Predicting Default Risk on Peer-to-Peer Lending Platform

1. Introduction

Peer-to-peer (P2P) lending is an alternative finance model that allows borrowers to connect directly with individual investors through online platforms, bypassing traditional banking intermediaries. This approach has expanded access to credit, particularly for individuals who struggle with conventional loan requirements, while attracting investors with the promise of competitive returns. However, with the profit potential comes the risk of loan defaults, making it crucial to understand the factors influencing whether a loan will be fully repaid. Key factor, such as a borrower's income, credit score, loan purpose, and term length, can significantly affect default risk, but the relationships between these variables and default probability are complex.

Using and assessing various predictive modeling techniques on a P2P lending dataset, Our analysis aims to provide P2P investors with actionable insights by identifying and quantifying the most significant predictors of loan default. By understanding these relationships, investors can develop more robust risk assessment frameworks and make more informed decisions in this rapidly evolving market, further offer practical value for participants in the P2P lending ecosystem.

2. Data Section

Dataset Overview

The P2P lending dataset under analysis comprises 1,044,488 loan records with 114 variables from 2015 to 2018, providing a comprehensive view of lending activities and borrower characteristics. The data encompasses crucial loan information such as loan amounts, interest rates, and term lengths. Each loan record includes detailed borrower information including employment titles, credit metrics, and financial indicators such as credit limits and debt settlements.

Data Pre-processing

Our data preparation process focused on establishing a clean and analytically robust dataset for default risk prediction. Columns with over 20% missing values were removed for data reliability. We first filtered the dataset to include only loans with definitive outcomes: "Fully Paid," "Charged Off," or "Default." We then created a binary classification variable where both "Charged Off" and "Default" statuses were coded as 1 (representing default), and "Fully Paid" was coded as 0. The overall default rate in the dataset was 21.12%, indicating that about one in five loans resulted in a default. Categorical variables, including income levels, FICO scores, home ownership status, and loan purpose, were transformed into dummy variables suitable for regression analysis. We also categorized continuous variables such as income and debt-to-income ratio (DTI) to better capture their nonlinear relationships with default risk. These preprocessing steps resulted in a refined dataset containing key borrower characteristics and loan features essential for default risk prediction.

3. Method section

To understand what factors affect default risk in peer-to-peer (P2P) lending, we used three main types of regression models:

- **Ordinary Least Squares (OLS) regression**
- **Logistic regression**
- **Regularization methods (Ridge and Lasso regression)**

Ordinary Least Squares (OLS) Regression

We started with OLS regression to explore the relationships between different borrower and loan characteristics and the likelihood of default. In this model, the outcome variable was binary: 1 for loans that defaulted and 0 for those that were fully repaid. While OLS is usually used for continuous outcomes, it provides an initial insight into which predictors are important and their influence on default probability. OLS assumes that relationships are linear (predictors have a straightforward, consistent effect) and that the spread of errors is constant.

Logistic Regression

Since default risk is a yes-or-no outcome, we also used logistic regression, which is better suited for binary results. Logistic regression calculates the probability that a loan will default, producing an outcome between 0 and 1 that represents this likelihood. Unlike OLS, it doesn't require the same assumptions about linearity, making it a good fit for our data.

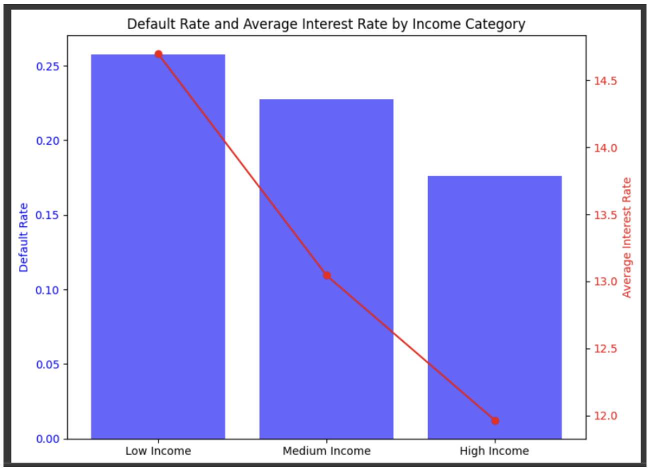
Ridge and Lasso Regression

We applied Ridge and Lasso regression to avoid overfitting and handle any overlap between predictors. These regularization methods add penalties to control the influence of predictors, helping keep the model simpler and more accurate. Ridge regression reduces the size of all predictor coefficients, while Lasso regression can shrink some predictors to zero, leaving only the most important ones.

4. Exploratory Data Analysis (EDA):

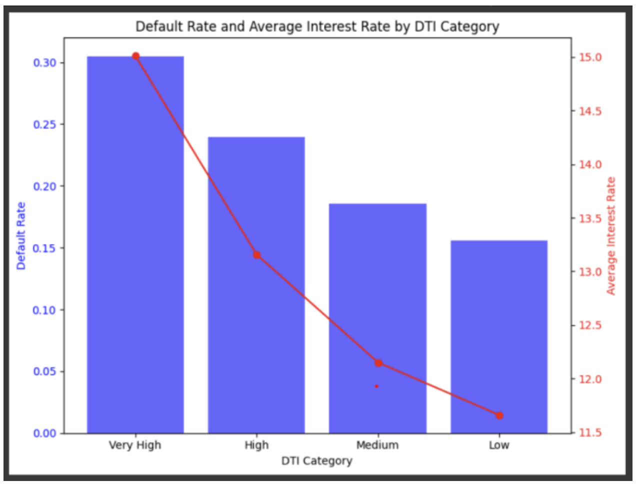
We performed an exploratory data analysis (EDA) on key borrowers and loan details to understand default risk in peer-to-peer lending. This revealed trends in how factors like income, debt-to-income ratio, credit score, home ownership, and loan purpose relate to default rates and interest rates, helping us select important factors for further analysis.

Default Rate and Average Interest Rate by Income Category



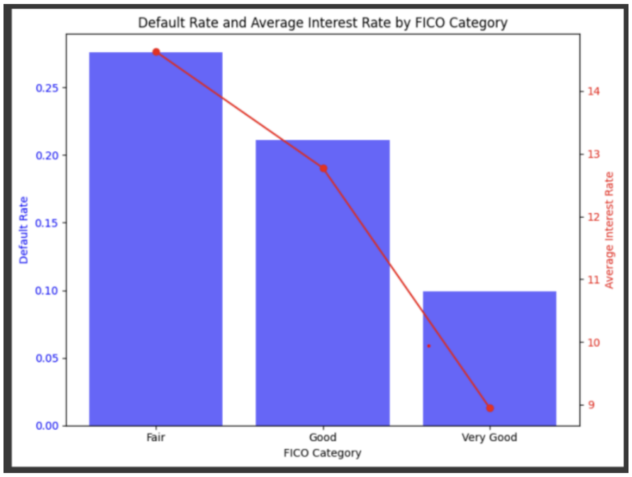
The chart shows that borrowers with lower incomes have the highest default rates, over 25%, while those with higher incomes have lower default rates, with high-income borrowers defaulting less than 15% of the time. Interest rates also decrease as income rises—low-income borrowers pay the highest rates, around 14.5%, compared to about 12% for high-income borrowers. This pattern suggests that lenders view low-income borrowers as higher risk and charge them more, which may increase their financial strain and default risk.

Default Rate and Average Interest Rate by Debt-to-Income (DTI) Category



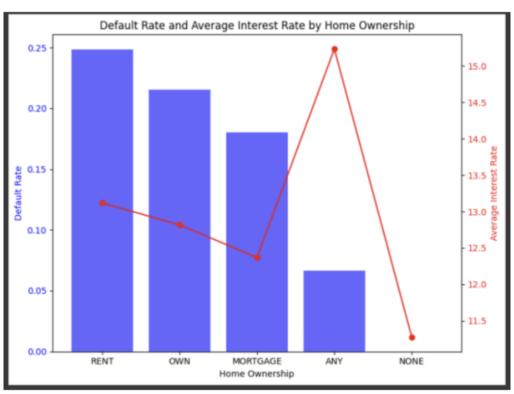
It has been study that Debt-to-income ratio and credit grade are significant predictors of default. (Serrano, 2015). The chart shows that borrowers with a "Very High" debt-to-income (DTI) ratio have the highest default rate, over 30%, while those with lower DTI ratios have progressively lower default rates. This suggests that borrowers with higher debt compared to their income struggle more with repayment. The red line shows that interest rates also decrease as DTI goes down—borrowers with a "Very High" DTI face rates around 15%, while those with a "Low" DTI have rates closer to 11.5%. This indicates that lenders see high DTI as a risk factor and charge higher rates, which may, in turn, add financial pressure and increase the risk of default.

Default Rate and Average Interest Rate by FICO Category



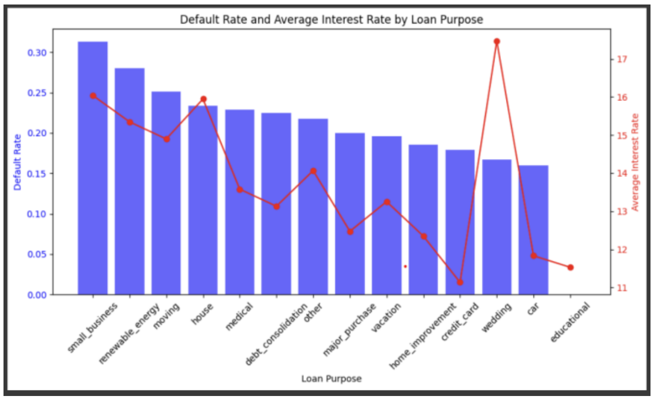
The chart shows that default rates decrease as FICO scores improve. Borrowers with a "Fair" FICO score have the highest default rate, above 25%, while those with "Good" and "Very Good" scores show significantly lower default rates. This suggests that higher credit scores are associated with lower default risk. The red line indicates that average interest rates also drop as FICO scores increase—borrowers in the "Fair" category face interest rates above 14%, while those in the "Very Good" category have rates closer to 9%. This pattern suggests that lenders consider FICO scores a key factor in assessing risk, offering better rates to those with stronger credit.

Default Rate and Average Interest Rate by Home Ownership



The chart shows that borrowers who rent have the highest default rate, close to 25%, while those with a mortgage or who own their homes have lower default rates. This suggests that homeowners, particularly those with mortgages, may have more financial stability and, therefore, lower default risk. The red line reveals a mixed trend in interest rates across home ownership categories. Renters and homeowners with mortgages face lower average interest rates, around 12–13%, while borrowers in the "ANY" category have the highest interest rates, above 15%. This pattern suggests that lenders may offer slightly better rates to those with stable housing arrangements, although the variation is less clear than with other factors.

Default Rate and Average Interest Rate by Loan Purpose



Iyer(2016) found out that studying soft information, such as borrower narratives, significantly improves default predictions. "Small business" loans have the highest default rate, over 30%, while loans for "education" and "car" purposes have the lowest. This suggests that business-related loans are riskier, leading to more defaults, while loans for more specific needs, like education and car purchases, are less risky. The red line shows that interest rates also vary, with "wedding" loans having the highest rates at over 17% and "educational" loans the lowest at around 11%. Lenders seem to charge higher rates for riskier purposes, making loan purpose a key factor in default risk and loan pricing.

Investment Decision

Based on our analysis of default and interest rates, we recommend focusing on lower-risk loan segments for a more stable return.

Primary Investment Criteria

- **FICO Score:** Target borrowers with "Very Good" and "Good" FICO scores, as they have lower default rates.
- **Income Level:** Invest in high-income borrowers who are generally more financially stable.
- **DTI Ratio:** Choose loans from borrowers with Low or Medium debt-to-income ratios to reduce risk.
- **Home Ownership:** Prioritize borrowers with "MORTGAGE" or "OWN" statuses, as they tend to be more financially stable.

Secondary Investment Criteria

- **Loan Purpose:** Focus on loans for home improvement and credit card refinancing among high-income, high-FICO borrowers, which are generally safer.
- **Loan Size:** Aim for medium-sized loans ($10,000 to $20,000) for a good balance of risk and return.
- **Loan Term:** Prefer shorter loan terms as longer loan terms are associated with increased default risk. (Emekter, 2015)

Risk Management Strategy

- **Diversification:** Spread investments across multiple loans that meet these criteria to limit the impact of any single default.
- **Conservative Approach:** Prioritize quality over higher returns, accepting slightly lower interest rates for loans with lower default risk, ensuring a more secure investment.

5. Machine Learning Model

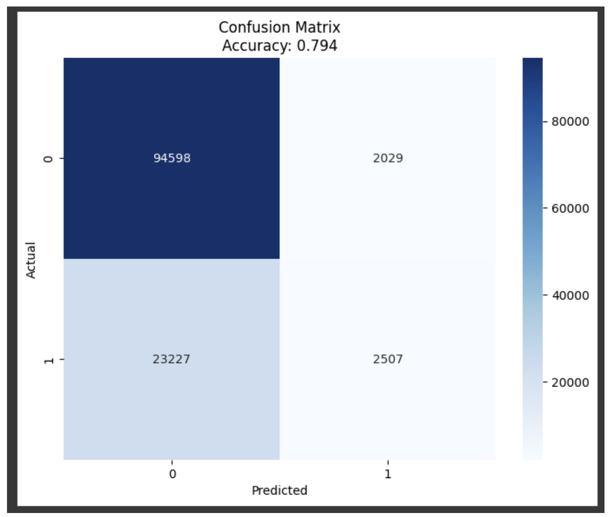
5.1 OLS Regression

While OLS is not typically used for binary classification, it can still provide insights into which features are most strongly associated with default risk.

Model Performance

Classification Report:

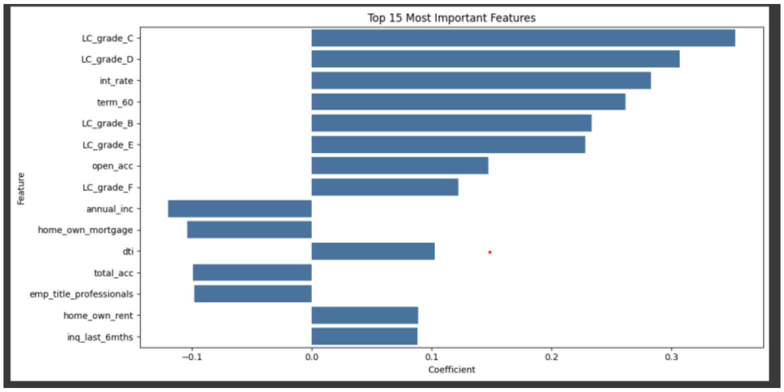
	precision	recall	f1-score	support
0	0.80	0.98	0.88	96627
1	0.55	0.10	0.17	25734
accuracy			0.79	122361
macro avg	0.68	0.54	0.52	122361
weighted avg	0.75	0.79	0.73	122361



The model's performance indicates that it effectively identifies non-default loans but struggles with predicting defaults. The classification report shows a high precision (0.80) and recall (0.98) for non-defaults, meaning the model reliably identifies loans that are likely to be repaid. However, the model performs poorly in detecting defaults, with a low precision of 0.55 and a recall of just 0.10, indicating that it misses many loans that should be flagged as default risks.

The confusion matrix further emphasizes these limitations, showing that the model accurately predicts a large number of non-defaults (94,598) but misclassifies a significant number of defaults (23,227) as non-defaults. This high false-negative rate is concerning, as it suggests that the model fails to capture a substantial portion of defaulted loans, limiting its effectiveness in identifying risk.

Correlation Analysis



We ran the correlation analysis and created a chart of the top 15 factors, showing which features have the biggest impact on the likelihood of a default:

- **Loan Grades** (e.g., LC_grade_C, LC_grade_D): Loan grades are strong indicators, with lower grades (like D and C) linked to a higher chance of default.
- **Interest Rate**: Higher interest rates are associated with a greater risk of default. This makes sense, as borrowers with higher interest payments may face more financial strain.
- **Loan Term**: Loans with a 60-month term are more likely to default compared to shorter-term loans, possibly because longer commitments can become harder to sustain over time.
- **Annual Income**: Higher income generally reduces default risk, as borrowers with more income are better able to meet loan payments.
- **Home Ownership**: Borrowers who own homes, especially those with a mortgage, have a lower risk of default than renters, suggesting they might have more financial stability.
- **Debt-to-Income Ratio (DTI)**: Higher DTI ratios are associated with a higher risk of default, indicating that borrowers with more debt compared to their income may struggle with payments.

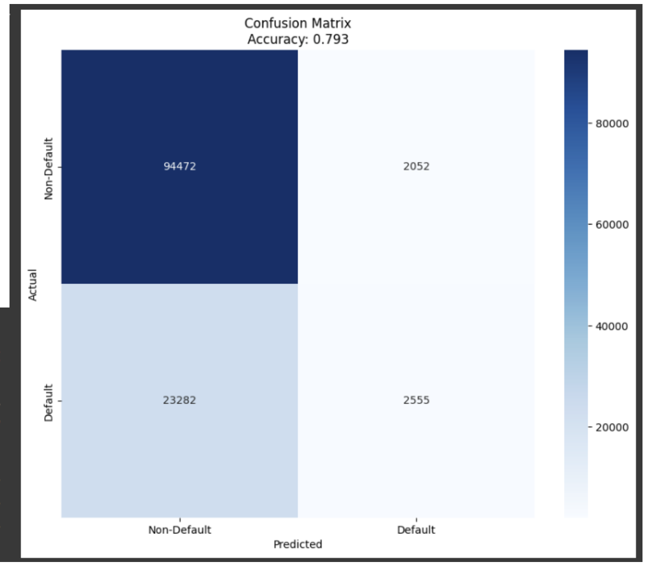
5.2 Logistic Regression

To improve on the OLS regression model, we applied a logistic regression model, which is more suitable for predicting binary outcomes, like loan defaults. This model helps estimate the probability of a loan default and is generally more reliable for classification tasks.

Model Performance

Classification Report:

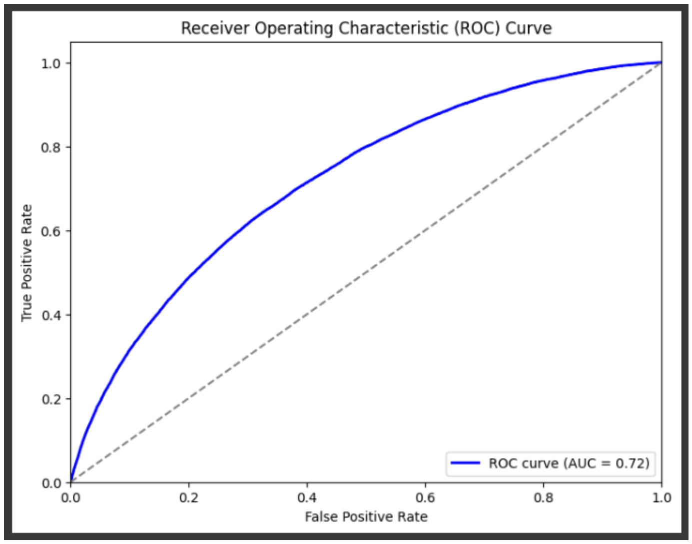
	precision	recall	f1-score	support
0	0.80	0.98	0.88	96524
1	0.55	0.10	0.17	25837
accuracy			0.79	122361
macro avg	0.68	0.54	0.52	122361
weighted avg	0.75	0.79	0.73	122361



The logistic regression model demonstrates high accuracy (79.3%) and performs well in identifying non-defaults (Class 0), achieving a precision of 0.80 and a recall of 0.98. This shows that the model reliably predicts loans that are likely to be repaid. However, it struggles with detecting defaults (Class 1), with a low recall of 0.10 and a precision of 0.55, indicating that it misses many loans at risk of default. This challenge is reflected in the low F1 score of 0.17 for defaults, underscoring the model's difficulty in balancing precision and recall for this class.

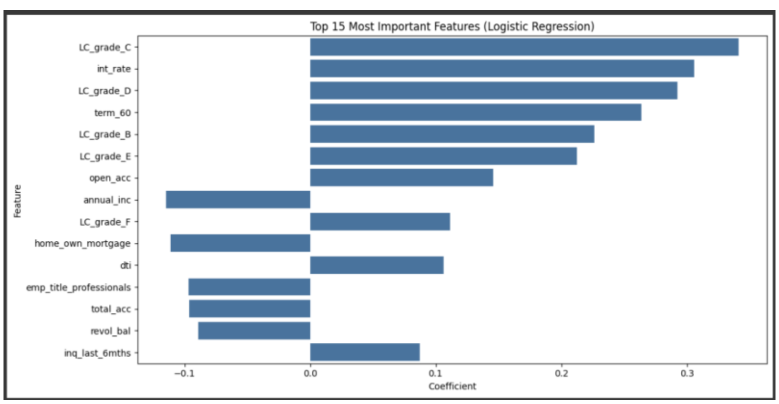
The confusion matrix also shows that the model correctly predicts a large number of non-defaults (94,472) but fails to identify many defaults, with 23,282 defaults misclassified as non-defaults. The Logistic Model is not enough to solve this problem, alternative models should be run to better capture default risk.

ROC Curve and AUC Score



The ROC curve provides a view of the model's ability to distinguish between non-defaults and defaults across different probability thresholds. The Area Under the Curve (AUC) score is 0.72, which indicates moderate discriminatory power. While an AUC of 0.72 suggests some ability to separate the two classes, it’s not exceptionally high, and there is room for improvement, especially in accurately identifying defaults.

Correlation Analysis - logistic regression



The feature importance chart and coefficients table reveal which factors are most influential in predicting default risk:

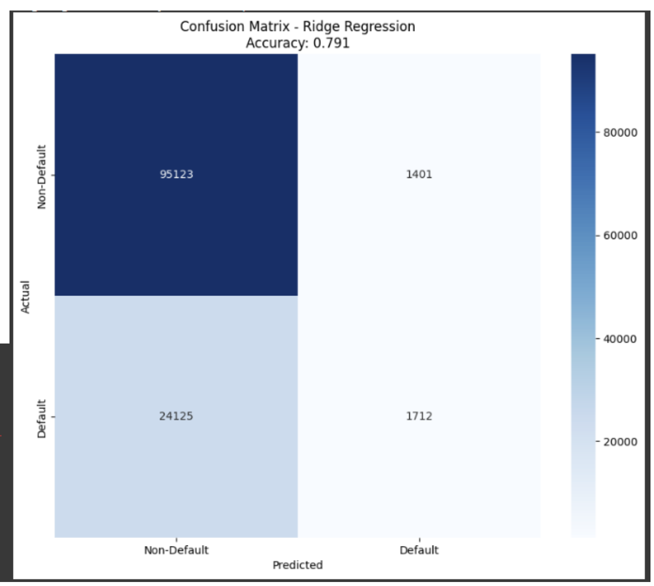
- Loan Grades (LC_grade_C, LC_grade_D, LC_grade_B, etc.): Loan grades are among the strongest predictors of default risk. Borrowers with lower grades (e.g., C, D) are more likely to default, as reflected by their positive coefficients and odds ratios above 1.

- Interest Rate (int_rate): Higher interest rates are associated with an increased risk of default. This aligns with the idea that borrowers with higher rates may face more financial strain.
- Loan Term (term_60): A loan term of 60 months is positively associated with default risk, suggesting that longer-term loans may be harder for borrowers to repay consistently.
- Annual Income (annual_inc): Higher annual income is negatively associated with default risk, meaning borrowers with higher incomes are less likely to default.
- Home Ownership (home_own_mortgage): Borrowers with a mortgage are less likely to default compared to renters or those without ownership, as indicated by the negative coefficient.
- Debt-to-Income Ratio (dti): Higher DTI ratios are associated with a higher likelihood of default, supporting the idea that borrowers with more debt relative to their income are at greater risk.

5.3 Ridge Regression

Model Performance

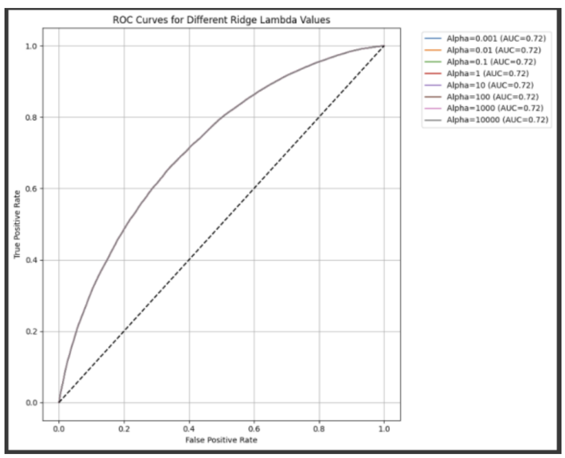
```
Classification Report for Ridge Regression:
              precision    recall  f1-score   support

           0       0.80      0.99      0.88     96524
           1       0.55      0.07      0.12     25837

    accuracy                           0.79    122361
   macro avg       0.67      0.53      0.50    122361
weighted avg       0.75      0.79      0.72    122361
```



The Ridge Regression model achieved an accuracy of 79.1%, compared to both the OLS and Logistic Regression models. It performs well in identifying non-default loans (Class 0), with a high precision of 0.80 and recall of 0.99, meaning it accurately classifies the majority of non-defaults. However, the model still struggles significantly with detecting defaults (Class 1), achieving a low recall of 0.07, which indicates that it correctly identifies only 7% of actual defaults. This low recall, coupled with a precision of 0.55 for defaults, highlights the model's limitations in capturing loans likely to default, likely due to class imbalance.

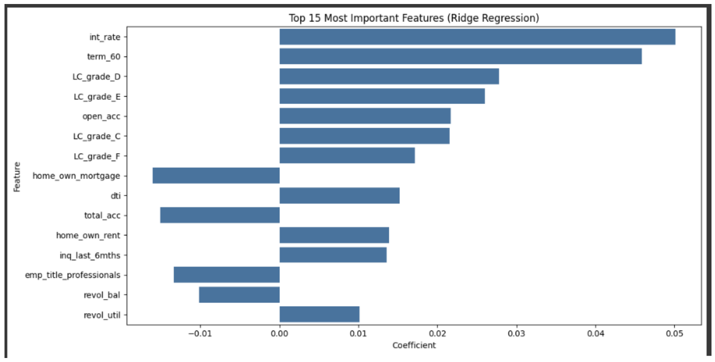
This limitation is also shown by the confusion matrix. While the model correctly identifies a large number of non-defaults (95,123), it misclassifies a substantial portion of defaults (24,125) as non-defaults. The small number of true positives (1,712) reinforces the need for alternative approaches or adjustments.

ROC Curve and AUC Score



The ROC curve and AUC score provide additional perspective on the model’s overall discriminatory power. With an AUC score of 0.72 across various alpha values, Ridge Regression demonstrates a moderate ability to distinguish between defaults and non-defaults. However, the AUC score also reflects that the model’s power to capture defaults remains limited, indicating room for improvement in separating the two classes more effectively.

Feature Importance and the Impact of Regularization



The Ridge Regression model highlighted key predictors of loan default risk. The top features included:

- **Interest Rate (int_rate)**: Higher interest rates were strongly associated with higher default risk, suggesting that borrowers facing higher borrowing costs are at greater risk.
- **Loan Term (term_60)**: A 60-month loan term was also linked to a higher likelihood of default, possibly due to the financial strain of a longer repayment period.
- **Loan Grades**: Categories like Lc_grade_C, Lc_grade D, and LC_grade_E were important predictors, with lower-rated grades showing a stronger association with defaults.
- **Debt-to-Income Ratio (dti)**: Higher DTI ratios were associated with increased default risk, as borrowers with high debt relative to their income may face financial difficulties.
- **Annual Income (annual_inc)**: Higher annual incomes were negatively associated with default, indicating that borrowers with higher earnings are less likely to default.

Performance Summary Across Alpha Values

Performance Summary for Different Alpha Values:

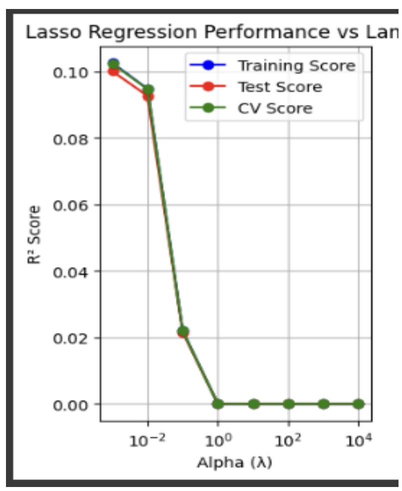
	Alpha	R2_Train	R2_Test	R2_CV	MSE_Train	MSE_Test
0	0.001	0.1031	0.1007	0.1029	0.1494	0.1498
1	0.010	0.1031	0.1007	0.1029	0.1494	0.1498
2	0.100	0.1031	0.1007	0.1029	0.1494	0.1498
3	1.000	0.1031	0.1007	0.1029	0.1494	0.1498
4	10.000	0.1031	0.1007	0.1029	0.1494	0.1498
5	100.000	0.1031	0.1007	0.1029	0.1494	0.1498
6	1000.000	0.1031	0.1007	0.1029	0.1494	0.1498
7	10000.000	0.1031	0.1007	0.1029	0.1494	0.1498

The performance summary across different alpha values further highlights Ridge Regression's behavior:

- **R^2 Scores**: The R^2 values for training, testing, and cross-validation remained low (around 0.10), indicating that the model explained only a small portion of the variance in loan default outcomes. This suggests that, while Ridge Regression controls for multicollinearity, it has limited explanatory power for this dataset.
- **MSE Scores**: Both training and testing MSEs were consistent across alpha values, remaining around 0.149. This stability shows that increasing regularization does not significantly impact error reduction, underscoring the limited effect of Ridge Regression on overall predictive performance.

5.4 Lasso Regression

Model Performance with Different Alpha Values



In the Lasso Regression model, we experimented with different alpha (λ) values to observe their impact on the model's predictive power. At lower alpha values, the model retained some predictive ability, with an R^2 score of around 0.10 on both training and test sets. However, as alpha increased, the R^2 score fell to zero, indicating a trade-off where higher regularization led to simpler models but at the cost of predictive accuracy. This pattern suggests that a minimal level of regularization might be best for capturing important relationships within the data without overly penalizing coefficients.

Mean Squared Error (MSE) Analysis

Comparison of Selected Features Across Alphas:

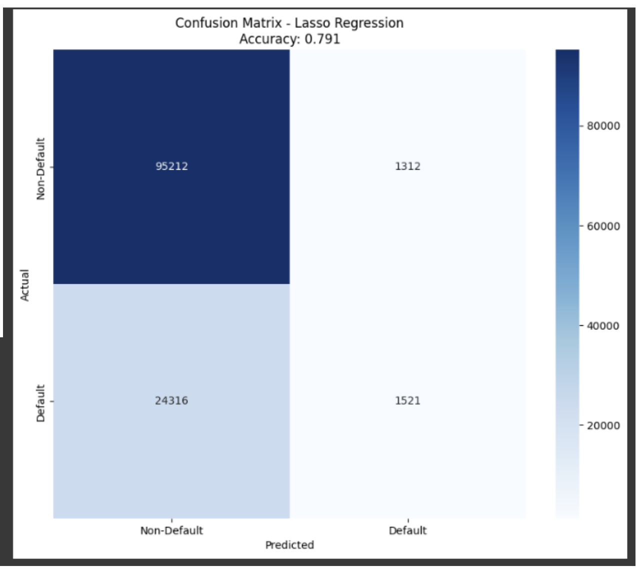
	Alpha	Lasso_Features	Lasso_Train_R2	Lasso_Test_R2	Lasso_MSE_Test
0	0.001	33	0.1027	0.1002	0.1499
1	0.010	8	0.0949	0.0927	0.1511
2	0.100	1	0.0220	0.0216	0.1630
3	1.000	0	0.0000	-0.0000	0.1666
4	10.000	0	0.0000	-0.0000	0.1666
5	100.000	0	0.0000	-0.0000	0.1666
6	1000.000	0	0.0000	-0.0000	0.1666
7	10000.000	0	0.0000	-0.0000	0.1666

Mean Squared Error (MSE) provided further insights into the model's performance. At the lowest alpha (0.001), the model achieved an MSE of approximately 0.1499, indicating reasonable accuracy. As alpha increased, MSE values also increased, reflecting a decrease in precision. This trend confirms that higher regularization weakens the model's predictive ability, suggesting that keeping alpha low might be necessary for Lasso to retain meaningful insights from the data.

Classification Report and Confusion Matrix

Classification Report for Lasso Regression:

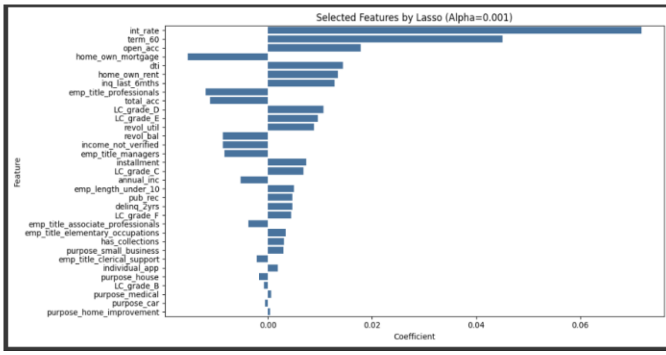
	precision	recall	f1-score	support
0	0.80	0.99	0.88	96524
1	0.54	0.06	0.11	25837
accuracy			0.79	122361
macro avg	0.67	0.52	0.49	122361
weighted avg	0.74	0.79	0.72	122361



The Lasso Regression model achieved an accuracy of 79.1%, similar to other models, and shows strong performance in identifying non-defaults (Class 0) with a precision of 0.80 and a recall of 0.99. This indicates that the model accurately identifies most non-default loans. However, it again struggles significantly with defaults (Class 1), achieving a low recall of 0.06, meaning it captures only 6% of actual defaults. Its precision of 0.54 suggests limited reliability when predicting defaults, likely due to the model's bias toward non-defaults.

The confusion matrix also shows that while the model correctly identifies a substantial number of non-defaults (95,212), it misclassifies a large portion of defaults (24,316) as non-defaults. This high false-negative rate indicates a considerable gap in the model's ability to capture loans likely to default, limiting its practical utility for risk prediction. The small number of true positives (1,521) further highlights the need for adjustments.

Feature Importance and Selection Stability




```
Feature Stability Analysis:

Feature Selection Stability (% of times selected):
int_rate                              37.5
term_60                               25.0
income_not_verified                   25.0
home_own_mortgage                     25.0
home_own_rent                         25.0
dti                                   25.0
inq_last_6mths                        25.0
emp_title_professionals               25.0
emp_title_associate_professionals     12.5
LC_grade_E                            12.5
dtype: float64
```

Lasso's feature selection properties allow us to see which variables are most predictive of default risk under different levels of regularization. At the lowest alpha (0.001), the model selected 33 features. Key predictors include:

- **Interest Rate (int_rate)**: This emerged as the most influential predictor, with higher interest rates correlating strongly with default risk, likely due to increased financial strain on borrowers.
- **Loan Term (term_60)**: A 60-month loan term was associated with higher default risk, suggesting longer repayment periods may add pressure on borrowers.
- **Open Accounts (open_acc)** and **Debt-to-Income Ratio (dti)**: Both variables were associated with default risk, indicating that borrowers with higher numbers of open accounts and higher debt-to-income ratios may be more vulnerable to default.
- **Home Ownership (home_own_mortgage)**: Borrowers with a mortgage appeared slightly less likely to default, possibly due to the stability associated with homeownership.

Feature Stability: Across different alpha values, certain features, especially **interest rate** and **loan term**, consistently showed up as significant predictors, underscoring their robust role in assessing default risk.

5.5 Random Forest Model

A research by Malekipirbazari and Aksakalli (2015) has shown that random Forests outperform traditional scoring models in predicting loan defaults in P2P lending, and offer superior predictive capabilities that inform sections on model comparison and selection.

Confusion Matrix and Classification Report

```
Confusion Matrix:
[[142844   1888]
 [ 36608   2201]]

Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.99      0.88    144732
           1       0.54      0.06      0.10     38809

    accuracy                           0.79    183541
   macro avg       0.67      0.52      0.49    183541
weighted avg       0.74      0.79      0.72    183541
```

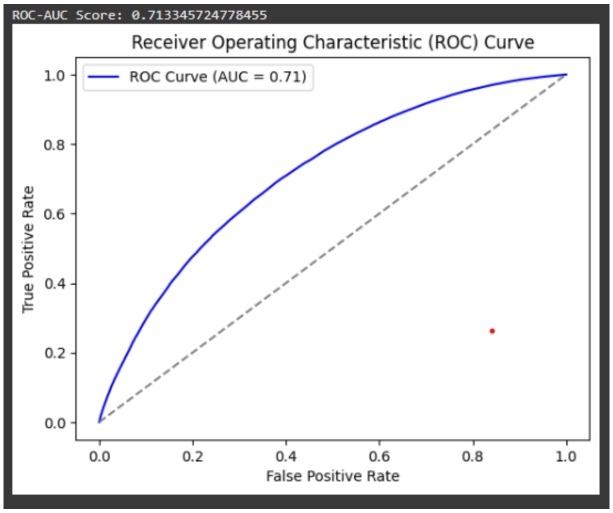
The confusion matrix reveals that the Random Forest model is effective at identifying non-default loans, with 142,844 correctly classified as non-default (true negatives) and only 1,888 misclassified as defaults (false positives). However, the model has difficulty with defaults, identifying only 2,201 true positives out of 38,809 actual defaults, leaving a high number of false negatives (36,608). This imbalance suggests that while the model is dependable for non-default predictions, it struggles to identify default cases accurately.

The classification report provides further clarity:

- **Non-Default Precision**: 0.80, indicating the model is quite accurate in identifying non-default loans.
- **Non-Default Recall**: 0.99, showing it captures almost all non-default cases.
- **Default Precision**: 0.54, meaning that when the model predicts a default, it's correct 54% of the time.
- **Default Recall**: 0.06, revealing the model only captures a small portion of actual defaults.

Overall, the model achieves an accuracy of 79%, but the low recall for defaults shows it could benefit from tuning to better capture high-risk cases.

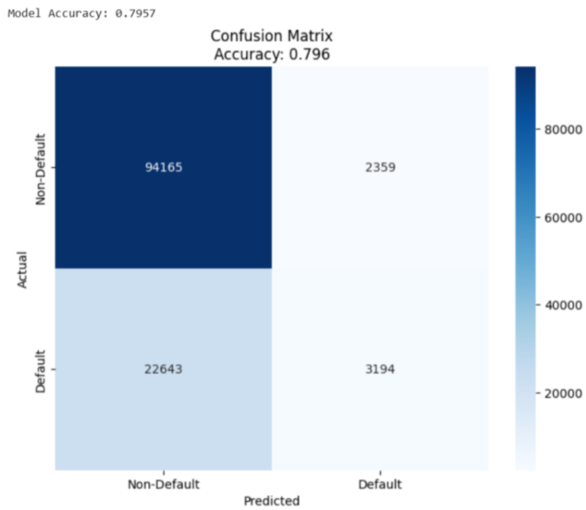
ROC Curve and AUC Score

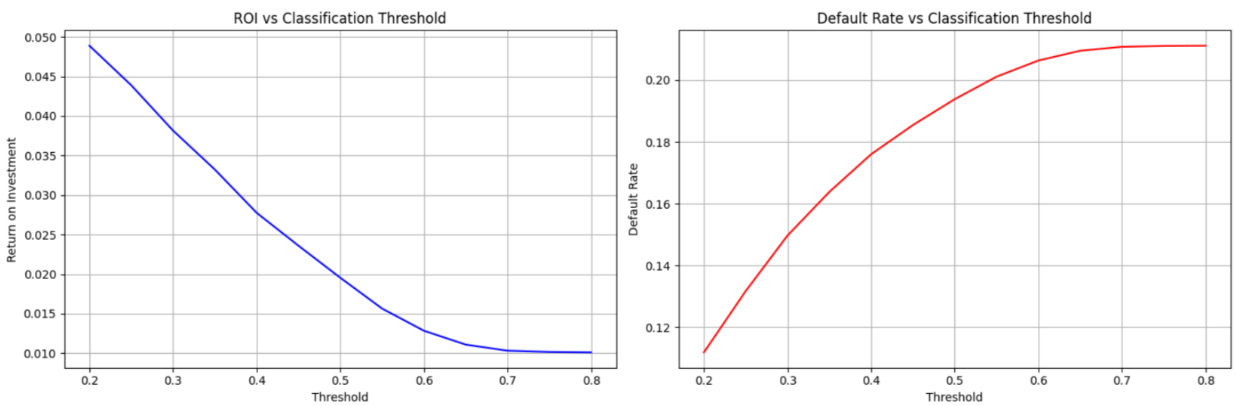


The ROC curve, with an AUC score of 0.71, shows that the model has a moderate ability to distinguish between defaults and non-defaults. Although an AUC of 0.71 is better than random guessing, it indicates room for improvement, particularly in identifying defaults more reliably.

5.6 LightGBM model

We tried a sophisticated approach to balancing risk and return in P2P lending decisions, using the LightGBM model.





With a 79.6% accuracy rate, it correctly identified over 94,000 good loans and 3,000 potential defaults. Setting the decision threshold at 0.20 achieved the best results: a 4.9% return on investment with an 11.2% default rate. Each approved loan averaged $641 in returns, leading to a total portfolio return exceeding $43 million.

However, the model faces some challenges. The high ROI volatility (36.5%) and negative risk-adjusted return measure suggest inconsistent performance. The results show a clear trade-off: stricter lending criteria reduce returns but also lower default risks. While the model approves 54.9% of loans at the optimal threshold, its high volatility indicates that additional safeguards might be needed for more stable real-world performance. This highlights the ongoing challenge of finding the right balance between being too cautious and too generous in lending decisions.

6. Conclusion

In this study, we applied multiple machine learning models, including logistic regression, ridge and lasso regression, random forest, and LightGBM, to predict loan default risk. Each model provided unique insights, with logistic regression offering a foundational understanding, while ensemble methods like random forest and LightGBM achieved greater accuracy in classifying defaults. Notably, LightGBM excelled due to its ability to handle complex data patterns and imbalances, showing promise as a reliable tool for lending decisions.

Throughout the analysis, we identified key factors influencing default risk. Variables such as loan grade, interest rate, loan term, and borrower income level emerged as significant predictors. Higher interest rates and longer loan terms were particularly associated with an increased likelihood of default, aligning with expected financial behaviors. Experimenting with threshold adjustments further highlighted the trade-offs between maximizing return on investment and controlling default rates, offering practical insights for risk management.

Overall, this analysis emphasizes the value of traditional statistical methods with advanced machine learning approaches for better predictive accuracy in lending. The combination of the regression models enables more comprehensive risk assessment, helping lenders make data-driven, informed decisions. Future research could focus on refining feature selection and experimenting with even more advanced models to capture additional nuances, ultimately enhancing loan portfolio management.

Nigmonov (2022) has stated that a higher interest rate and inflation increase the probability of default in the P2P lending market. For future research and model refinement, we suggest Incorporating macroeconomic indicators, such as unemployment rates or inflation rates to improve default predictions and enhance those models.

References

1. Malekipirbazari, Milad, and Vural Aksakalli. 2015. "Risk Assessment in Social Lending via Random Forests." Expert Systems with Applications 42 (10): 4621–31.
2. Nigmonov, Asror, Syed Shams, and Khorshed Alam. 2022. "Macroeconomic Determinants of Loan Defaults: Evidence from the U.S. Peer-To-Peer Lending Market." *Research in International Business and Finance* 59 (January): 101516.
3. Emekter, Riza, Yanbin Tu, Benjamas Jirasakuldech, and Min Lu. 2014. "Evaluating Credit Risk and Loan Performance in Online Peer-To-Peer (P2P) Lending." *Applied Economics* 47 (1): 54–70.
4. Serrano-Cinca, Carlos, Begoña Gutiérrez-Nieto, and Luz López-Palacios. 2015. "Determinants of Default in P2P Lending." Edited by Mikael Bask. *PLOS ONE* 10 (10): e0139427.
5. Iyer, Rajkamal, et al. *Inferring Asset Quality: Determining Borrower Creditworthiness in Peer-To-Peer Lending Markets*. 1 July 2011,