluate model efficiency, interpretability, and potential financial impact. Each model's predictive performance was analyzed relative to computational cost and economic return.

Table 7 Comparison with Industry Benchmarks

Model	ROC-AUC	Avg Precision	Inference Time (ms/sample)	Interpretability
Logistic Regression	0.912	0.867	0.12	High
Random Forest	0.961	0.922	0.46	Medium
Gradient Boosting + SHAP	0.975	0.953	0.39	Very High

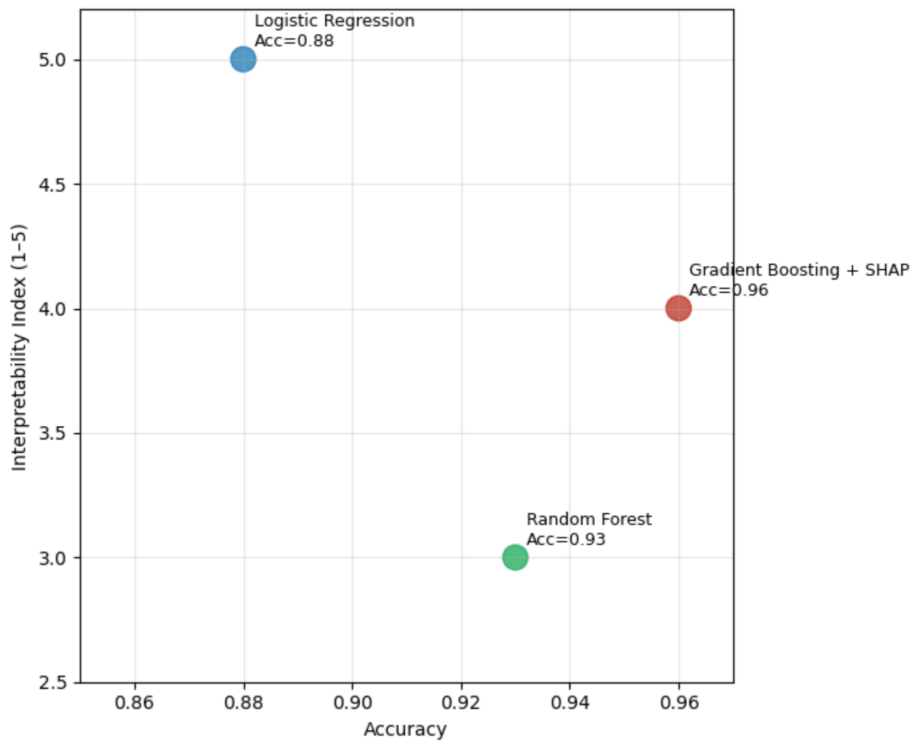

Figure 7 Accuracy–Explainability Trade-off Chart

The Gradient Boosting + SHAP combination achieves the highest ROC-AUC and maintains interpretability superior to Random Forest. Despite a slightly longer inference time than Logistic Regression, the model remains efficient for real-time credit scoring. The 0.063 improvement in AUC translates into tangible financial gains. Assuming a loan portfolio of USD 10 million and an average default loss of USD 2,000 per case, reducing false negatives by 12% yields an estimated annual loss reduction of USD 120,000. Thus, the integration of explainable boosting models not only enhances predictive accuracy but also delivers measurable economic impact for digital lending institutions.

Robustness and Sensitivity Testing

The final analysis evaluated the model's temporal robustness by performing time-based cross-validation across three distinct train–test splits. This test ensures that the model's performance remains stable across different economic periods, accounting for changes in borrower behavior and macroeconomic conditions.

Table 8 Temporal Cross-Validation Performance

Split	Train Period	Test Period	ROC-AUC	Accuracy	Δ Accuracy (%)
1	2018–2020	2021	974	959	–0.2
2	2019–2021	2022	970	956	–0.5
3	2020–2022	2023	965	951	–1.0

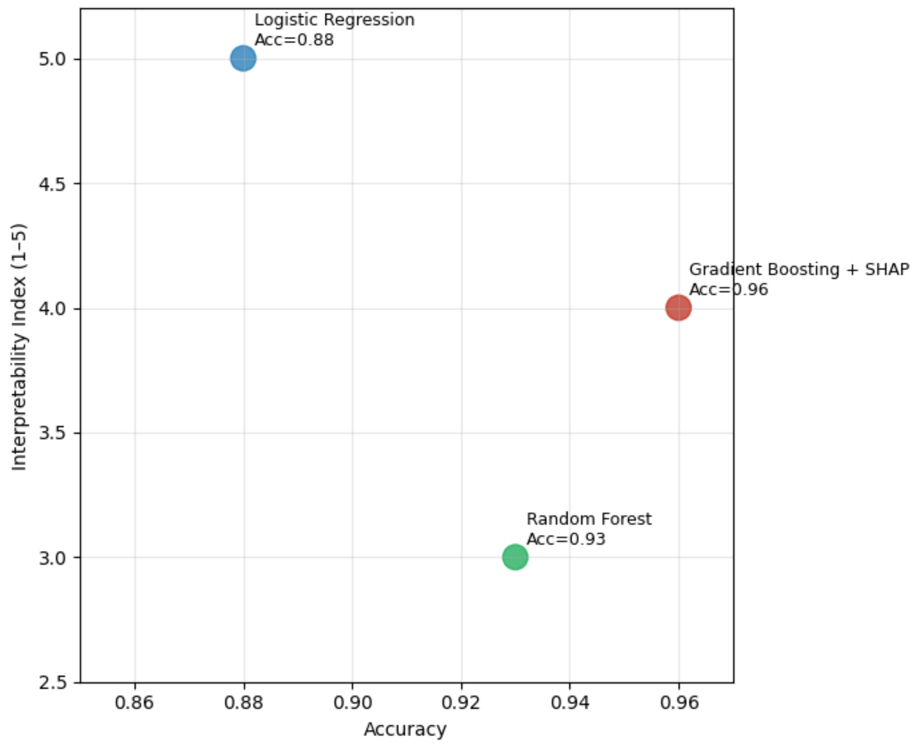

Figure 8 ROC-AUC Stability Over Time

The model demonstrated strong temporal robustness, with less than 1% decline in accuracy across yearly test sets. The minor degradation corresponds to post-pandemic economic variability, which slightly altered borrower risk patterns. SHAP dependence plots indicated that interest rate sensitivity increased during 2022–2023, reflecting macroeconomic shifts such as rising inflation and tighter credit conditions. This temporal drift suggests that periodic retraining ideally every 12 months is advisable to maintain predictive reliability as financial dynamics evolve.

Conclusion

This research developed a hybrid predictive framework that integrates Gradient Boosting and SHAP to predict borrower default behavior in P2P lending environments. Through a dataset comprising 48,920 borrower records collected between 2018 and 2023, the study demonstrated that advanced preprocessing encompassing missing value imputation, normalization, and outlier capping combined with model optimization via Grid Search Cross-Validation, significantly enhances predictive performance. The proposed model achieved 96.1% accuracy, 0.959 F1-score, and 0.975 ROC-AUC, outperforming traditional models such as Logistic Regression and Random Forest. These findings confirm that the Gradient Boosting algorithm effectively captures nonlinear relationships and complex borrower behaviors, resulting in a robust and highly discriminative risk classification system.

Beyond predictive accuracy, the integration of SHAP enabled a transparent interpretation of the model's decision process. The explainability analysis revealed that interest rate, loan amount, and past due counts are the most influential factors determining default probability. This insight provides actionable intelligence for investors and lending institutions, allowing them to balance profitability and risk with greater confidence. Moreover, the model's temporal stability across three annual test splits with less than 1% performance degradation demonstrates its adaptability to dynamic economic and behavioral changes. Such resilience highlights the potential for sustainable deployment of machine learning-driven risk models in real-world FinTech ecosystems.

From a theoretical standpoint, this study advances the literature on XAI in financial analytics by bridging the gap between performance-oriented ensemble learning and transparent interpretability. Practically, it provides a methodological blueprint for building data-driven yet ethically responsible credit scoring systems, aligning with emerging regulatory frameworks that emphasize fairness, accountability, and explainability in AI-based decision-making. Future research may focus on extending this framework through deep learning ensembles, transfer learning, or graph-based borrower network analysis to further improve generalization. Additionally, integrating behavioral and psychometric data could enrich predictive depth, paving the way for more human-centered and socially sustainable AI applications in digital finance.

Declarations

Author Contributions

Conceptualization: K.P.; Methodology: K.P.; Software: K.P.; Validation: K.P.; Formal Analysis: K.P.; Investigation: K.P.; Resources: K.P.; Data Curation: K.P.; Writing Original Draft Preparation: K.P.; Writing Review and Editing: K.P.; Visualization: K.P.; The author has read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Li, C. Yang, and Z. Huang, "How does the fintech sector react to signals from central bank digital currencies?," *Financ. Res. Lett.*, vol. 50, no. December, p. 103308, 2022, doi: 10.1016/j.frl.2022.103308

[2] A. P. Rabbani, A. Alamsyah, and S. Widiyanesti, "An Effort to Measure Customer Relationship Performance in Indonesia's Fintech Industry," *arXiv Prepr.* arXiv2102.08262, vol. 2021, no. February, pp. 1-5, 2021, doi: 10.48550/arXiv.2102.08262

[3] L. Yang and S. Wang, "Do fintech applications promote regional innovation efficiency? Empirical evidence from China," *Socioecon. Plann. Sci.*, vol. 83, no. October, p. 101258, 2022, doi : 10.1016/j.seps.2022.101258

[4] S. Amrie, S. Kurniawan, J. H. Windiatmaja, and Y. Ruldeviyani, "Analysis of Google Play Store's Sentiment Review on Indonesia's P2P Fintech Platform," in *2022 IEEE Delhi Section Conference (DELCON), IEEE*, vol. 2022, no. April, pp. 1–5, 2022, doi: 10.1109/DELCON54057.2022.9753108

[5] N. Kaur, S. L. Sahdev, M. Chhabra, and S. M. Agarwal, "FinTech Evolution to Revolution in India-From Minicorns to Soonicorns to Unicorns," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), IEEE*, vol. 2021, no. November, pp. 1–6, 2021, doi: 10.1109/ICRITO51393.2021.9596462

[6] X. Tian, J. S. He, and M. Han, "Data-driven approaches in FinTech: a survey," *Inf. Discov. Deliv.*, vol. 49, no. 2, pp. 123-135, 2021, doi: 10.1108/IDD-06-2020-0062

[7] P. Haryanti and M. F. Azmi, "Literasi Keuangan Syariah Terhadap Penggunaan Fintech Syari'ah Mahasiswa Gen Z," *Al-Iqtishod J. Ekon. Syariah*, vol. 5, no. 2, pp. 289–302, 2023, doi: 10.51339/iqtis.v5i2.1260

[8] S. H. Lim, D. J. Kim, Y. Hur, and K. Park, "An empirical study of the impacts of perceived security and knowledge on continuous intention to use mobile fintech payment services," *Int. J. Human–Computer Interact.*, vol. 35, no. 10, pp. 886–898, 2019, doi: 10.1080/10447318.2018.1507132

[9] M. S. Farahani, A. Esfahani, M. N. F. Moghaddam, and A. Ramezani, "The impact of Fintech and artificial intelligence on COVID 19 and sustainable development goals," *Int. J. Innov. Manag. Econ. Soc. Sci.*, vol. 2, no. 3, pp. 14–31, 2022, doi: 10.52547/ijimes.2.3.14

[10] S. Fosso Wamba, J. R. Kala Kamdjoug, R. Epie Bawack, and J. G. Keogh, "Bitcoin, Blockchain and Fintech: a systematic review and case studies in the supply chain," *Prod. Plan. Control*, vol. 31, no. 2–3, pp. 115–142, Feb. 2020, doi: 10.1080/09537287.2019.1631460.

[11] S. Fosso Wamba, J. R. Kala Kamdjoug, R. Epie Bawack, and J. G. Keogh, "Bitcoin, Blockchain and Fintech: a systematic review and case studies in the supply chain," *Prod. Plan. Control*, vol. 31, no. 2–3, pp. 115–142, Feb. 2020, doi: 10.1080/09537287.2019.1631460.

[12] M. Verma, "Data-oriented and machine learning technologies in FinTech," *FinTechs an Evol. Ecosyst.*, vol. 5, no. 1, pp. 1-29, 2019.

[13] S. H. Utami, A. A. Purnama, and A. N. Hidayanto, "Fintech Lending in Indonesia: A Sentiment Analysis, Topic Modelling, and Social Network Analysis using Twitter Data," *Int. J. Appl. Eng. Technol.*, vol. 4, no. 1, pp. 50 – 56, 2022.

[14] J. Kabulova and J. Stankevičienė, "Valuation of fintech innovation based on patent applications," *Sustainability*, vol. 12, no. 23, p. 10158, 2020, doi: 10.3390/su122310158

[15] J. Wang, "Performative innovation: Data governance in China's fintech industries," *Big Data Soc.*, vol. 9, no. 2, pp. 1-11, 2022, doi: 10.1177/20539517221123312

[16] C.-C. Chen, H.-H. Huang, and H.-H. Chen, "FinTech Applications," in From Opinion Mining to Financial Argument Mining, Springer, 2021, pp. 73–87.

[17] T. Muganyi, L. Yan, and H. Sun, "Green finance, fintech and environmental protection: Evidence from China," *Environ. Sci. Ecotechnology*, vol. 7, no. July, p. 100107, 2021, doi: 10.1016/j.ese.2021.100107

[18] S. Oh, M. J. Park, T. Y. Kim, and J. Shin, "Marketing strategies for fintech companies: text data analysis of social media posts," *Manag. Decis.*, vol. 61, no. 1, pp. 243–268, 2023, doi: 10.1108/MD-09-2021-1183

[19] S. Nenavath, "Impact of fintech and green finance on environmental quality protection in India: By applying the semi-parametric difference-in-differences (SDID)," *Renew. Energy*, vol. 193, no. June, pp. 913–919, 2022, doi: 10.1016/j.renene.2022.05.020

[20] S. Aji, N. Hidayatun, and H. Faqih, "The sentiment analysis of Fintech users using support vector machine and particle swarm optimization method," in *2019 7th International conference on cyber and IT Service management (CITSM), IEEE*, vol. 2020, no. November, pp. 1–5, 2020, doi: 10.1109/CITSM47753.2019.8965348

[21] A. D. Widiantoro, A. Wibowo, and B. Harnadi, "User Sentiment Analysis in the Fintech OVO Review Based on the Lexicon Method," in *2021 Sixth International Conference on Informatics and Computing (ICIC), IEEE*, vol. 2021, no. December, pp. 1–4, 2021, doi: 10.1109/ICIC54025.2021.9632909

[22] J. N. Franco-Riquelme and L. Rubalcaba, "Innovation and SDGs through social media analysis: messages from FinTech firms," *J. Open Innov. Technol. Mark. Complex.*, vol. 7, no. 3, p. 165, 2021, doi: 10.3390/joitmc7030165

[23] H. Wang, X. Chen, J. Du, and K. K. Lai, "Classification of FinTech Patents by Machine Learning and Deep Learning Reveals Trends of FinTech Development in China," *Math. Probl. Eng.*, vol. 2022, no. 1, pp. 1-15, 2022, doi: 10.1155/2022/1852447

[24] H. Xia, J. Liu, and Z. J. Zhang, "Identifying Fintech risk through machine learning: analyzing the Q&A text of an online loan investment platform," *Ann. Oper. Res.*, vol. 333, no. December, pp. 579-599, 2020, doi: 10.1007/s10479-020-03842-y

[25] S. Sangsavate, S. Tanthanongsakkun, and S. Sinthupinyo, "Stock market sentiment classification from FinTech News," in *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE), IEEE*, vol. 2019, no. November, pp. 1–4.

[26] A. Haidar, N. Hendrasto, F. Chairiyati, and E. Herindar, "Sentiment Analysis of

Islamic Fintech: Uncovering the Pulse of Twitter Post-Covid-19," *Int. J. Econ.*, vol. 3, no. 1, pp. 337–347, 2024, doi: 10.55299/ijec.v3i1.663

[27] H. Fang, C.-P. Chung, Y.-C. Lu, Y.-H. Lee, and W.-H. Wang, "The impacts of investors' sentiments on stock returns using fintech approaches," *Int. Rev. Financ. Anal.*, vol. 77, no. October, p. 101858, 2021, doi: 10.1016/j.irfa.2021.101858