Issues in Information Systems
Volume 26, Issue 3, pp. 166-182, 2025

DOI: https://doi.org/10.48009/3_iis_2025_2025_114

Using machine learning to support peer-to-peer loan funding decision support

Mousumi Munmun, *Metropolitan State University, mousumi.munmun@metrostate.edu*
Queen E. Booker, *Metropolitan State University, queen.booker@metrostate.edu*
Simon Jin, *Metropolitan State University, simon.jin@metrostate.edu*

Abstract

This study investigates the application of artificial intelligence (AI), specifically machine learning (ML) techniques, in enhancing decision-making processes on peer-to-peer (P2P) lending platforms. Traditional financial institutions use AI in their decision-making processes. However, the use of AI in supporting investor decisions in decentralized P2P environments remains limited. This research addresses a critical inefficiency in the P2P lending process, where creditworthy loan applications approved by platforms often go unfunded by investors, by developing a second-tier decision support model. Using historical data from the Prosper platform, multiple ML models, including Logistic Regression, Support Vector Machines, Neural Networks, and Random Forests, were evaluated for their ability to predict unfundable loans. Findings indicate that AI-powered models can significantly improve funding decision accuracy and reduce investor burden. Random Forest demonstrated the highest suitability among the models assessed across multiple evaluation metrics. These results highlight the potential of AI to enhance the efficiency and reliability of alternative lending ecosystems.

Keywords: artificial intelligence, machine learning, financial technology, peer-to-peer lending

Introduction

Artificial Intelligence (AI) has changed the landscape of financial decision-making, introducing unprecedented capabilities in automation, pattern recognition, and predictive analytics. One of the most significant applications of AI in financial technology (FinTech) is within the peer-to-peer (P2P) lending sector, where machine learning a subset of AI is used to enhance decision-making at multiple stages of the lending process. P2P lending represents an alternative financial model that facilitates direct connections between individual borrowers and investors via online platforms, bypassing traditional financial intermediaries such as banks (Arner et al., 2015; Boot et al., 2021). These platforms assess the creditworthiness of loan applicants and determine which applications advance for potential funding by investors. Consequently, the lending process involves two critical decision points: the first by the platform, which evaluates creditworthiness, and the second by investors, who ultimately decide whether to fund the loan.

Each platform employs proprietary algorithms to assess credit risk and select loan applications for investor consideration. However, not all approved loan applications by the platform receive funding. This disconnect can frustrate borrowers whose applications are approved but ultimately unfunded. Additionally, a surplus of unfundable yet approved applications burdens investors, who must navigate numerous listings to identify desirable opportunities (Li et al., 2016; Liu et al., 2021; Yan et al., 2018). This study explores the potential of implementing a second-tier decision support system to filter further loan applications presented to investors. Specifically, the research investigates whether machine learning algorithms can accurately predict which creditworthy loans are unlikely to be funded, thereby streamlining the investor's decision-making process.

This study evaluates various machine learning models to determine their effectiveness in enhancing the efficiency of the P2P loan funding process. The primary research questions guiding this study are:

1. *To what extent can machine learning algorithms predict unfundable loans during the second decision-making stage?*
2. *Which model offers the highest predictive accuracy and reliability?*

A key contribution of this research lies in its focus on P2P loan applications approved by platforms but ultimately unfunded. By identifying Common features of these unsuccessful applications, this study seeks to develop a model capable of reducing investor burden and improving overall platform efficiency. The remainder of this paper is structured as follows: a review of the relevant literature on machine learning applications in loan decision support systems; a description of the research methodology; presentation and discussion of results; limitations of the study; and directions for future research.

Literature Review

This study evaluates the application of artificial intelligence, specifically machine learning (ML) models, in predicting unfundable loans during the second stage of peer-to-peer (P2P) lending—namely, the investor decision-making phase following platform-level creditworthiness approval. Although AI has been widely adopted across various domains in financial services, its usage in supporting investor decisions in P2P lending remains underexplored. To better understand the capabilities of AI in this context, it is essential to examine the machine learning algorithms that serve as the foundation for predictive decision-making. The following section reviews widely used models in P2P lending, such as Logistic Regression, Decision Trees, Support Vector Machines, and Neural Networks, assessing their predictive performance, suitability, and limitations.

Machine Learning Models for Loan Prediction

Machine learning, a subfield of artificial intelligence grounded in computational learning theory, employs data-driven algorithms to identify patterns and make predictions (Gollapudi, 2016). ML models adaptively adjust parameters based on training data to improve predictive performance. The two dominant ML paradigms are supervised and unsupervised learning (Mahesh, 2020). Supervised learning approaches, including Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVM), and Neural Networks (NN), have been extensively applied in loan approval research. Unsupervised techniques, such as K-means clustering, are less commonly used but have been seen experimental adoption in financial analytics (James et al., 2021; Sinap, 2024). While much of ML-based loan prediction research has focused on traditional banking systems, studies examining P2P lending remain comparatively limited (Munmun, 2023).

Performance of Prominent ML Models for Loan Approval

Logistic Regression (LR)

Logistic regression is widely employed in binary classification tasks like loan approval and rejection. Studies report high accuracy rates for LR models (78.5%–91%) in predicting loan outcomes (Dosalwar et al., 2021; Rath et al., 2021; Shukla et al., 2021). However, LR performance diminishes in cases involving unbalanced or sparse datasets, high-dimensional feature spaces, and missing values (Li et al., 2017; Khan et al., 2021). Hybrid variants incorporating iterative computation or ensemble techniques have improved performance (Zeng et al., 2017; Wei et al., 2018).

Neural Networks (NN)

Neural networks, particularly deep neural networks (DNNs), have shown promising results in capturing complex, non-linear relationships in loan data. Comparative studies indicate that NNs can outperform LR, SVM, and DT under certain conditions (Wu et al., 2019; Abakarim et al., 2018). Nevertheless, their effectiveness is contingent upon large, well-labeled datasets and substantial computational resources (Zhang et al., 2019; Karthiban et al., 2019). Recent evaluations present mixed findings, with some researchers ranking NN above SVM and DT (Lusinga et al., 2021; Pimcharee & Surinta, 2022), while others report inferior performance (Chen et al., 2021; Peiris, 2022).

Support Vector Machines (SVM)

SVMs are particularly effective for small and moderately complex datasets, especially in binary classification problems. Reported accuracies range from 62% to 98% (Li et al., 2017; Abedin et al., 2019). SVM's use of kernel functions facilitates performance with semi-structured data, a typical feature of P2P lending platforms (Zhao, 2022; Thomas et al., 2022). However, SVM performance is highly sensitive to hyperparameter tuning, and it has been ranked inconsistently across comparative studies (Sam et al., 2021; Orji et al., 2022; Dabas et. al., 2024).

Decision Trees (DT) and Random Forest (RF)

Decision trees - more specifically ensemble variants such as Random Forests are among the most extensively studied models in loan prediction tasks. RF generally outperforms single DT models, with accuracy rates reported between 60% and 99% (Kelen & Emanuel, 2019; Hamayel et al., 2021). Ensemble methods such as gradient boosting further enhance performance, achieving accuracy levels as high as 98% in some studies (Karthiban et al., 2019; Kumar et al., 2019). Nonetheless, DT-based models may suffer from overfitting and lack of generalization capacity compared to more robust ensemble or neural architectures (Shukla et al., 2021; Naik & Manerkar, 2022).

Table 1 shows the summary of the most frequently used models used in loan approval prediction.

Table 1: Summary of The Models Used in Loan Approval Prediction

ML Model	Accuracy Range	Strengths	Weaknesses
Logistic Regression (LR)	78.5% – 91%	Strong for binary classification, easy to implement	Struggles with small/unbalanced data, missing values
Neural Networks (NN)	67% – 98%	Handles non-linear patterns, adaptive learning	Requires large datasets, high computational cost
Support Vector Machines (SVM)	62% – 98%	Works well with small datasets, strong pattern recognition	Sensitive to hyperparameter tuning
Decision Trees (DT) & Random Forest (RF)	60% – 99%	Simple, interpretable, strong ensemble models	Prone to overfitting, lower performance in small datasets

Based on prior research, no single ML model consistently outperforms others in all scenarios. LR remains a popular choice, while RF provides higher accuracy in ensemble-based approaches. SVM is effective for small datasets, and NN shows promise in complex, large-scale applications. The results of the literature review support the gap in current research regarding using ML to predict loan fundability in the P2P industry while also indicating there are no clear "leaders" to apply to this decision support step.

Research Methodology

The research questions were:

1. *How effectively can machine learning models predict unfundable loans in the second stage of predictive modeling? and,*
2. *Which model demonstrates the highest accuracy and reliability?*

Following the method used by Ilyas & Chu (2019), data was collected, cleaned, and prepared for analysis. Preparation for analysis included running the descriptive statistics to validate no outliers, and appropriate means and variances for the independent variables for logistic regression. Each ML was evaluated using suitability measures, considering Accuracy, Recall, Precision, F-value, confusion matrix, and Area Under the ROC Curve (AUC) (Hossin & Sulaiman, 2015). Each predictive model was validated using the entire dataset. T-tests were used to compare the results of each model to the others.

The population for this study is all the United States (US)-based loan applications that had been funded by the Prosper lending platform between 2018 and 2022. Individual loan data were collected from internal secondary data on the SEC's EDGAR to examine the relationship between lenders' loan funding decisions and loan characteristics. This data incorporated both loan listings and sales reports for Prosper. As stated by Lo (2015), SEC collects loan data from the P2P lending industry The SEC is authorized to collect data from various organizations, including those in the P2P industry, enhancing the validity and reliability of the loan data used in this study. (The role of the SEC, 2023). The downloaded data had 105,666 instances of loans approved by the platform, and whether the loan was funded or not by investors on the platform. The data set was distinctive from traditional loan approval data in that the number of funded loans exceeded the number of unfunded loans.

Data Description

Each instance included the loan's characteristics and the borrower's credit profile. The original dataset had a total of 28 independent variables. A definition for the dependent variable and the independent variables is as follows:

- **Funded**: Investor invested in the loan. (dependent variable)
- **Duration**: How long it takes to get funding.
- **Loan ID**: The number that uniquely identifies the listing to the public as displayed on the website.
- **Term**: The length of the loan expressed in months.
- **Borrower APR**: The Borrower's Annual Percentage Rate (APR) for the loan.
- **Borrower Rate**: The Borrower's interest rate for this loan.
- **Lender Yield**: The Lender yield on the loan. Lender yield is equal to the interest rate on the loan less the servicing fee.
- **Borrower State**: The state of the address of the borrower at the time the Listing was created.
- Occupation: The Occupation selected by the Borrower at the time they created the listing.

- **Employment Status**: The employment status of the borrower at the time they posted the listing.
- **Employment Status Duration**: The length in months of the employment status at the time the listing was created
- **Homeowner**: A Borrower will be classified as a homeowner if they have a mortgage on their credit profile or provide documentation confirming they are a homeowner.
- **Prosper Score:** A custom risk score was built using historical Prosper data to assess the risk of Prosper borrower listings.
- **Prosper Ratings:** Every loan application is assigned a Prosper Rating by Prospers's proprietary system that allows to maintain consistency in their evaluation. Prosper Ratings allow potential investors to easily consider a loan application's level of risk because the rating represents an estimated average annualized loss rate range to the investor.
- **FICO score**: A person's credit score calculated with software from Fair Isaac Corporation (FICO).
- **First Recorded Credit Line**: The date the first credit line was opened.
- **Current Credit Lines:** Number of current credit lines at the time the credit profile was pulled.
- **Open Credit Lines:** Number of open credit lines at the time the credit profile was pulled.
- **Total Credit Lines past 7 years:** Number of credit lines in the past seven years at the time the credit profile was pulled.
- **Open Revolving Accounts:** Number of open revolving accounts at the time the credit profile was pulled.
- **Open Revolving Monthly Payment:** Monthly payment on revolving accounts at the time the credit profile was pulled.
- **Inquiries Last 6 Months Number of inquiries**: in the past six months at the time the credit profile was pulled.
- **Total Inquiries**: Total number of inquiries at the time the credit profile was pulled.
- **Current Delinquencies:** Number of accounts delinquent at the time the credit profile was pulled.
- **Amount Delinquent:** Dollars delinquent at the time the credit profile was pulled.
- **Delinquencies Last 7 Years:** Number of delinquencies in the past 7 years at the time the credit profile was pulled.
- **Public Records Last 10 Years**: Number of public records in the past 10 years at the time the credit profile was pulled.
- **Public Records Last 12 Months**: Number of public records in the past 12 months at the time the credit profile was pulled.
- **Revolving Credit Balance:** The amount of revolving credit at the time the credit profile was pulled.
- **Bankcard Utilization**: The percentage of available revolving credit that is utilized at the time the credit profile was pulled.
- **Debt To Income Ratio:** The debt-to-income ratio of the borrower at the time the credit profile was pulled. This value is Null if the debt-to-income ratio is not available. This value is capped at 10.01 (any debt-to-income ratio larger than 1000% will be returned as 1001%).
- **Income Range**: The income range of the borrower at the time the listing was created.
- **Stated Monthly Income:** The monthly income the borrower stated at the time the listing was created.

Data Preparation

To prepare the dataset for analysis, descriptive statistics were performed to identify missing data, potential outliers, and skewness. 1,206 records were removed due to missing values and outliers leaving104,460 instances for analysis of which 87,086 (83%) of loans were funded, and 17,374 (17%) of loans were unfunded. One variable from the original dataset – Public Record in Last Twelve Months had 99% missing data and was dropped from the dataset. Other missing values were handled by an impute linear method using R software (Li et al., 2014). Dummy variables were created for categorical variables with at least one value used as the control value. Numeric variables that were not normally distributed were transformed using log n, square root, and cube root as appropriate. After data cleaning and transforming the independent variables, all the numeric continuous independent variables were normally distributed. To ensure the data was normalized, each variable had a skewness score below $\pm$0.5 and a kurtosis score below $\pm$2.

This study employed a widely used sampling strategy for data segmentation (Jaafari et al., 2019), randomly selecting 70% of loan acceptance observations for training and reserving the remaining 30% for testing. The entire original dataset was used for validation. The dependent variable "Funded and Unfunded" was not equally represented: 87,086 (83%) of data was Funded, and 17,374 (17%) of data was Unfunded. SMOTE was the selected method for balancing the funded/unfunded data set for compatibility with a wide range of MLs and recommended (Chawla et al., 2002; Fernández et al., 2018), creating a 50:50 Funded and Unfunded ratio for analysis for training the ML models.

This study uses the three-fold validation resampling, which is the recommended strategy by Valavanis & Kosmopoulos (2010) as a commonly used process for machine learning modeling. *Three-fold cross-validation* is iterative; fitting and evaluation are performed with each of the three subsets (Valavanis & Kosmopoulos, 2010).

Model Evaluation Measurement

This study measures Accuracy, Sensitivity, Classification Error, Precision, F measure, and AUC. In addition, this study also performed descriptive data analysis, inferential data analysis, and assumption testing. The descriptive data analysis included measures of frequency, measures of central tendency, measures of dispersion or variation, and measures of position. Inferential data analysis was carried out after the descriptive analysis. Inferential data analysis included regression data analysis, ANOVA (Analysis of Variance), one tail and two-tail t-tests. A one-tail and two-tail t-test were performed to test the hypotheses of this study. According to the rules, the lower the p-value, the more substantial the evidence that the null hypothesis is false (Shaffer, 1995). This study adhered to the 0.01 p-value threshold, considered highly statistically significant, as the basis for determining whether to accept or reject the research hypotheses.

Results

The results present the actual funded/unfunded number versus predicted funded/unfunded ratio, model performance evaluation, and t-test to determine suitability. Each model is measured at three stages of the process and at three different thresholds: 0.7, 0.6, and 0.5. The stages are model training, model testing, and model validating. Table 2 shows training results for actual observation versus each model's prediction at different thresholds. The actual funded number is set at 60,960, and the actual unfunded number is also set at 60,960 for a 50:50 ratio.

Table 2 shows that all the models performed best when classifying a loan as potentially "unfunded" when the threshold was set at .5. Results for the other thresholds were low for unfunded loan prediction. The standalone models performed poorly for training data at .7 and .6 thresholds.

Table 2: Actual Versus Predicted Results for Training Data

Training Data								
			Threshold					
			0.7		0.6		0.5	
Model	Actual Funded	Actual Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded
LR	60,960 (50%)	60,960 (50%)	88%	12%	73%	27%	53%	47%
SVM	60,960 (50%)	60,960 (50%)	94%	6%	73%	27%	52%	48%
ANN	60,960 (50%)	60,960 (50%)	82%	18%	72%	28%	66%	34%
RF	60,960 (50%)	60,960 (50%)	60%	40%	57%	43%	53%	47%

Table 3 shows the testing model result for actual observation versus each model's prediction at different thresholds. The actual funded observation is 26,126, and the unfunded observation is also set at 26,126 for a 50:50 ratio. In the testing stage, all the models performed best when classifying a loan as potentially "unfunded" when the threshold was set at 0.5. LR (12% to 47%), SVM (6% to 48%) improved significantly when the threshold changed from 0.7 to 0.5. RF (40%) performance does not change when the threshold changes from 0.7 to 0.6; it increases by 1% from 0.6 to 0.5.

Table 3: Actual Versus Predicted Results for Testing Data

Testing Data								
			Threshold					
			0.7		0.6		0.5	
Model	Actual Funded	Actual Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded
LR	26,126 (50%)	26,126 (50%)	88%	12%	73%	27%	53%	47%
SVM	26,126 (50%)	26,126 (50%)	94%	6%	73%	27%	52%	48%
ANN	26,126 (50%)	26,126 (50%)	82%	18%	72%	28%	66%	34%
RF	26,126 (50%)	26,126 (50%)	60%	40%	60%	40%	59%	41%

Table 4 shows the validating model result for actual observation versus each model's prediction at different thresholds. The actual funded observation is 87,086, and the unfunded observation is also set at 87,086 for a 50:50 ratio. In the validating stage, all the models performed best when classifying a loan as potentially "unfunded" when the threshold was set at 0.5. LR (12% to 47%), SVM (6% to 48%) improved significantly when the threshold changed from 0.7 to 0.5.

Table 4: Actual Versus Predicted Results for Validation Data

Validation Data								
			Threshold					
			0.7		0.6		0.5	
Model	Actual Funded	Actual Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded
LR	87,086	87,086	88%	12%	73%	27%	53%	47%
SVM	87,086	87,086	94%	6%	73%	27%	52%	48%
ANN	87,086	87,086	82%	18%	72%	28%	66%	34%
RF	87,086	87,086	60%	40%	58 %	42%	55%	45%

Table 5 shows the model result for the original, unbalanced dataset values versus each model's prediction at different thresholds for original data. The actual funded observation is 87,086, and the unfunded observation is 17,374. In the original data validating stage, LR, SVM models performed best when classifying a loan as potentially "unfunded" when the threshold was set at 0.6. LR (8% to 20%), SVM (4% to 20%) improved significantly when the threshold changed. RF model performed best when classifying a loan as potentially "unfunded" when the threshold was set at 0.5. RF (1% to 8%) improved when the threshold changed. ANN model performed best when classifying a loan as potentially "unfunded" when the threshold was set at 0.7. ANN (49% to 24%) improved when the threshold changed.

Table 5: Actual Versus Predicted Results for Original Data

Original Data								
			Threshold					
			0.7		0.6		0.5	
Model	Actual Funded	Actual Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded	Predicted Funded	Predicted Unfunded
LR	87,086 (83%)	17,374 (17%)	96,142 (92%)	8,318 (8%)	83,744 (80%)	20,716 (20%)	64,322 (61%)	40,138 (39%)
SVM	87,086 (83%)	17,375 (17%)	1,00,672 (96%)	3,788 (4%)	83,286 (80%)	21,174 (20%)	62,368 (60%)	42,092 (40%)
ANN	87,086 (83%)	17,376 (17%)	79,783 (76%)	24,677 (24%)	66,015 (63%)	38,445 (37%)	53,071 (51%)	51,389 (49%)
RF	87,086 (83%)	17,377 (17%)	1,03,816 (99%)	644 (1%)	1,01,383 (97%)	3,077 (3%)	95,634 (92%)	8,826 (8%)

Table 6 shows the Type 1 and Type 2 Errors from the model predictions using the original data. All the models had the fewest Type I Errors when the threshold is set at 0.5. All the models had the fewest Type II Errors when the threshold is set at 0.7.

Table 6: Summary Results for Original Data Type I and Type II Error Performance

Models	Threshold	T I E	T II E
LR	0.5	0.40964	0.34312
LR	0.6	0.62461	0.16299
LR	0.7	0.81409	0.05843
SVM	0.5	0.41280	0.36619
SVM	0.6	0.63820	0.17096
SVM	0.7	0.91269	0.02608
ANN	0.5	0.32750	0.45593
ANN	0.6	0.44561	0.33086
ANN	0.7	0.60556	0.20467
RF	0.5	0.50207	0.00201
RF	0.6	0.82480	0.00038
RF	0.7	0.96311	0.00003

Prediction performance measures for loan funded and unfunded decision-making is tested by using Accuracy (Acc), Recall, Precision (Pre), F-measure (F mes), Specificity (Spe), Type I Error (T I E), Type II Error (T II E), and AUC. For the training model data, the LR, SVM, and ANN models performed best across the criteria of Accuracy, Recall, F-measure, AUC, and Type I Error values when the threshold is set at 0.5. The rest of the criteria, Precision, Specificity, and the lowest Type II Error values, the LR, SVM, and ANN models performed best when the threshold is set at 0.6. For the training model data, the RF model performed best across all eight criteria when the threshold was set at 0.5. Table 7 shows the training model data results.

Table 7: Summary Results for Training Data Model Performance

Models		Prediction Performances							
Model	Threshold	Acc	Pre	Recall	Spe	F mes	AUC	T I E	T II E
LR	0.5	0.62940	0.63730	0.60061	0.65819	0.61841	0.67841	0.19970	0.17091
LR	0.6	0.60390	0.69350	0.37239	0.83542	0.48458	0.65701	0.31380	0.08229
LR	0.7	0.55504	0.73756	0.17090	0.93919	0.27750	0.63821	0.41455	0.03041
SVM	0.5	0.62483	0.62897	0.60875	0.64090	0.61870	0.68145	0.19562	0.17955
SVM	0.6	0.61122	0.70317	0.38493	0.83751	0.49751	0.65374	0.30753	0.08125
SVM	0.7	0.53187	0.78579	0.08762	0.97611	0.15766	0.58135	0.45619	0.01194
ANN	0.5	0.59862	0.64532	0.43796	0.75928	0.52179	0.64323	0.28102	0.12036
ANN	0.6	0.58842	0.65957	0.36547	0.81137	0.47033	0.63203	0.31727	0.09432
ANN	0.7	0.56452	0.67951	0.24423	0.88481	0.35931	0.61356	0.37789	0.05760
RF	0.5	0.97211	1.00000	0.94423	1.00000	0.97131	0.99999	0.02789	0.00000
RF	0.6	0.92676	1.00000	0.85353	1.00000	0.92098	0.99999	0.07324	0.00000
RF	0.7	0.90448	1.00000	0.80896	1.00000	0.89439	0.99999	0.09552	0.00000

For the testing model data, the LR, SVM, ANN, and RF models performed best across the criteria of Accuracy, Recall, F-measure, AUC, and Type I Error values when the threshold is set at 0.5. Rest of the criteria, Precision, Specificity, and the lowest Type II Error values, the LR, SVM, ANN, and RF models performed best when the threshold is set at 0.6. Table 8 shows the testing model data results.

Table 8: Summary Results for Testing Data Model Performance

Models		Prediction Performances							
Model	*Threshold*	Acc	Pre	Recall	Spe	F mes	AUC	T I E	T II E
LR	0.5	0.63075	0.63831	0.60342	0.65808	0.62038	0.67874	0.19829	0.17096
LR	0.6	0.60541	0.69610	0.37419	0.83664	0.48673	0.65856	0.31291	0.08168
LR	0.7	0.55780	0.74529	0.17561	0.93998	0.28425	0.63874	0.41219	0.03001
SVM	0.5	0.61456	0.61894	0.59615	0.63297	0.60733	0.66001	0.20193	0.18351
SVM	0.6	0.60000	0.68388	0.37190	0.82809	0.48180	0.62978	0.31405	0.08596
SVM	0.7	0.53066	0.76897	0.08765	0.97367	0.15737	0.60001	0.45617	0.01317
ANN	0.5	0.59722	0.64246	0.43845	0.75599	0.52120	0.64336	0.28077	0.12200
ANN	0.6	0.58719	0.65674	0.36534	0.80904	0.46950	0.63387	0.31733	0.09548
ANN	0.7	0.56507	0.67955	0.24627	0.88387	0.36152	0.61396	0.37687	0.05806
RF	0.5	0.90354	0.99207	0.81360	0.99349	0.89401	0.94416	0.09320	0.00325
RF	0.6	0.90123	0.99857	0.80361	0.99885	0.89054	0.94087	0.09820	0.00057
RF	0.7	0.89646	0.99981	0.79308	0.99985	0.88453	0.93570	0.10346	0.00008

For the validation model performances, Table 9 shows LR prediction performance as the best with Acc (0.62980), Recall (0.60145), F-mes (0.61900), AUC (0.67851), and T I E (0.19927) at threshold 0.5. At threshold 0.7, LR prediction performance is the best with Pre (0.73990), Spe (0.93943), and T II E (0.03029). For the LR model, threshold 0.5 is chosen over the .6 and .7 thresholds because five categories are the highest scoring versus three for threshold 0.7. The SVM prediction performance is the best with Acc (0.61559), Recall (0.59741), F-mes (0.60847), AUC (0.66217), and T I E (0.20130) at threshold 0.5. At threshold 0.7, the SVM prediction performance is the best with Pre (0.77062), Spe (0.97391), and T II E (0.01304). For the SVM model, threshold 0.5 is chosen over .6 and .7 thresholds because five categories are highest scoring versus three categories for threshold 0.7.

ANN prediction performance is the best with Acc (0.59820), Recall (0.43811), F-mes (0.52161), and T I E (0.28095) at threshold 0.5. At threshold 0.7, ANN prediction performance is the best with Pre (0.67952), Spe (0.88453), and T II E (0.05774). For the ANN model, threshold 0.5 is chosen over the .6 and .7 thresholds because five categories are the highest scoring versus three categories for threshold 0.7. RF prediction performance is the best with Acc (0.95154),

Recall (0.90504), F-mes (0.94918), AUC (0.99119), and T I E (0.04748) at threshold 0.5. At threshold 0.7, RF prediction performance is the best with Pre (0.99994), Spe (0.99995), and T II E (0.00002). For the RF model, threshold 0.5 is chosen over the .6 and .7 thresholds because five categories are the highest scoring versus three categories for threshold 0.7.

Table 9: Summary Results for Validating Data Model Performance

Models		Prediction Performances							
Model	***Threshold***	**Acc**	**Pre**	**Recall**	**Spe**	**F mes**	**AUC**	**T I E**	**T II E**
LR	0.5	0.62980	0.63761	0.60145	0.65815	0.61900	0.67851	0.19927	0.17092
LR	0.6	0.60436	0.69428	0.37293	0.83578	0.48522	0.66578	0.31353	0.08211
LR	0.7	0.55587	0.73990	0.17231	0.93943	0.27953	0.63256	0.41384	0.03029
SVM	0.5	0.61559	0.61995	0.59741	0.63376	0.60847	0.66217	0.20130	0.18312
SVM	0.6	0.60112	0.68582	0.37321	0.82903	0.48337	0.66148	0.31340	0.08548
SVM	0.7	0.53078	0.77062	0.08765	0.97391	0.15740	0.63461	0.45618	0.01304
ANN	0.5	0.59820	0.64445	0.43811	0.75830	0.52161	0.64327	0.28095	0.12085
ANN	0.6	0.58805	0.65872	0.36543	0.81067	0.47008	0.64675	0.31728	0.09467
ANN	0.7	0.56468	0.67952	0.24484	0.88453	0.35997	0.63589	0.37758	0.05774
RF	0.5	0.95154	0.99785	0.90504	0.99805	0.94918	0.99119	0.04748	0.00098
RF	0.6	0.91910	0.99959	0.83855	0.99966	0.91202	0.94789	0.08072	0.00017
RF	0.7	0.90207	0.99994	0.80419	0.99995	0.89145	0.94876	0.09790	0.00002

For the LR model, threshold 0.5 is chosen over the .6 and .7 thresholds because five categories are the highest scoring versus three categories for threshold 0.7. Usually, individual models optimize at different thresholds, although 0.5 provides the most optimization across models. Table 10 shows the best model based on the optimum threshold.

Table 10: Best Model Based on Optimum Threshold

Models		Prediction Performances									
Model	Best Threshold (most suitable)	Acc	Data _PV L	Pre_ PVL	Pre	Recall	Spe	F-mes	AUC	T I E	T II E
LR	0.5	0.63	0.50	0.53	0.64	0.60	0.66	0.62	0.68	0.20	0.17
RF	0.5	0.95	0.50	0.55	0.998	0.91	0.99	0.95	0.99	0.05	0.00098
SVM	0.5	0.62	0.50	0.52	0.62	0.60	0.63	0.61	0.66	0.20	0.18
ANN	0.5	0.60	0.50	0.66	0.64	0.44	0.76	0.52	0.64	0.28	0.12

Models Suitability Evaluation

As mentioned earlier, based on the literature suitability measures were developed. For this study, MLa suitability requires the following: Accuracy score over 90%, Specificity score over 85%, Type I Error score under 10%, Type II Error score under 10%, Recall score over 85%, Precision score over 85%, F measure score over 85%, and an AUC near to 1 (Chen et al., 2021; Demraoui et al., 2022; Karthiban et al., 2019; Lusinga et al., 2021; Li et al., 2017; Ndayisenga, 2021; Orji et al., 2022; Pimcharee & Surinta, 2022). See Table 11 for details. Based on this study's model evaluation measurements, LR, SVM, and ANN are not suitable models, while RF is a suitable model for funded and unfunded loan decision-making.

Table 11: Models Suitability Evaluation

	Model Performances								Suitable Model?
Best Models	Acc ≥ 0.90	Pre≥ 0.85	Recall≥ 0.85	Spe≥ 0.85	F mes≥ 0.85	AUC near to 1	T I E≤ 0.10	T II E≤ 0.10	
LR	0.62980	0.63761	0.60145	0.65815	0.61900	0.67851	0.19927	0.17092	No
RF	0.95154	0.99785	0.90504	0.99805	0.94918	0.99119	0.04748	0.00098	Yes
SVM	0.61559	0.61995	0.59741	0.63376	0.60847	0.66217	0.20130	0.18312	No
ANN	0.59820	0.64445	0.43811	0.75830	0.52161	0.64327	0.28095	0.12085	No

Discussion

This study explored the use of machine learning (ML) to support investor-level decision-making within peer-to-peer (P2P) lending platforms, specifically, predicting whether a loan will be funded after receiving platform-level approval. This focus represents a shift from conventional research on creditworthiness assessment to the less-explored domain of investor behavior modeling. The results show that ensemble models, particularly Random Forest (RF), significantly outperform other techniques in identifying unfundable loans. These findings highlight important issues of model divergence and identity in decision support systems and contribute to the growing literature on FinTech and artificial intelligence in financial services.

Model Performance

Among the models tested, RF consistently delivered the highest prediction accuracy (95.15%), recall (90.50%), and AUC (0.991), with error rates well below acceptable thresholds. This performance supports earlier findings by Zeng et al. (2017) and Chen et al. (2021), who demonstrated the robustness of ensemble methods in loan classification and credit approval tasks. RF's ability to manage nonlinearity and variance through bagging and tree aggregation makes it especially effective in handling imbalanced and multidimensional data—common characteristics in investor behavior datasets. y contrast, the lower performance of Logistic Regression (LR), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) represent a divergence from studies like Wu et al. (2019) and Lusinga et al. (2021), which reported competitive performance from these models in loan approval prediction. This divergence underscores the importance of application context: models that perform well in platform-level approval may not generalize to the investor funding stage. For example, ANN performance in this study may have been constrained by training data limitations or insufficient model depth, factors less prevalent in platform-side datasets but more critical in investor decision modeling.

Investor Identity and Second-Tier Decision Support

Investor behavior in P2P lending is heterogeneous, shaped by individual risk preferences, financial goals, and interpretations of borrower attributes. This study proposes that a second-tier ML system, trained on historical funding outcomes, can serve as a proxy for investor identity patterns that are otherwise difficult to model explicitly. By aligning model predictions with observed funding behavior, the system implicitly captures shared investor preferences without requiring explicit investor profiling. This approach builds on conceptual insights from Yan et al. (2018), who emphasized the need for systems that bridge platform evaluation criteria and investor trust formation. The high performance of the RF model suggests that ML can identify recurring traits of loans that, while creditworthy, are unlikely to attract investor support. These include potentially "soft" features such as borrower occupation, income range, or regional lending trends—variables that investors may weigh differently than platform algorithms. By modeling this identity-driven divergence, the proposed system helps reconcile platform and investor evaluations.

Practical Implications for P2P Lending Platforms

The integration of ML-based decision support at the investor stage offers several benefits for P2P platforms:

- **Improved efficiency**: Reducing the volume of unfundable loan listings streamlines investor navigation and shortens decision cycles.
- **Enhanced user experience**: Investors receive more targeted recommendations, and borrowers face less frustration from platform-approved but investor-rejected listings.
- **Stronger platform credibility**: By aligning platform approval with funding likelihood, platforms may improve investor trust and borrower satisfaction.

For practitioners, these findings indicate that deploying an ensemble-based prediction system can reduce manual screening tasks and improve the overall funding ratio—key metrics for platform growth and sustainability.

Contributions to Information Systems Literature

This study contributes to the growing body of information systems research on intelligent financial technologies by introducing a second-tier predictive framework that models investor behavior, not just borrower creditworthiness. It also illustrates how data balancing techniques such as SMOTE can improve model training on real-world class-imbalanced datasets. These contributions extend existing work on AI-driven decision support systems by focusing on decentralized financial ecosystems and investor behavior modeling. It also contributes to the growing body of literature on AI in FinTech by addressing a relatively underexplored area: the second-stage decision-making process in P2P lending. While prior studies have focused primarily on loan approval prediction at the platform level, this study shifts the focus to investor behavior and the likelihood of funding. The findings align with and extend previous work (e.g., Zeng et al., 2017; Chen et al., 2021) by demonstrating that ensemble models like Random Forest not only perform well in traditional credit scoring but also in decentralized lending platforms.

Limitations

Although SMOTE was effective for addressing data imbalance, future studies should explore advanced resampling methods such as Borderline-SMOTE, ADASYN, or GAN-based oversampling to improve realism in synthetic data generation. Furthermore, while this study evaluated four widely used ML models, expanding the model set to include XGBoost, LightGBM, and explainable AI tools (e.g., SHAP, LIME) could improve interpretability and adoption in real-world applications.he dataset used in this study was sourced exclusively from the Prosper platform. While Prosper is one of the major players in the P2P lending space, its borrower demographics, platform policies, and investor behaviors may not be representative of the broader P2P industry. Future studies should consider multi-platform datasets to improve validity and generalizability.

Conclusion and Future Research

This study addresses predicting whether a loan will be funded after platform-level approval using machine learning. These findings indicate that ML can serve as a valuable tool for investors by helping them identify loans with a higher likelihood of funding and improving the overall efficiency of P2P lending platforms. The study contributes to the literature by addressing the P2P lending from the investor-level funding perspective rather than the platform-level loan approval. It also highlights the importance of model selection and data balancing techniques in achieving reliable predictions in imbalanced datasets.

There are several areas for future research. First, expanding the dataset to include additional P2P platforms would enhance the validity of the findings and allow for cross-platform comparisons. Next, the exploration of more advanced data balancing methods could improve model robustness and generalizability. Finally, the inclusion of additional ML models may yield further performance gains. By addressing these areas, future research can build on the foundation of this study to develop more sophisticated, scalable, and interpretable decision support systems for the evolving landscape of decentralized finance.

References

Abakarim, Y., Lahby, M., & Attioui, A. (2018). Towards An Efficient Real-time Approach to Loan Credit Approval Using Deep Learning. *2018 9th International Symposium on Signal, Image, Video, and Communications (ISIVC)*. https://doi.org/10.1109/isivc.2018.8709173

Abedin, M. Z., Guo-Tai, C., & Moula, F. E. (2019). Weighted SMOTE-Ensemble Algorithms: Evidence from Chinese Imbalance Credit Approval Instances. *2019 2nd International Conference on Data Intelligence and Security (ICDIS)*. https://doi.org/10.1109/icdis.2019.00038

Arner, D. W., Barberis, J., & Buckley, R. P. (2015). The Evolution of Fintech: A New Post-Crisis Paradigm? *Social Science Research Network*. https://doi.org/10.2139/ssrn.2676553

Boot, A. W. A., Hoffmann, P., Laeven, L., & Ratnovski, L. (2021). Fintech: What's Old, What's New? *Journal of Financial Stability*, *53*, 100836 https://doi.org/10.1016/j.jfs.2020.100836

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Chen, Z., Lin, G., & Jin, X. (2021). Credit Approval Prediction Based on Boruta-GBM Model. *2021 7th International Conference on Systems and Informatics (ICSAI)*. https://doi.org/10.1109/icsai53574.2021.9664026

Dabas, M., Manikanta, A., Kumar, P. S., Reddy, K. A., & Aravind, K. (2024). Loan Approval Prediction using Machine Learning. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *8*(12), 1–7. https://doi.org/10.55041/IJSREM39586

Demraoui, L., Eddamiri, S., & Hachad, L. (2022). Digital Transformation and Costumers Services in Emerging Countries: Loan Prediction Modeling in Modern Banking Transactions. *Lecture Notes on Data Engineering and Communications Technologies*, 627–642. https://doi.org/10.1007/978-3-030-90618-4_32

Dosalwar, S., Kinkar, K., Sannat, R., & Pise, N. (2021). Analysis of Loan Availability Using Machine Learning Techniques. *International Journal of Advanced Research in Science, Communication and Technology*, 15–20. https://doi.org/10.48175/ijarsct-1895

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863-905.

Gollapudi, S. (2016). *Practical Machine Learning*. Packt Publishing Ltd.

Hamayel, M., Mohsen, M. a. A., & Moreb, M. (2021). Improvement Of Personal Loans Granting Methods in Banks Using Machine Learning Methods and Approaches in Palestine. *International Conference in Information Technology*. https://doi.org/10.1109/icit52682.2021.9491636

Hossin, M., & Sulaiman, M. R. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Ilyas, I. F., & Chu, X. (2019). *Data cleaning*. Morgan & Claypool. http://dx.doi.org/10.1145/3310205

Jaafari, A., Panahi, M., Pham, B. T., Shahabi, H., Bui, D. T., Rezaie, F., & Lee, S. (2019). Meta Optimization of an Adaptive Neuro-Fuzzy Inference System with Grey Wolf Optimizer And Biogeography-Based Optimization Algorithms For Spatial Prediction of Landslide Susceptibility. *Catena, 175,* 430-445.

James, G. M., Witten, D., Hastie, T., & Tibshirani, R. (2021). Unsupervised Learning. *Springer Texts in Statistics*, 497–552. https://doi.org/10.1007/978-1-0716-1418-1_12

Karthiban, R., Ambika, M., & Kannammal, K. E. (2019). A Review on Machine Learning Classification Technique for Bank Loan Approval. *International Conference on Computer Communication and Informatics*. https://doi.org/10.1109/iccci.2019.8822014

Kelen, Y. P. K., & Emanuel, A. W. R. (2019). Comparison of Classification Methods using Historical Loan Application Data. *International Conference on Information Technology*. https://doi.org/10.1109/icitisee48480.2019.9003786

Khan, A., Bhadola, E., Kumar, A., & Singh, N. (2021). Loan Approval Prediction Model a Comparative Analysis. *Advances and Applications in Mathematical Sciences*, *20*(3). https://www.mililink.com/upload/article/1759044670aams_vol_203_january_2020_a10_p427-435_afrah_khan_and_nidhi_singh.pdf

Kumar, R., Jain, V., Sharma, P. S., Awasthi, S., & Jha, G. (2019). Prediction of Loan Approval Using Machine Learning. *International Journal of Advanced Science and Technology*, *28*(7), 455-460.

Li, J., Hsu, S., Chen, Z., & Chen, Y. (2016). Risks of P2P Lending Platforms in China: Modeling Failure Using a Cox Hazard Model. *Chinese Economy*, *49*(3), 161–172. https://doi.org/10.1080/10971475.2016.1159904

Li, Y., Li, Z., & Li, L. (2014). Missing traffic data: Comparison of imputation methods. *IET Intelligent Transport Systems*, *8*(1), 51-57. https://doi.org/10.1049/iet-its.2013.0052

Li, Z., Tian, Y., Li, K., Zhou, F., & Yang, W. (2017). Reject Inference in Credit Scoring Using Semi-Supervised Support Vector Machines. *Expert Systems with Applications*, *74*, 105–114. https://doi.org/10.1016/j.eswa.2017.01.011

Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised Learning: Generative or Contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 1. https://doi.org/10.1109/tkde.2021.3090866

Lo, B. (2015). It Ain't Broke: The Case for Continued SEC Regulation of P2P Lending. *Harv. Bus. L. Rev. Online*, *6*, 87. https://rb.gy/18nwm

Lusinga, M., Mokoena, T., Modupe, A., & Mariate, V. (2021). Investigating Statistical and Machine Learning Techniques to Improve the Credit Approval Process in Developing Countries. *AFRICON*. https://doi.org/10.1109/africon51333.2021.9570906

Mahesh, B. (2020). Machine Learning Algorithms- a Review. *International Journal of Science and Research (IJSR).[Internet]*, *9*, 381-386. https://t.ly/Scmc

Munmun, M. (2023). *Heterogeneous ensemble learning and loan funding decision support for P2P lending in the USA* (Publication No. 30691775.) [Doctoral dissertation, Metro State University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/2908963653

Naik, S., & Manerkar, G (2022). Education Loan Prediction Analysis. *International Journal of Innovative Science and Research Technology*, *7(4)*, 258-262. https://ijisrt.com/assets/upload/files/IJISRT22APR259.pdf

Ndayisenga, T. (2021). *Bank Loan Approval Prediction Using Machine Learning Techniques* [Doctoral dissertation, University of Rwanda]. http://www.dr.ur.ac.rw/handle/123456789/1437

Peiris, M. P. C. (2022). *Credit Card Approval Prediction by Using Machine Learning Techniques* [Doctoral dissertation, University of Colombo School of Computing]. https://dl.ucsc.cmb.ac.lk/jspui/handle/123456789/4593

Pimcharee, K., & Surinta, O. (2022). Data Mining Approaches in Personal Loan Approval. *Engineering Access*, *8*(1), 15-21. doi: 10.14456/mijet.2022.2. https://ph02.tci-thaijo.org/index.php/mijet/article/view/244392

Orji, U. E., Ugwuishiwu, C. H., Nguemaleu, J. C. N., & Ugwuanyi, P. O. (2022). Machine Learning Models for Predicting Bank Loan Eligibility. *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*. https://doi.org/10.1109/nigercon54645.2022.9803172

Rath, G. B., Das, D., & Acharya, B. (2021). Modern Approach for Loan Sanctioning in Banks Using Machine Learning. *Springer eBooks*, 179–188. https://doi.org/10.1007/978-981-15-5243-4_15

Sam, D., Suresh, K. C., Kanya, N., Tamilselvi, C., & Tejasria, M. V. S. L. (2021) An Improved Bank Customer Churn and Loan Prediction Model Using Supervised Machine Learning Approach. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*, 3.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*(1), 561-584. https://doi.org/10.1146/annurev.ps.46.020195.003021

Shukla, S., Maheshwari, A., & Johri, P. (2021). Comparative Analysis of Ml Algorithms & Stream Lit Web Application. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. https://doi.org/10.1109/icac3n53548.2021.9725496

Sinap, V. (2024). A Comparative Study of Loan Approval Prediction Using Machine Learning Methods. *Gazi Üniversitesi Fen Bilimleri Dergisi*, *12*(2), 644–663. https://doi.org/10.29109/gujsc.1455978

The role of the SEC. (2023). Investor.gov. https://www.investor.gov/introduction-investing/investing-basics/role-sec

Thomas, T. C., Sridhar, J. P., Chandrashekar, M. J., Upadhyaya, M., & Aurelia, S. (2022). Developing a Website for a Bank's Machine Learning-Based Loan Prediction System. *International Journal of Mechanical Engineering, 7(3)*, 475-483. https://kalaharijournals.com/resources/56_MARCH%20ISSUE.pdf

Valavanis, I., & Kosmopoulos, D. (2010). Multiclass defect detection and classification in weld radiographic images using geometric and texture features. *Expert Systems with Applications, 37*(12), 7606-7614. http://dx.doi.org/10.1016/j.eswa.2010.04.082

Wei, X., Gotoh, J., & Uryasev, S. (2018). Peer-To-Peer Lending: Classification in the Loan Application Process. *Risks*, *6*(4), 129. https://doi.org/10.3390/risks6040129

Wu, M., Huang, Y., & Duan, J. (2019). Investigations on Classification Methods for Loan Application Based on Machine Learning. *International Conference on Machine Learning and Cybernetics*. https://doi.org/10.1109/icmlc48188.2019.8949252

Yan, Y., Lv, Z., & Hu, B. (2018). Building Investor Trust in the P2P Lending Platform with a Focus on Chinese P2P Lending Platforms. *Electronic Commerce Research*, *18*, 203-224. http://dx.doi.org/10.1007/s10660-017-9255-x

Zeng, X., Liu, L., Leung, S. C. H., Du, J., Wang, X., & Li, T. (2017). A Decision Support Model for Investment on P2P Lending Platform. *PLOS ONE*, *12*(9), e0184242. https://doi.org/10.1371/journal.pone.0184242

Zhang, C., Zhang, H., & Hu, X. (2019). A Contrastive Study of Machine Learning on Funding Evaluation Prediction. *IEEE Access*. https://doi.org/10.1109/access.2019.2927517

Zhao, Y. (2022). Credit Card Approval Predictions Using Logistic Regression, Linear SVM and Naïve Bayes Classifier. *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*. https://doi.org/10.1109/mlke55170.2022.00047