ITESO, Universidad
Jesuita de Guadalajara

MODELOS DE CRÉDITO

Credit Model for Personal Loans in the U.S.

Prof. Rodolfo Slay Ramos

José Pablo Jiménez Mendoza
Camila Alejandra Lopez Camacho
Marcos Ramos Alvarez
Paulo Adrian Villa Dominguez
Pablo Emilio Sanchez Galicia
Francisco Alexander Sotomayor Pineda

Credit Model Selection

Credit models are a widely used tool by financial (and non-financial) institutions to measure and assess the risk of default or non-payment of a debt by a borrower. These models rely on different classification techniques to label applicants based on their probability of default, forming the backbone of risk assessment and lending processes. Financial institutions, car dealerships, and even manufacturing companies that offer credit options to their clients rely heavily on these models to minimize risks and make informed lending decisions.

The goal of this project is to develop a predictive credit model powered by Machine Learning. By utilizing such a model, the aim is to improve the accuracy and reliability of loan approval predictions, ultimately aiding in making data-driven lending decisions. For those unfamiliar with these types of models, Machine Learning algorithms draw from databases to identify patterns and make predictions. They are the driving force behind the rapidly advancing technology known as Artificial Intelligence (AI).

This project uses the "PS4E9 || Original Data || Loan Approval Prediction" dataset, which contains 32,581 entries with various applicant and loan-related details. A custom-designed ML model will classify applicants as "Eligible" or "Not Eligible" for a personal loan at an interest rate determined by the team. The geographical scope of this project focuses on the United States due to the quality and availability of data from its financial institutions, making it an ideal region for model training and analysis.

By combining advanced data analysis and feature engineering techniques, this project aims to create a robust and accurate credit model that can efficiently classify loan applicants and help financial institutions make informed lending decisions.

Interest Rate Formation

The formation of an interest rate for a personal loan is not simply a percentage decided on a whim by the risk manager of a bank. This interest rate, which borrowers pay in addition to the loan amount, is influenced by multiple underlying interest rates that are constantly fluctuating. As a result, the interest rate on a personal loan today may not be the same as it was a year ago or

even a few months ago, even if the loan is from the same bank, for the same amount, and to the same borrower. The formation of an interest rate for any type of loan involves six key components, which vary depending on the country, the prevailing economic context, the lending institution, and the borrower. These components will be listed below and explored in more detail throughout this investigation.

1. Base Rate
2. Inflation Premium
3. Credit Risk Premium
4. Liquidity Premium
5. Administrative Costs
6. Profit Margin

These components interact to determine the final interest rate that borrowers face. Each one plays a specific role in reflecting the risks, costs, and desired returns associated with issuing a loan. This section serves as an introduction to these factors, which will be analyzed more thoroughly in the following parts of this research.

Reference Rate / Base Rate

The base rate in the United States not only serves as a benchmark for American institutions but also for many countries around the world. To get an idea of its impact, numerous central banks globally announce changes to their own base rates within days of the U.S. Federal Reserve's (commonly known as the Fed) announcements. The Federal Reserve is the central bank of the United States and is responsible for steering the country's monetary policy, which is crucial given that the U.S. is a major global economic power. This responsibility gives the Fed significant influence over global financial trends.

The primary objective of the Fed is to control inflation, and one of the key tools for achieving this goal is managing the reference rate, which is the minimum rate at which a bank lends money

to its customers. Changes in this rate help regulate factors such as inflation and either stimulate or contract economic growth based on whether the rate is raised or lowered.

The most recent adjustment by the Fed to the reference rate was on September 18, 2024, when it was reduced by 50 basis points, bringing it down to 5%. This marks the first rate cut by the U.S. central bank after more than a year of keeping it steady at 5.5%. Before that, during the first half of 2023, the rate followed an upward trend. In the graph below, the evolution of the U.S. interest rate over the past two years can be seen, illustrating the Fed's efforts to control rising inflation in the United States, which will be discussed further later on.

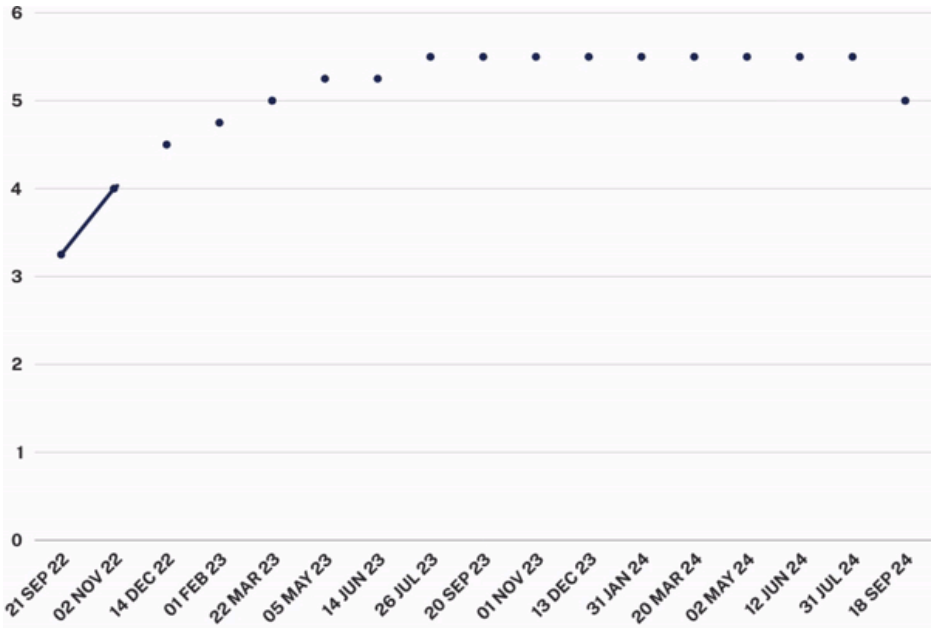

Image 1. Fed Fund Rates - Source. investing.com

Inflation Premium

The inflation premium is the additional cost or interest on a loan that compensates for the loss in the value of money over time due to inflation. Lenders apply this premium based on the expected inflation rate during the loan's duration. Many investors and economists "remove" this factor from the interest rate to determine the real interest rate. This concept is clearly illustrated by the Fisher Equation, which states that the nominal interest rate is equal to the sum of the real interest rate plus the inflation rate:

$$i = r + \pi$$

Where:

- i = Nominal interest rate
- r = Real interest rate
- π = Inflation rate

Inflation is measured based on changes in the prices of a selected basket of basic goods and services. Each country determines this selection based on the consumption patterns of its residents. In the United States, the current inflation rate is at 2.4%, indicating a downward trend over the last six months after reaching its peak of 3.5% in March of this year. Below is a graph showing the evolution of inflation in the U.S. over the past year.

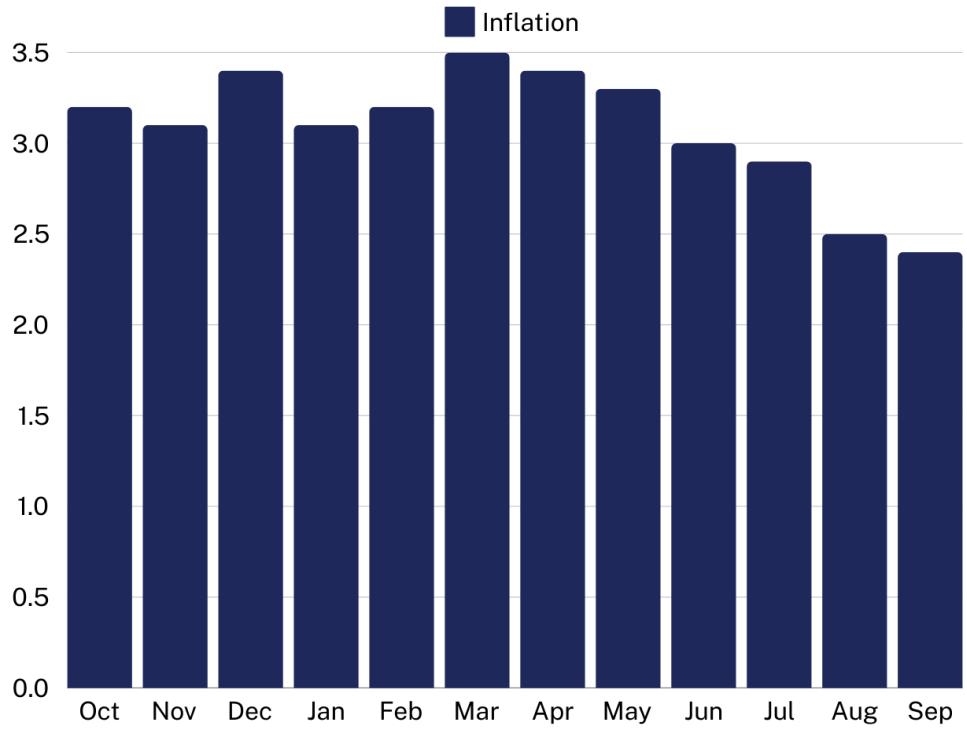

Image 2. United States Inflation Rate - Source. tradingeconomics.com

Credit Risk Premium

The credit risk premium is an additional charge applied to loans as compensation for the risk of borrower default. Lenders impose this premium to safeguard against the possibility of not recovering the full loan amount, including interest, due to the borrower's failure to meet payment obligations. Among the various components that form the interest rate, the credit risk premium is one of the few that is directly controlled by the lender, allowing them to adjust rates based on the perceived risk.

This premium is closely tied to the borrower’s creditworthiness, which is assessed through credit scores, financial history, and other key factors. Borrowers with a strong credit history, stable income, and a good track record of timely payments are considered low-risk and typically receive lower interest rates. On the other hand, borrowers with a history of late payments, high debt levels, or other red flags are deemed higher risk, and as a result, are charged a higher credit risk premium to compensate for the increased likelihood of default.

Lenders use this mechanism to balance potential losses with profitability, ensuring that the additional risk taken on when lending to high-risk individuals is adequately compensated. This differentiation in the premium encourages borrowers to maintain a good credit profile while allowing lenders to manage their risk exposure effectively.

Liquidity Premium

The liquidity premium is another crucial component influencing interest rate formation. This term refers to the additional compensation lenders demand for lending money in markets or situations where access to liquid capital is limited or more costly. Essentially, this premium is added to compensate for the risk that assets cannot be quickly converted to cash without negatively affecting their market value.

The liquidity premium varies depending on the type of credit and the nature of the borrower. For example, personal loans typically have a lower premium compared to mortgage loans or long-term loans since the former are shorter in term and have lower liquidity risk. However, during times of economic uncertainty or credit restrictions, this premium can increase

significantly. Financial institutions facing difficulties in accessing stable funding sources may raise credit costs to compensate for this lack of liquidity.

One way to mitigate the liquidity premium is to ensure that the borrower has access to additional lines of credit or an alternative funding source that reduces the perceived risk. Studies show that institutions that better manage their liquidity tend to offer more competitive interest rates and attract lower-risk borrowers.

Administrative Costs

Administrative costs also form part of the set of components defining the interest rate on a personal loan. These costs refer to the operating expenses a credit institution must cover to process and manage loans. They include, but are not limited to:

- **Credit evaluation:** Expenses incurred to assess applicants' creditworthiness.
- **Staff and technology:** Salaries of personnel involved in credit management, as well as the technological platforms used for risk management and loan processing.
- **Regulations and compliance:** Costs associated with adhering to local and federal regulations in the U.S., which require a series of formal procedures and documentation.

These costs vary significantly depending on the institution's size and the level of process automation. Institutions that have invested in technology and automation often succeed in reducing their administrative costs, allowing them to offer better loan conditions to clients. However, it is essential to note that while these costs decrease with digitalization, they cannot be entirely eliminated.

A recent study suggests that the average administrative cost to issue personal loans in the U.S. represents about **1-2% of the total loan amount**, depending on the lender's operational efficiency.

Profit Margin

The profit margin is the component that ensures the lender earns a profit after covering all the costs associated with issuing and maintaining a loan. This margin varies according to the

financial institution's policies, investor return expectations, and market competition. In personal loans, the profit margin is typically higher than in other financial products, such as mortgage loans, due to the higher perceived risk and shorter loan terms.

Lenders set the profit margin based on several factors:

1. Risk Level: The higher the perceived risk in the borrower's profile, the higher the profit margin to compensate for potential losses.
2. Market Conditions: In a market with low interest rates or high competition, profit margins tend to compress, forcing institutions to lower them to attract more customers.
3. Financial Objectives: Each lender has specific return-on-investment goals, and the profit margin is a key tool for achieving these objectives.

An excessively high profit margin can make the loan unattractive to borrowers, while a margin that is too low could jeopardize the institution's profitability. For this reason, setting an appropriate profit margin is essential to balancing profitability and competitiveness in the lending market. Lenders must carefully evaluate market dynamics and borrower risk to establish a margin that aligns with their financial goals while remaining competitive.

Determining our interest rate

Once we have checked all the components that influence our interest rate and a quick view of their actual status, we can determine our interest rate. Each one of the components cooperates to form a new interest rate. We can divide those components into two different groups: the ones lenders can control and the ones that don't. Components like the base rate and the inflation premium are components that rely completely on the economic conditions of the economy; they can go up when economic conditions are not favorable and go down when there is a pleasant economic climate. On the other hand, credit risk premium and liquidity are components that rely on the lender and can vary the most depending on the financial institution. Administrative costs and profit margin are components that do rely on the lender but typically are the same for all of them, swinging between 1-2% for both.

We propose the next configuration of the components:

Component	Percentage contributed
Base Rate	5%
Inflation premium	0.5%
Credit Risk Premium	3.5%
Liquidity Premium	1.5%
Administrative Costs	1%
Profit Margin	1%
Interest Rate	**12.5%**

Credit risk is the premium that can vary the most, because it depends on the credit score the borrower has, the higher the credit score the probability of default is smaller and vice versa. The previous credit risk premium is considered for a borrower with a high credit score between 720-850. The average FICO score in the United States was 715 during 2023 (Horymski, 2024).

Interest Rate Breakdown - Justification

1. **Base Rate (5%)**

The Base Rate of 2.5% represents the minimum return expected in a risk-free environment, typically aligned with the yield on secure government securities like U.S. Treasuries. This rate compensates for the opportunity cost of lending funds without taking on significant risk. By setting the base rate at 5%, we align it with historical returns in stable economic conditions, providing a solid foundation for the interest rate calculation.

2. **Inflation Premium (0.5%)**

The Inflation Premium of 0.5% has been assigned to account for the expected loss of purchasing power over time due to inflation. This value considers current inflationary expectations and market conditions, aiming to safeguard the lender's return against erosion by rising prices. In the

current economic scenario, where inflation is relatively controlled, a 0.5% premium is appropriate for reflecting anticipated inflation trends.

3. **Credit Risk Premium (3.5%)**

The Credit Risk Premium of 3.5% compensates the lender for the possibility of borrower default. This value is selected based on the borrower's credit profile and reflects a moderate level of risk. By setting the premium at 3.5%, the lender ensures that the interest rate adequately covers the probability of credit losses while remaining competitive in the market.

4. **Liquidity Premium (1.5%)**

The Liquidity Premium of 1.5% addresses the additional cost associated with investing in assets that cannot be quickly or easily converted to cash without a loss of value. This premium reflects the illiquidity risk present in certain financial instruments or investment conditions. Given the prevailing market environment, a 1.5% premium provides a balanced approach to account for this risk.

5. **Administrative Costs (1%)**

The Administrative Costs have been set at 1%, which covers expenses related to loan origination, processing, monitoring, and servicing. This allocation considers operational costs such as personnel, technology platforms, and regulatory requirements. By maintaining administrative costs at 1%, the financial institution can manage its operational expenses efficiently while remaining competitive in the market.

6. **Profit Margin (1%)**

The Profit Margin of 1% is intended to ensure a reasonable gain for the lending institution, enabling it to operate sustainably and meet its financial targets. This margin strikes a balance between market competitiveness and institutional profitability, allowing for steady growth and investment in future operations.

Risk factor analysis

When granting a loan, financial institutions must consider various risk factors that influence the interest rate and determine the applicant's credit risk. In this model, five categories of risk are evaluated: Credit Risk, Market Risk, Operational Risk, Legal and Regulatory Risk, and Country Risk. Below is a detailed description of the quantification process for each of these factors and their impact on the final interest rate.

Credit Risk

Credit risk refers to the likelihood of a borrower defaulting on a loan, potentially causing financial loss to the lender. To quantify this risk, we calculate the Probability of Default (PD). This begins with a logistic regression model, which establishes an initial risk estimate, and is subsequently refined using an XGBoost model that captures more complex patterns in the data for a more accurate individual credit risk assessment. The resulting PD provides a tailored risk score for each applicant, allowing for a more precise evaluation of their risk profile.

Based on the PD, a Credit Risk Premium is added to the base rate. For instance, applicants with a high PD (over 5%) may incur an additional premium of 2-3%, while those with a low PD might see a smaller premium of around 1% or potentially none at all. This approach ensures that the interest rate is adjusted to reflect the specific risk posed by each borrower.

Market Risk

Market risk refers to external economic conditions that can impact a borrower's ability to repay their loan. Key indicators of market risk include GDP growth, interest rate volatility, and unemployment rates. This risk is quantified through macroeconomic metrics such as GDP growth and unemployment levels, as higher unemployment and lower economic activity generally signal increased market risk. Additionally, interest rate volatility, assessed through the standard deviation of rates over a defined period, provides insights into the potential for significant fluctuations.

Based on these indicators, a Market Risk Premium is incorporated into the interest rate. For example, during times of economic uncertainty, an additional premium of 0.5-1% is added to the base rate to account for increased market risk.

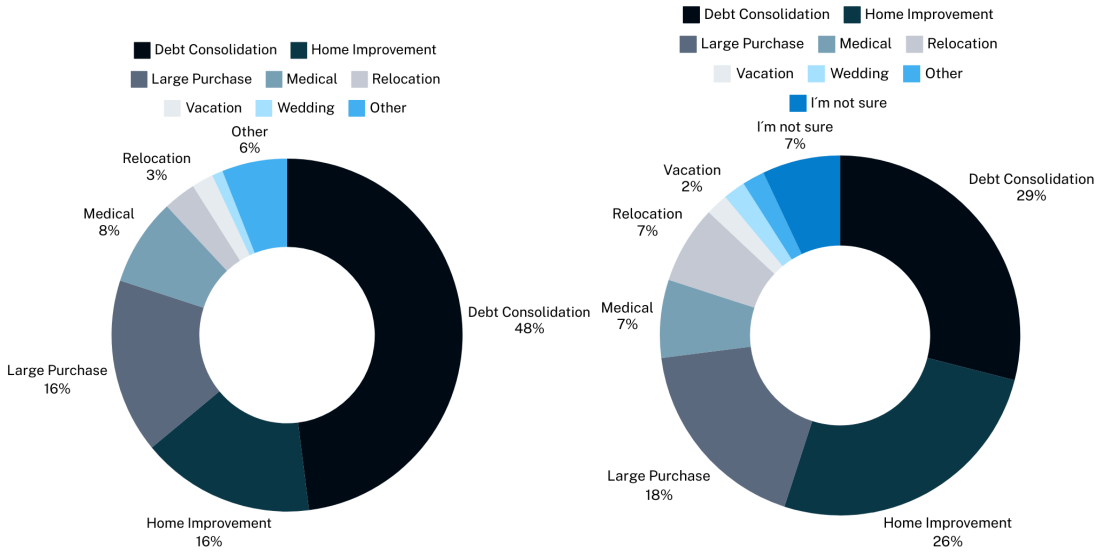

Image 4. Debt Consolidation is th eMost Common Use for Personal Loan - Source. Investopedia

The previous graph is an example of how the personal loans have been used over the past year. The data showed on the graph that 48% of the people that have applied for a personal loan are to consolidate other debt, which makes it easier that any external economical condition can affect the ability to repay.

Operational Risk

Operational risk involves potential losses stemming from internal processes, system inefficiencies, or human errors, all of which can increase the costs of issuing and managing loans. This risk is quantified by tracking operational costs as a percentage of the total issued credit, alongside regular internal audits to assess risks tied to process inefficiencies. Additionally, the institution's level of digitalization plays a key role in mitigating operational risk, as automated systems typically reduce errors and enhance efficiency.

To account for these risks, an Operational Premium of 0.5-1% is added to the interest rate. Financial institutions with high levels of digitalization or efficient processes may reduce or even eliminate this premium, reflecting their ability to minimize operational costs.

Legal and Regulatory Risk

Legal and regulatory risk refers to the potential costs associated with complying with national and local regulations. Changes in legislation or compliance requirements can introduce unforeseen expenses that impact the financial institution. This risk is quantified by tracking compliance costs as a percentage of operational expenses, alongside analyzing how new or changing regulations may impose additional financial burdens on the credit issuance process. Typically, increased regulation leads to higher compliance costs, which in turn affect the lender's overall operating expenses.

To address these risks, a Regulatory Premium of 0.5-1% is added to the interest rate if compliance costs are substantial or if the regulatory environment is perceived as unstable. This premium is especially relevant when regulations directly impact specific loan types or certain borrower categories.

Country Risk

Country risk refers to the political or economic instability in the borrower's country that could impact their ability to repay the loan. For loans within the United States, this factor typically has minimal influence but may be adjusted if lending expands to international markets. Country risk is quantified using sovereign credit ratings from agencies such as S&P and Moody's, along with additional sovereign risk indicators. Metrics such as inflation trends and currency stability provide further insights, allowing for a more precise evaluation of specific country risks.

In cases involving high-risk countries, a Country Risk Premium of up to 2% may be added to the interest rate to offset potential losses. However, in stable countries like the United States, this premium is usually zero or minimal, reflecting lower levels of political and economic uncertainty.

Conclusion

This project has developed a predictive credit model that utilizes machine learning techniques to improve the precision and dependability of loan approval decisions. Through a comprehensive analysis of interest rate formation and the integration of critical risk factors, this model demonstrates a balanced, data-driven approach tailored to personal loans.

In the **Interest Rate Formation** analysis, six principal components were identified: Base Rate, Inflation Premium, Credit Risk Premium, Liquidity Premium, Administrative Costs, and Profit Margin. Each component plays an essential role in defining the interest rate to reflect both broader economic conditions and borrower-specific risks. By structuring these premiums, particularly those for credit risk and liquidity, the model effectively aligns the interest rate with individual applicant risk profiles, thereby balancing profitability goals with borrower affordability.

The **Credit Risk Model** employs both logistic regression and XGBoost to assess applicants' eligibility, which allows for baseline accuracy while capturing complex patterns within applicant data. This dual approach achieved a model accuracy rate of 92%, indicating strong predictive performance. The inclusion of critical metrics such as Probability of Default (PD) and Loss Given Default (LGD) further enabled precise, risk-adjusted interest rates for different applicant profiles. By configuring weights for each risk factor, high-risk borrowers are allocated higher premiums, thus safeguarding the lending institution against potential defaults.

Limitations and Future Directions: While this model successfully leverages available data for risk assessment, incorporating additional metrics—such as real-time market conditions or more detailed applicant financial histories—could further refine predictive accuracy. Regular recalibration of premiums in response to changing market and regulatory conditions is recommended to maintain the model’s effectiveness over time.

In conclusion, this project achieves its objective of creating an adaptable, industry-aligned credit model that enhances lending decisions. The methodologies applied here set a solid foundation for ongoing improvements, with future iterations potentially incorporating expanded economic

indicators and advanced machine learning techniques to continually advance credit risk assessment.

Appendix

Project Workflow and Python Code Outline

The project follows a structured workflow that includes data loading, exploratory data analysis, preprocessing, model building, evaluation, and risk assessment. The Python code for these steps utilizes custom classes for data loading (`DataLoader`), EDA and preprocessing (`LoanEDA`), modeling (`LoanApprovalModel`), and risk analysis (`RiskAnalysis`). The following outlines the classes and functions used:

- *DataLoader*: Loads and summarizes the dataset, checks for missing values, and displays basic information about the data.
- *LoanEDA*: Handles exploratory data analysis by plotting histograms, KDE plots, violin plots, and categorical feature distributions. It also replaces outliers and imputes missing values.
- *LoanApprovalModel*: Benchmarks a logistic regression model, trains an XGBoost model, and evaluates model performance using various metrics such as precision, recall, and F1-score.
- *RiskAnalysis*: Performs a detailed analysis of the Probability of Default (PD) distribution, calculates risk assessments, and visualizes risk by category, income range, and other factors.

Data Summary and Imputation Strategy

The dataset contains 32,581 entries with 12 features. The columns include numerical features such as `person_age`, `person_income`, `person_emp_length`, `loan_amnt`, `loan_int_rate`, `loan_percent_income`, and `cb_person_cred_hist_length`, as well as categorical features like

`person_home_ownership`, `loan_intent`, `loan_grade`, and `cb_person_default_on_file`. The target variable, `loan_status`, indicates whether a loan application was approved or not.

Missing data were identified in the `person_emp_length` and `loan_int_rate` columns. These missing values were imputed using the median to ensure data consistency. Additionally, outliers in `person_age`, `person_income`, and `person_emp_length` were replaced with the median to reduce skewness.

Model and Evaluation Metrics

An initial logistic regression model was established as a baseline with an AUC score of 0.7955. The refined XGBoost model achieved higher accuracy, with the final evaluation showing an overall accuracy of 92%, supported by balanced precision and recall across both loan approval classes. This indicates the model's effectiveness in distinguishing between high and low-risk applicants.

Credit Risk Assessment and Visualization

The model includes a comprehensive risk assessment framework, which calculates the Probability of Default (PD) for each applicant. Key visualizations include:

- *PD Distribution*: Illustrates the concentration of low and high-risk borrowers within the dataset.
- *Risk Assessment by Category*: Displays how risk varies across different loan grades, identifying groups with higher risk levels.
- *Risk Concentration Curve (Lorenz Curve)*: Demonstrates the cumulative distribution of risk relative to the applicant pool, highlighting areas of risk concentration.
- *Risk by Income Range*: Shows how risk assessments differ by applicant income levels, helping to pinpoint trends and vulnerabilities.

The Total Portfolio Risk was calculated at $7,597,802.16, providing a benchmark for further portfolio adjustments and risk mitigation strategies.

Credit Risk Metrics

Two critical metrics were integrated into the analysis:

- *Probability of Default (PD)*: Indicates the likelihood of a borrower defaulting on their loan, serving as a basis for personalized risk assessment.
- *Loss Given Default (LGD)*: Set at 40% for this analysis, representing the estimated loss in the event of a default.

References and Sources of Data

- Dataset Source: "PS4E9 || Original Data || Loan Approval Prediction" dataset, containing 32,581 loan applications with detailed applicant and loan-specific metrics.
- Credit Risk Models and Machine Learning Techniques: Referenced from standard financial risk assessment literature and XGBoost implementation guidelines.

Bibliographic References:

Federal Reserve Bank of Richmond. (2023). *Efficiency in Consumer Lending: A Cost-Analysis of Administrative Practices*. https://www.richmondfed.org/publications/research

Finance Rate on Personal Loans at Commercial Banks, 24 Month Loan. (2024). Stlouisfed.org. https://fred.stlouisfed.org/series/TERMCBPER24NS

Petersen, M. A., & Rajan, R. G. (2020). *The Economics of Lending and Borrowing*. *Journal of Finance, 75*(3), 1207-1235. https://doi.org/10.1111/jofi.12945

Peterson, L. (2023). *Investopedia's 2023 Personal Loan Borrower Survey*. Investopedia. https://www.investopedia.com/investopedia-2023-personal-loan-borrower-survey-8383574

Mishkin, F. S. (2019). *The Economics of Money, Banking, and Financial Markets* (12th ed.). Pearson.https://www.pearson.com/store/p/the-economics-of-money-banking-and-financial-markets/P100000585280

VanSomeren, L. (2024). *Best Personal Loan Lenders for October 2024: Picks for Every Need*. Investopedia. https://www.investopedia.com/best-personal-loans-4773300#comparison-table

BBVA. (2024, 24 octubre). *¿Qué es la Tasa de Referencia?* BBVA México. Recuperado 21 de octubre de 2024, de https://www.bbva.mx/educacion-financiera/t/tasa_de_referencia.html

Buckley, D. (2023, 27 octubre). *The 5 Components of Interest Rates - DayTrading.com*. daytrading.com. https://www.daytrading.com/components-interest-rates

Ceizyk, D. (2024, 17 octubre). *What is the average personal loan interest rate?* Bankrate. Recuperado 27 de octubre de 2024, de https://www.bankrate.com/loans/personal-loans/average-personal-loan-rates/#credit

Horymski, C. (2024, 22 enero). *What Is the Average Credit Score in the US?* Recuperado 27 de octubre de 2024, de https://www.experian.com/blogs/ask-experian/what-is-the-average-credit-score-in-the-u-s/

Investing.com. (2024, 18 septiembre). *Estados Unidos - Anuncio de la decisión de política monetaria*. Recuperado 21 de octubre de 2024, de https://mx.investing.com/economic-calendar/interest-rate-decision-168

Keizer, J., & Keizer, J. (2021, 5 marzo). Determinants of Interest Rates. *CFA, FRM, and Actuarial Exams Study Notes*. https://analystprep.com/study-notes/actuarial-exams/soa/fm-financial-mathematics/determinants-of-interest-rates/

TRADING ECONOMICS. (s. f.). *United States inflation rate*. Recuperado 23 de octubre de 2024, de https://tradingeconomics.com/united-states/inflation-cpi