



# **Predictive Modelling for Loan Approval: A Data-Driven Approach to Binary Classification**

Shweta Anoop

20023484@mydbs.ie

Applied Research Project

submitted in partial fulfilment of the requirements for the degree of

Master's in Business Analytics at

Dublin Business School

Dublin, Ireland

**Supervisor:** Nithya Govindaraju

August 2025

## Contents

<b>Chapter 1 - Introduction</b> .....	6
<b>1.1 Problem Statement</b> .....	6
<b>1.2 Motivation for the Study</b> .....	6
<b>1.3 Research Question</b> .....	7
<b>1.4 Research Objectives</b> .....	7
<b>Chapter 2 – Literature Review</b> .....	9
<b>2.1 Introduction</b> .....	9
<b>2.2 Traditional Credit Scoring Models</b> .....	9
<b>2.3 Emergence of Machine Learning at Credit Risk</b> .....	10
<b>2.4 Binary Classification for Loan Approval</b> .....	10
<b>2.5 Dealing with Class Imbalance</b> .....	11
<b>2.6 Ensemble Models in Loan Prediction</b> .....	11
<b>2.7 Explainability and Trust in AI Models</b> .....	12
<b>2.8 Fairness and Ethics in Credit Scoring</b> .....	12
<b>2.9 Preprocessing in Credit Risk Datasets</b> .....	12
<b>2.10 Handling Categorical Variables with SMOTENC</b> .....	13
<b>2.11 Comparative Performance of ML Models in Credit Scoring</b> .....	13
<b>2.12 Interpretability and Black-box Modeling in Finance</b> .....	14
<b>2.13 Fairness Auditing with SHAP and Feature Importance</b> .....	15
<b>2.14 Loan Prediction and Deployment Readiness</b> .....	15
<b>2.15 Trends in Ethical AI and Fintech Regulation</b> .....	16
<b>2.16 Deep Learning in Credit Risk Assessment</b> .....	16

2.17 Real-Time Loan Approval Systems .....	17
2.18 Financial Technology (FinTech) and ML Integration .....	18
2.19 Emerging Issues: Fairness, Regulation, and Governance .....	18
2.20 Research Gap and Synthesis .....	19
<b>Chapter 3: Methodology</b> .....	<b>20</b>
3.1 Introduction .....	20
3.2 Dataset Description .....	20
3.3 Exploratory Data Analysis .....	20
3.4 Data Preprocessing.....	21
3.5 Model Development.....	21
3.5.1 Logistic Regression.....	21
3.5.2 Random Forest Classifier .....	22
3.5.3 XGBoost Classifier .....	22
3.5.4 Gaussian Naive Bayes .....	22
3.6 Model Evaluation .....	22
3.7 Explainable AI (XAI) Integration.....	23
3.7.1 Feature Importance.....	23
3.7.2 SHAP Analysis .....	23
3.8 Model Export and Deployment Readiness.....	24
3.9 Ethical Considerations.....	24
<b>Chapter 4 - Results and Discussion</b> .....	<b>25</b>
4.1 Data Overview and Initial Analysis.....	25
4.1.1 Dataset Snapshot and Feature Composition .....	25
4.1.2 Data Types and Structure .....	26
4.1.3 Descriptive Statistics of Numerical Features .....	27

4.1.4 Target Variable Distribution .....	28
4.1.5 Missing Values Check .....	29
4.2 Exploratory Data Analysis .....	30
4.2.1 Loan Status Distribution.....	30
4.2.2 Gender and Loan Approval Outcomes.....	31
4.2.3 Education and Loan Approval Trends.....	33
4.2.4 Home Ownership and Loan Status Patterns.....	35
4.2.5 Loan Intent and Approval Outcomes .....	37
4.2.6 Age Distribution and Its Association with Loan Status .....	39
4.2.7 Credit Score Distribution and Approval Outcomes .....	40
4.2.8 Proportional Income Commitment and Loan Status.....	42
4.2.9 Interest Rate Patterns and Approval Outcomes.....	44
Chapter 5 - Conclusion and Future Work.....	56
Future Work .....	56
References.....	58

## List of Figures

Figure 1 - Sample Records from the Dataset .....	25
Figure 2 - Dataset Info Summary .....	27
Figure 3 - Summary Statistics .....	28
Figure 4 - Summary Statistics .....	29
Figure 5 - Loan Status Distribution .....	30
Figure 6 - Loan status by Gender .....	32
Figure 7 - Loan Status by Education Percentage .....	34
Figure 8 - Loan Status by Homeownership .....	36
Figure 9 - Loan status by Loan Intent Percentage .....	38
Figure 10 - Age Distribution by Loan Status .....	39
Figure 11 - Credit score distribution by loan status .....	41
Figure 12 - Loan Percent Income by Loan status .....	43
Figure 13 - Interest rate distribution by loan status .....	45
Figure 14 - Random forest performance matrices .....	46
Figure 15 - Logistic Regression performance matrices .....	47
Figure 16 - Logistic Regression performance matrices .....	48
Figure 17 - Consolidated ROC Curve .....	49
Figure 18 - Random Forest Feature Importance .....	50
Figure 19 - XG Boost Feature Importance .....	51
Figure 20 - Logistic Regression Importance .....	52
Figure 21 - Mutual Information Naive Bayes .....	53
Figure 22 - Force Plot .....	54
Figure 23 - Streamlit Application .....	55

## **Chapter 1 - Introduction**

### **1.1 Problem Statement**

Loan approval is a crucial decision-making process in the financial institutions since its result determines the achievement of goals of individual borrowers and the organization of making profits and reducing the exposure to risk. As the number of applications to take out a loan increases, and the character of borrowers becomes more diverse and complex, traditional methods of evaluation are progressively unable to generate estimates that are at the same time precise and fair. The traditional approaches to methodologies are based on strict rules and fixed parameters of thresholds; they pursue some form of predictive accuracy that cannot be matched by the heterogeneous reality of a wide degree of applicants. This limitation has highlighted the need for more intelligent and adaptive systems for loan decision making.

One of the major challenges in credit data is class imbalance, where the number of approved and rejected applications is uneven. This can lead to biased models that favor the majority class.

### **1.2 Motivation for the Study**

Predictive modeling with machine learning is one of the most tremendous developments in the framework of applied statistics. The algorithms of machine learning can identify and discriminate patterns in historical data and then use the identified patterns to make probabilistic predictions about new data. In the lending process where loans are approved, the algorithmic models are used to determine the likelihood of loan approval or rejection by analyzing the past records of credit-risk. Binary classification models are especially useful in this regard since it allows placing each application into one of two mutually exclusive categories approved or rejected. This study aims to develop and evaluate such a predictive model using real-world credit data to improve the accuracy, efficiency, and fairness of loan approval decisions.

This research study was selected following the proven ability of the use of data to enhance decision-making procedures in the financial services sector. Specifically, predictive modeling will serve to reduce the occurrence of bad debts to ensure that the right people are creditworthy

and those at high risk are not. Online credit ratings can increase access to credit because their evaluation system can be fairer to applicants who, with a traditional system, can be rejected. In a more competitive financial world, the ability of an institution to manage credit risk is essential to its growth and long-term sustainability. By incorporating a diverse range of variables into the modeling process, this research attempts to bridge the gap between academic innovation and practical application.

The current research has a lot of implications for both the lending and borrowing communities. On the part of lenders, the implementation of data-driven strategies is anticipated to increase the pace of processing and reinforce the operational efficiency, thus reducing financial loss. At the same time, it is expected that these practices will increase customer satisfaction by promoting transparency and equity during the lifecycle of the loan. Predictive modeling is a methodological tool that finds its application in the context of borrowers with non-traditional financial history since it allows making a more in-depth evaluation of credit risk. The current research contributes to the growing field of financial technology with the suggestion of a methodology that can be reproduced to assess creditworthiness, which is based on machine-learning approaches.

### **1.3 Research Question**

How effectively can machine learning-based binary classification models, enhanced with data balancing techniques like SMOTENC, improve the accuracy and fairness of loan approval predictions using real-world credit risk data?

### **1.4 Research Objectives**

1. To examine the current approaches to the assessment of credit risk and determine the shortcomings of the classical models of credit approval of loans.
2. To clean the real-life credit data through encoding categorical variables and class imbalance problems using SMOTENC.
3. To apply and contrast the results of different machine learning techniques like Logistic Regression, Random Forest, XGBoost and Naive Bayes to predict the status of getting loan granted/not granted.

4. To compare the two models based on performance metrics, such as accuracy, precision, recall, F1-score and AUC-ROC to establish the most effective algorithm.
5. To ensure ethical factors have been considered by evaluating the fairness and transparency of the model prediction, especially those that historically did not receive fair treatment.
6. To create a predictive framework that would be replicable and deployable to enable real-time, fair and scalable loan approval decision-making in financial institutions.

## Chapter 2 – Literature Review

### 2.1 Introduction

Loan approval is a traditional area of stake in decentralization in financial services, and data-driven methodologies are becoming widely used to limit human bias and improve credit risk evaluation for lenders. While they work in some cases, traditional rule-based credit scoring systems are not designed to capture the non-linear and complex relationships between borrower characteristics, sorted individual traits, and payer behavior.

The increased number and variation of applicant data make machine learning (ML) the most practical choice for both automation and fairness improvements in credit decision-making systems. Collectively, these results show that although ensemble models and boosting algorithms can achieve maximum predictive scores, a blend of statistics (mutual information) and explainable approaches (logistic regression) can help present a more balanced explanation of the decision predictors, making it fair and transparent in the process of loan issuance.

### 2.2 Traditional Credit Scoring Models

The statistical models used to assess credit risk have traditionally been linear in nature, with logistic regression and scorecards being the most common. Persons and Sinko (1997) further emphasized the flexibility and predictability of logistic grades in relation to loan default or approval. Thomas et al(2002). It was expanded on by Pass et al. (2002) with the use of credit scoring cards, which scored certain traits, such as income and age, to assess creditworthiness.

Unfortunately, those models all rely on straight-line relationships and struggle to accommodate the new behaviors emerging in borrower behavior. In addition, these traditional models represent a bottleneck in terms of scalability and generalization. The general lack of flexibility of scorecards has also been noted by Crook, Edelman and Thomas (2007), who draw the conclusion that it leads to greater inefficiency in economies that have relatively low fractionalized financial income or a shorter credit history. This shortcoming exemplifies the necessity for adaptive and intelligent systems.

### **2.3 Emergence of Machine Learning at Credit Risk**

The contribution of machine learning has disrupted the traditional way in which credit institutions evaluate the risk profile of a borrower. Lessmann et al. (2015). Another benchmark study was performed by Song et al. (2015), comparing 15 classification algorithms on credit scoring datasets, and they found that ensemble methods (like random forests and gradient boosting) show superior predictions compared to logistic regression-based models in both accuracy and AUC-ROC.

There are many other examples of applying ML techniques to predict defaults using transaction data; for instance Khandani, Kim and Lo (2010) used largely state-of-the-art ML methods, including neural networks and decision trees. In doing so, they showed that machine learning models not only made better predictions but also enabled risk management for financial institutions.

Related studies have cited the effectiveness of machine learning approaches as well, with Brown and Mues (2012) comparing random forests to logistic regression and naive Bayes for retail default scoring purposes and finding that machine learning methods were more flexible and effective, particularly in heterogeneous noisy data implicating hybrid variables.

### **2.4 Binary Classification for Loan Approval**

The financing decision activities among other applications of Binary classification are concerned with characterizing each loan application as either approved or not approved. Such an arrangement is especially appropriate to logistic regression, decision trees, support vector machines (SVM), and ensemble models such as random forest and XGBoost.

Malekipirbazari and Aksakalli (2015) present a peer-to-peer lending risk classification model based on random forests with high classification potential in order to determine whether it classifies as a good borrower or a high-risk borrower. Logistic regression is likely to be a benchmark where it is highly interpretable to classify the status of loans approval (Siddiqi, 2006), and more complex models such as XGBoost achieve the same with additional non-linear relationships (Chen & Guestrin, 2016).

## **2.5 Dealing with Class Imbalance**

Class imbalance, one of the major barriers in real-world loan data, is the complete opposite: the ratio of the loans that get approved greatly exceeds the ratio of rejected loans or the latter to the former. This imbalance can cause biased models that will miss the instances of minority classes.

To balance the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was introduced by Chawla et al. (2002) and used to generate synthetic one of the minority classes. Such future developments as SMOTENC (Synthetic Minority Oversampling Technique on Nominal and Continuous features) overcame the original SMOTE restriction on categorical variables.

Fernandez et al. (2018) assessed more than 20 different resampling techniques and determined that SMOTENC was one of the best to be applied to a mixture-type-feature dataset. Buda, Maki, and Mazurowski (2018) also showed that oversampling can be an effective choice when considering classification performance on imbalanced binary outcomes in general and tree-based ones specifically, proving that such an approach can enhance performance offers significant improvements.

## **2.6 Ensemble Models in Loan Prediction**

Random Forest and XGBoost ensemble models have been reported to perform spectacularly in credit risk modeling, because of their stability and overfitting propensity reduction. Random Forests by Breiman (2001) is a methodology of bagging, which consists in an average of the results of various decision trees, and XGBoost, proposed by Chen and Guestrin (2016), is a method of boosting that involves the sequential refinement of the model using the results of previous attempts.

The two models are inherently applied to mixed-type data and provide in-built feature importance measurements. These models are always more accurate and generalized when compared to different financial datasets (Lessmann et al., 2015).

## **2.7 Explainability and Trust in AI Models**

As more black-box models are deployed across areas of high stakes, such as lending, the need to explain them has become critical. LIME (Local Interpretable Model-agnostic Explanations) was proposed by Ribeiro, Singh, and Guestrin (2016) and SHAP (SHapley Additive exPlanations) by Lundberg and Lee (2017), which has become a standard tool to explain models.

SHAP values separate the prediction of a model in terms of individual feature changes, so lenders can comprehend and rationalize the decision. Hall et al. (2021) highlighted the need to explain difficulty in the context of processing discrimination algorithms and thus ensuring adherence with regulations regarding discrimination algorithms, especially the GDPR in the EU and the Equal Credit Opportunity Act in the U.S.

## **2.8 Fairness and Ethics in Credit Scoring**

Loan approval discriminatory ML models are increasing in concern. Barocas, Hardt, and Narayanan (2019) pointed out how machine learning systems can support historical discrimination and biased data. Such tools as SHAP can be used to determine whether any features, such as gender or marital status, have an outsized impact on the outcomes of a model. Mehrabi et al. (2021) surveyed measures of fairness in ML, and measures of mitigating fairness revealed that in most cases, fairness is measured at the expense of accuracy. Such trade-offs must be closely dealt with in cases of loan approval to ensure that the operational objectives and ethics are not neglected.

## **2.9 Preprocessing in Credit Risk Datasets**

Data preprocessing plays a central role during the credit risk modeling process, especially where there is mixed-type data in the presence of numerical and categorical variables. Categorical variable systems such as loan intent or home ownership need to be encoded so that they can be used by most of the machine learning models. According to Zheng and Casari (2018) some of the commonly used methods include label encoding and one-hot encoding. Because of this, nevertheless, they warn about the possibility of one-hot encoding leading to an extremely large feature dimensionality, particularly when the features are high-cardinality.

Kotsiantis et al. (2006) stated that to enhance predictive accuracy in classification tasks, missing values, scaling, and noise are to be handled. The error or the nulls in the financial datasets can be attributed to a mistake of the user or an error in the system or the application forms that are not filled in. Without remedy, the attendant error imposes bias or limits generalization of the model.

The standardization of the data is frequently suggested, especially when it comes to numerical characteristics of the data, like the credit score and income, so that different variables could be scaled consistently. Jain et al. (2019) proved that logistic regression and SVM algorithms have better convergence and performance after feature normalization or standardization. Another typical transformation, StandardScaler, already enforces both a mean of zero and variance normalized to one, which helps with gradient-based optimization procedures.

## **2.10 Handling Categorical Variables with SMOTENC**

Features that have class imbalances usually do not respond well to standard oversampling strategies such as SMOTE. Extensions such as Borderline-SMOTE and Adaptive Synthetic Sampling (ADASYN), proposed by Chawla et al. (2011), showed some improvement to categorical attributes, although they did not do too well in this aspect. The introduction of SMOTENC (Synthetic Minority Oversampling Technique for Nominal and Continuous) by Lund, Lee, and Chawla (2018) marked a significant shift. SMOTENC can produce real-looking synthetic instances using the distribution of categorical variables, including gender or education.

In the survey on imbalanced learning techniques (He and Garcia, 2009), it was concluded that SMOTENC is specifically useful when working with categorical data, where applications are required in marketing, credit risk, and medical decision-making. Furthermore, it was observed that the combination of SMOTENC and stratified k-fold cross-validation was reported to be more robust and prevents overfitting of a synthetic minority sample (Bellinger, Reiff-Marganiec and Sterling 2015).

## **2.11 Comparative Performance of ML Models in Credit Scoring**

The comparative assessment of classification algorithms using real data on credit scenarios has been profusely reported. A comparison of a decision tree, logistic regression, a neural network,

and a support vector machine was done by Baesens et al. (2003) based on credit scoring information. From an overall perspective, the ensemble methods are far better than the single classifier in prediction accuracy and stability.

Martens et al. (2009) further demonstrated that not only do random forest and tree-based models exceed the performance of other techniques, but feature importance values can be estimated that can be useful information in interpretation. On the other hand, both support vector machines and deep neural networks often function as opaque models, necessitating further explanation through post-hoc analysis.

The current challenges in the credit scoring community have recently shown XGBoost as one of the most efficient methods (Zhang et al., 2020). It is more production-ready and accounts internally to regularization and missing data, which speaks to its more production-ready financial use cases. According to Wang et al. (2018), they performed a benchmarking study in which they indicated that XGBoost yields higher accuracy and AUC-ROC values than logistic regression and random forest in most cases, especially on the dataset with noisy or highly non-linear relationships.

## **2.12 Interpretability and Black-box Modeling in Finance**

Lack of interpretability of high-performing ensemble and deep learning models represents one of the important criticisms of these models. This is a significant threat to the financial sector given that regulators demand transparency and audibility. Lipton (2018) defined two classes of interpretability of models: intrinsic interpretability (e.g., logistic regression) and post hoc interpretability (e.g., SHAP and LIME are related to tree-based models).

Doshi-Velez and Kim (2017) explained that interpretability should not only be considered a technical necessity but also a socio-technical one in fields such as healthcare and finance. Their framework determines the extent to which the descriptions given by a model can be understood by those stakeholders who are human and how it can assist in sound decision-making.

SHAP (SHapley Additive exPlanations) has become one of the most popular in recent years, as the use of the cooperative game theory is its theoretical basis (Lundberg & Lee, 2017). SHAP values provide both local and global interpretability in that they assign each predictive feature an important score based on a particular prediction. It worked well to model how fair it is within financial applications to evaluate bias in characteristics such as income or gender (Chen et al., 2020).

### **2.13 Fairness Auditing with SHAP and Feature Importance**

Fairness in models is becoming an essential part of ethical AI. Warned by Barocas and Selbst (2016), historical data frequently harbor societal biases, which are projected into an associated machine learning application unless reportedly audited. In the case of such biases, SHAP values are useful to observe because without them, the practitioner would not notice that a sensitive attribute impacts outcomes unfairly.

Mehrabi et al. (2021) introduced fairness-aware pipelines, which combine model explanation and statistical parity testing and adversarial debiasing. Their research indicates that any fairness intervention could become a part of existing ML workflows, particularly in the financial sector.

Kearns and Roth (2019) explored optimal actions in predictive models in which fairness trade-offs were examined game-theoretically. Their argument is that fairness needs to be weighed with predictive performance, particularly when the application is high stakes. Practically, that implies that although a model, such as XGBoost, could be quite accurate, the application should be conditional on the performance of different metrics of fairness regularly checked and published.

### **2.14 Loan Prediction and Deployment Readiness**

In terms of practical use, the readiness to deploy models is the same as practicing to train models. Trained models must serialize and have the same processing pipeline as that of training. As evidenced by Sculley et al. (2015), effective productionization of ML models involves the proper architectural pipeline, model versioning, and monitoring.

Kumar and Singh (2020) proposed a credit scoring deployment pipeline that included model export (joblib, e.g.), consistent scaling (StandardScaler, e.g.) and an API for real time decision. This is typically integrated into an institution's credit management system or loan Origination backend.

Additionally, the model drift and data drift need to be periodically tested, particularly in turbulent financial markets. The process of continual monitoring with the A/B testing and shadow deployment (Hastie, Tibshirani, & Friedman, 2017) allows making sure that the model remains effective when applied to new, incoming data without raising the issue of fairness or losing accuracy.

### **2.15 Trends in Ethical AI and Fintech Regulation**

Regulatory environment regarding AI in finance is changing fast. Accountability, fairness, and transparency to AI systems are emphasized in the EU Artificial Intelligence Act, the U.S. Equal Credit Opportunity Act (ECOA), and the Proposed Digital Personal Data Protection Bill of India.

In a study by Rahwan et al. (2019), the authors suggested the so-called Society-in-the-loop framework of algorithmic governance, which attributes the value integration of machine decision-making with those existing in society. This will imply that in the fintech sector, credit approval systems will be created not only to be effective, but also influential towards society.

Fairness toolkits in AI like the AI Fairness 360 by IBM, What-If Tool by Google, or Fairlearn by Microsoft are becoming more common in the financial services sector to audit and reduce algorithmic prejudices. Such tools tend to be coupled with SHAP, LIME, and adversarial debiasing models, so that their use can be within ethical guidelines.

### **2.16 Deep Learning in Credit Risk Assessment**

Deep Learning has had a good prospect of achieving sophisticated patterns in financial parameters within recent years. As opposed to classic models which train using hand-creation of features, deep neural networks (DNNs) automatically learn feature representations thus being highly useful in structured data as well as unstructured data.

Sirignano et al. (2016) are based on the deep learning model of predicting defaults, where millions of anonymized mortgage loan histories are used. The neural network compared to the conventional approach of credit scoring (both in terms of accuracy and its ability to capture subtle interplays between borrower profiles and macro conditions).

Liu et al. (2021) investigated the use of the long short-term memory (LSTM) networks to foresee credit card default, in particular, credit card applicants with a very small historical record of payments. They discovered that sequential models are more appropriate to describe repayment behavior.

Nevertheless, DL models have major interpretability shortcomings even though they have many strengths, particularly under controlled settings bankrolled contexts such as the financial realm. According to Rudin (2019), one should be more eager to pay attention to promoting the use of interpretable models in high-stakes fields, since black-box ones, correctly functioning, might not be useful for human justification.

### **2.17 Real-Time Loan Approval Systems**

Real-time loan approval systems are picking up as digital banks urge to give instant decisions. These solutions are based on effective integration of ML pipelines and transactional APIs that enable decisions to be made in milliseconds.

According to Wang and Siau (2019), the architecture of such systems includes delivery of models through a REST API, incorporation in the frontend forms, fraud detectors, and scoring engines. This makes it easy to have smoother user experiences, particularly in the mobile-first markets. Latency, resource footprint and scalability are other constraints of models in real-time requirement. Such models as logistic regression or gradient boosting machines are typically used because they incur less computation during inference and can be more easily deployed inside of containerized systems, like Docker or Kubernetes.

Modularity and flexibility are also promoted by microservice architecture. By applying such frameworks as FastAPI or Flask, it is possible to deploy the models, which have version control and monitoring endpoints. MLOps tools like MLflow (Zaharia et al., 2018) or TFX (Baylor et

al., 2017) provide support with experiment tracking, artifact-logging, and rollstock management of production.

## **2.18 Financial Technology (FinTech) and ML Integration**

ML has been adopted in the FinTech industry as a central facilitator of personalization, automation, and risk-mitigation. What banks have always used credit scoring with, digital lenders and neobanks have entered the fray with alternate data including mobile usage, utility bills, and social behavior.

In a comparison between traditional FICO scores and ML-based FinTech lending models, Jagtiani and Lemieux (2019) found that alternative models introduced a higher inclusiveness of credit-invisible people without affecting equally high repayment performance.

The Indian firms such as CASHe, KreditBee, and PaySense process the thin-file borrowers via ML-based processing with transactional and behavioral data. The Reserve Bank of India states that these methods have proved to be a prospective way to enhance access to credit and have led to concerns regarding data privacy and bias in algorithms (Reserve Bank of India, 2021).

Zhang et al. (2022) examined how to integrate the ML models with the blockchain-based lending where the immutability of records and trustless credit scoring are guaranteed. Indicative of these integrations which are still in their early phases, is that FinTech is driving things to the edge, where conventional modeling and networking go into decentralized finance (DeFi) realms.

## **2.19 Emerging Issues: Fairness, Regulation, and Governance**

The call of ML to become part of credit decision-making has led to an amplified concern about ethics, governance, and compliance. Both the EU's General Data Protection Regulation (GDPR) and the U.S.'s Fair Credit Reporting Act (FCRA) mandate explainability and non-discrimination in automated decision-making.

Binns et al. (2018) provided an example of counterfactual explanations, which serve as a means of regulatory transparency by answering the question: What would need to change to achieve a different outcome? In ML-based credit scoring, this enables the borrowers to know how they can maximize the likelihood of their approval.

Algorithmic auditing has been crystallized into an accepted best practice. Raji et al. (2020) introduced an audit scheme that includes data provenance, model training, bias verification, and deployment surveillance. Their solution is consistent with the increasing demands on model cards and transparency datasheets when implementing AI (Mitchell et al., 2019).

Zliobaite (2017) also pointed out that fairness must not only be statistical but contextual, considering the societal impact of erroneous rejections, especially for underbanked populations.

## **2.20 Research Gap and Synthesis**

The overview provided in the literature shows that the field of credit risk modeling has dynamically changed over the years: scorecards and logistic regression gave way to ensemble ML models, and explainable deep learning models incorporated modern deep learning. Methods such as SMOTENC and SHAP have now reached maturity to aid modelers with the needs of fair, interpretable, and accurate modeling of mixed-type, imbalanced data.

However, a clear research gap persists in combining all three key facets:

Handling Imbalanced Real-World Credit Data,

Ensuring Interpretability and Fairness via XAI, and

This research provides a replicable framework that is suitable for real-time deployment.

Whereas there have been some studies on one or two of these dimensions, less frequently have methods been developed that combine all three in a complete, end-to-end pipeline based on actual credit application data.

This research bridges the gap by creating and testing` binary classifiers on real-world credit risk information by using SMOTENC for balancing, and SHAP for added interpretability. It also includes real-time deployment and fairness auditing concerns, that cause methodology to be both practically deployable and ethically correct.

## **Chapter 3: Methodology**

### **3.1 Introduction**

The chapter describes the approach taken to designing a predictive model in loan approval classification that is based on real-life data regarding credit. It involves preprocessing of data, explorative data analysis (EDA), the use of class balancing methods, model training, model assessment using appropriate measures, and the insertion of Explainable AI (XAI) to improve interpretability, and fairness. The objective is to design a repetitive, equitable, and efficient binary classification method that can aid financial institutions to access the credit risk.

### **3.2 Dataset Description**

The data that will be employed in doing the analysis is collated and deanonymized records of loan applicant data that have both numerical and categorical data points that have significance in the assessment of credits. The available data (each observation is one application to a loan) contains such characteristics as `person_age`, `person_income`, `loan_amnt`, `loan_intent`, `person_home_ownership`, `credit_score`, and `loan_percent_income`. The target variable is `loan_status`, which is binary and where 1 show that the loan was approved and 0 indicates not approved. It is a theory-based dataset and highly leaves an impression of a scenario in real-life credit risk, in which the allocation of approvals has unequal distribution.

### **3.3 Exploratory Data Analysis**

Exploratory Data Analysis (EDA) was used to give a clearer image of the structure and the distribution the dataset possesses. The descriptive data analysis revealed the presence of the class bias regarding the approved and non-approved loans and the presence of continuous variables and categorical variables. Numerical data variables were summarized using statistical metrics and a bar plot and stack visualization were created to help test the interdependence between loan status and the feature of gender, education, home ownership, and loan intent. Histograms/density plots were used to explore distributions of age, credit score and loan amount, whereas scatter plots/violin plots offered additional information on the relationship between financial variables and loan status.

This comparison served to identify outliers, learn the patterns that were used to engineer the features, and select the model. The visualization of the class imbalance also affirmed that the resampling technique was necessary to prevent bias in the model prediction.

### **3.4 Data Preprocessing**

The process of preprocessing the dataset was essential to changing the original dataset into one applicable to machine learning models. Initially all categorical variables were converted into numeric values via LabelEncoder, which transforms the labels in the string format usually needed by most of the models in scikit-learn. `loan_intent`, `person_gender`, `person_home_ownership`, and `person_education` are the encoded feature.

After that, the SMOTENC (Synthetic Minority Oversampling Technique for Nominal and Continuous data) algorithm was used to resolve the problem of the class imbalance. Unlike typical SMOTE, SMOTENC operates with categorical and continuous features, which makes it viable in relation to the chosen dataset. It synthesized samples of the minority class (e.g., not-approved loans) using it in categorical word groups of features, thus balancing the classes and reducing bias.

After resampling, the StandardScaler was applied to ensure that each variable was scaled identically, as it was able to boost the performance of the model, especially with regard to logistic regression and XGBoost algorithms, which are sensitive to the scale of the features.

### **3.5 Model Development**

Multiple supervised machine learning models were made and trained on the preprocessed dataset. These four models (Logistic Regression, Random Forest, XGBoost, and Gaussian Naive Bayes) selected as models to be evaluated were chosen due to the ease of application in binary classification problems and because they have different learning mechanisms.

#### **3.5.1 Logistic Regression**

Logistic regression was used as a benchmark model since it is basic and straightforward to interpret. It estimates the probability of the membership to the classes by estimating a logistic

function based on the input data. The regularization parameter C and solver type were both tuned via grid search cross-validation to optimize the model's performance on the validation set.

### **3.5.2 Random Forest Classifier**

Random Forest is an ensemble learning protocol that creates several decision trees and combines their results. It is resistant to overfitting, and it provides variable importance information. Grid search was used to tune the model in respect of parameters such as n\_estimators, max\_depth and min samples leaf with stratified 3-fold cross-validation to confirm stability of performance of the model.

### **3.5.3 XGBoost Classifier**

XGBoost is a gradient boosting system which refines the model's performance with more errors. Its capability of investing missing values and unbalanced information by adopting scale\_pos\_weight made it the perfect candidate. The scale\_pos\_weight parameter was calculated with the ratio of the classes in the training data. To enhance generalization, grid search was used to tune three parameters learning\_rate, max depth and subsample.

### **3.5.4 Gaussian Naive Bayes**

The Naive Bayes classifier uses Bayes inference under the independent feature assumption adopted in the scenario. It is a simple algorithm that, quite frequently, works well with high-dimensional data. It was fit using the var\_smoothing parameter, and by its inclusion, an informative reference was had to the more complex models.

## **3.6 Model Evaluation**

A variety of metrics were used to test each trained model on the test set. These were accuracy, precision, recall, the F1 score, and the confusion matrix. The overall performance was evaluated in terms of accuracy, whereas precision and recall allowed gaining information about the rate of false positives and false negatives. The F1 score was a trade-off measure that addressed both.

The Random Forest and XGBoost models were the most accurate and scored high on the F1-scores and thus serve to be deployed. Logistic regression, slightly less in performance, was kept as having interpretability as well. Naive Bayes performed well but showed instability across different metrics.

### **3.7 Explainable AI (XAI) Integration**

Since tree-based models like Random Forest and XGBoost are black boxes, Explainable AI (XAI) methods were included in the paper to explain model predictions and make sure that the model worked transparently.

#### **3.7.1 Feature Importance**

Both XGBoost and Random Forest offered embedded feature importance scores, which were graphed to determine the range of variables that impacted the decisions concerning the approval of loans most strongly. Any variables of the features, including those assessed as credit score, intent to take a loan, loan amount, and personal income, were among the most prevailing predictors.

#### **3.7.2 SHAP Analysis**

To boost interpretability further, SHAP (SHapley Additive exPlanations) values were calculated. SHAP values will decompose the prediction into individual prefixes to display the contribution of each feature. Global SHAP summaries were applied so as to separate the most influential features in all the predictions, whereas the local SHAP plots were applied to clarify particular model decisions on the part of specific applicants.

Such interpretability supported fairness audits, which were the analysis of such sensitive features as gender, education level, and home ownership to verify that their contribution to the decisions was not disproportionate. This was critical to ethical deployment in the field of financial services.

### **3.8 Model Export and Deployment Readiness**

The last trained model (Random Forest) has been saved with the help of the joblib module to be deployed. Together with the model, fitted scalers and label encoders were also serialised. This was to ensure that whatever new data was subjected to inference was done in the same procedure as the training data ensuring consistency.

This setup allows the model to be used in real-time applications such as online loan approval systems, integrating seamlessly with existing financial platforms.

### **3.9 Ethical Considerations**

Ethical considerations were maintained throughout this research by anonymizing the data and paying special attention to determining the role of sensitive features in it. Despite a dataset having variables like education and gender, XAI tools were used to measure the fairness of the model regularly to check the possible biases. The idea was to prevent the patterns of discrimination and guarantee that the examples of models were recommended on financial grounds but not demographics.

Also, model design and logic transparency are made possible by SHAP and feature importance analyses allow user trust and regulatory compliance.

## Chapter 4 - Results and Discussion

### 4.1 Data Overview and Initial Analysis

#### 4.1.1 Dataset Snapshot and Feature Composition

To initiate the modeling process, the dataset with 45,000 records of loan applications was analyzed. In every row, there is a distinct application, and the feature table has 13 entry features, and a target variable (`loan_status`) is binary, taking the value of 1 when the loan is approved and 0 when the loan is not approved.

Dataset Head:

```
  person_age person_gender person_education person_income person_emp_exp \
0      22.0      female      Master      71948.0      0
1      21.0      female      High School      12282.0      0
2      25.0      female      High School      12438.0      3
3      23.0      female      Bachelor      79753.0      0
4      24.0      male      Master      66135.0      1

  person_home_ownership loan_amnt loan_intent loan_int_rate \
0          RENT      35000.0      PERSONAL      16.02
1          OWN      1000.0      EDUCATION      11.14
2      MORTGAGE      5500.0      MEDICAL      12.87
3          RENT      35000.0      MEDICAL      15.23
4          RENT      35000.0      MEDICAL      14.27

  loan_percent_income cb_person_cred_hist_length credit_score \
0          0.49          3.0          561
1          0.08          2.0          504
2          0.44          3.0          635
3          0.44          2.0          675
4          0.53          4.0          586

  previous_loan_defaults_on_file loan_status
0          No          1
1          Yes          0
2          No          1
3          No          1
4          No          1
```

*Figure 1 - Sample Records from the Dataset*

The features include:

- **Demographic:** person\_age, person\_gender, person\_education
- **Financial:** person\_income, person\_emp\_exp, loan\_amnt, loan\_int\_rate, loan\_percent\_income, credit\_score
- **Credit History:** cb\_person\_cred\_hist\_length, previous\_loan\_defaults\_on\_file
- **Loan Characteristics:** loan\_intent, person\_home\_ownership
- **Target Variable:** loan\_status

#### 4.1.2 Data Types and Structure

The dataset consists of:

- **6 numerical float columns**
- **3 integer columns**
- **5 categorical (object-type) columns**

```

RangeIndex: 45000 entries, 0 to 44999
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   person_age                                45000 non-null  float64
1   person_gender                             45000 non-null  object
2   person_education                          45000 non-null  object
3   person_income                             45000 non-null  float64
4   person_emp_exp                            45000 non-null  int64
5   person_home_ownership                    45000 non-null  object
6   loan_amnt                                 45000 non-null  float64
7   loan_intent                               45000 non-null  object
8   loan_int_rate                             45000 non-null  float64
9   loan_percent_income                      45000 non-null  float64
10  cb_person_cred_hist_length               45000 non-null  float64
11  credit_score                             45000 non-null  int64
12  previous_loan_defaults_on_file           45000 non-null  object
13  loan_status                              45000 non-null  int64
dtypes: float64(6), int64(3), object(5)

```

*Figure 2 - Dataset Info Summary*

The diverse structure required the application of encoding techniques and SMOTENC, as detailed in subsequent sections.

### 4.1.3 Descriptive Statistics of Numerical Features

Summary Statistics:

	person_age	person_income	person_emp_exp	loan_amnt	\
count	45000.000000	4.500000e+04	45000.000000	45000.000000	
mean	27.764178	8.031905e+04	5.410333	9583.157556	
std	6.045108	8.042250e+04	6.063532	6314.886691	
min	20.000000	8.000000e+03	0.000000	500.000000	
25%	24.000000	4.720400e+04	1.000000	5000.000000	
50%	26.000000	6.704800e+04	4.000000	8000.000000	
75%	30.000000	9.578925e+04	8.000000	12237.250000	
max	144.000000	7.200766e+06	125.000000	35000.000000	

	loan_int_rate	loan_percent_income	cb_person_cred_hist_length	\
count	45000.000000	45000.000000	45000.000000	
mean	11.006606	0.139725	5.867489	
std	2.978808	0.087212	3.879702	
min	5.420000	0.000000	2.000000	
25%	8.590000	0.070000	3.000000	
50%	11.010000	0.120000	4.000000	
75%	12.990000	0.190000	8.000000	
max	20.000000	0.660000	30.000000	

	credit_score	loan_status
count	45000.000000	45000.000000
mean	632.608756	0.222222
std	50.435865	0.415744
min	390.000000	0.000000
25%	601.000000	0.000000
50%	640.000000	0.000000
75%	670.000000	0.000000
max	850.000000	1.000000

*Figure 3 - Summary Statistics*

**Observations:**

- There are outliers in person\_income (max 7.2M) and person\_age (max 144 years) and these can influence model training and were handled through scaling and EDA.
- The **credit score** distribution appears fairly centralized, with a mean score of around 633.
- On loan\_percent\_income, there is mostly a conservative borrowing profile; however, there are also high-ratio applicants (up to 66% of their income is spent on loans).

**4.1.4 Target Variable Distribution**

The loan\_status variable is **heavily imbalanced**:

- **Approved (1):** ~22%
- **Not Approved (0):** ~78%

This is confirmed by:

Mean of loan\_status: 0.222, which is approximately 22% approvals

Such imbalance was confirmed to warrant the use of SMOTENC to balance the dataset by producing synthetic samples of minority class examples without disturbing the integrity of the categorical features.

#### 4.1.5 Missing Values Check

A complete examination of null values across all 14 features yielded the following:

Missing Values:

person_age	0
person_gender	0
person_education	0
person_income	0
person_emp_exp	0
person_home_ownership	0
loan_amnt	0
loan_intent	0
loan_int_rate	0
loan_percent_income	0
cb_person_cred_hist_length	0
credit_score	0
previous_loan_defaults_on_file	0
loan_status	0

*Figure 4 - Summary Statistics*

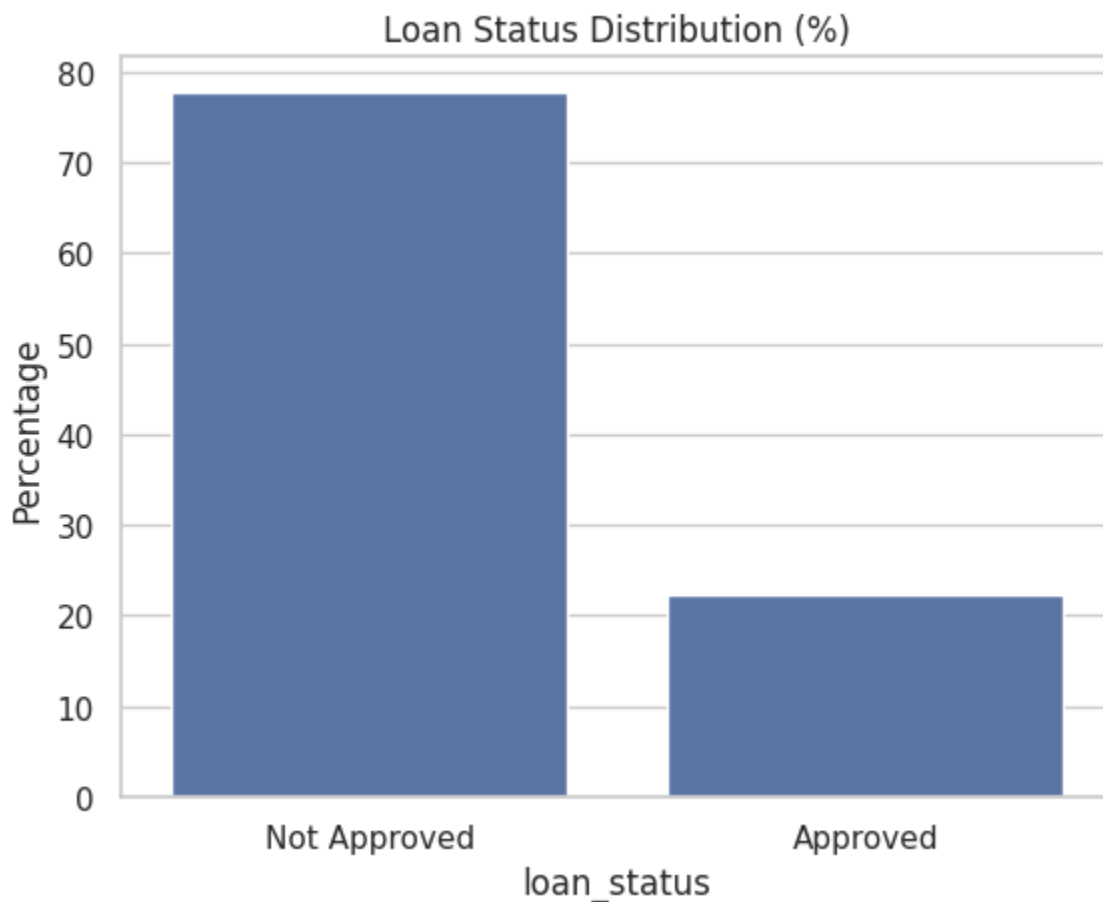
All columns: 0 missing entries

**Implication:**

- **No imputation** was required, ensuring consistent data integrity throughout the pipeline.

## 4.2 Exploratory Data Analysis

### 4.2.1 Loan Status Distribution



*Figure 5 - Loan Status Distribution*

One of the initial elements in terms of knowing how the data is structured is the analysis of the distribution of the target variable `loan_status` that specifies whether loan application was approved (coded as 1) or not approved (coded as 0). The bar chart in Figure 5 shows the relative percentages of each of the classes of the total data.

The visualization immediately reveals a high imbalance between the two classes in the given dataset. Approximately 22 percent of the loan application applications available in the data were accepted out of the 78 percent in the data set. This means that one out of four applications was accepted whereas three were rejected. Such biased distribution will have serious consequences to future machine learning algorithms.

Classification models learned on imbalanced data, without any methods of correction, have a bias in favor of the more populous class. Here a model trained naively on the original distribution may have high accuracy, simply because it classifies applications mostly as non-approved, and at the same time it may fail to point out good cases that it should approve. This is particularly problematic in financial decision-making, where fairness and opportunity in access to credit are essential.

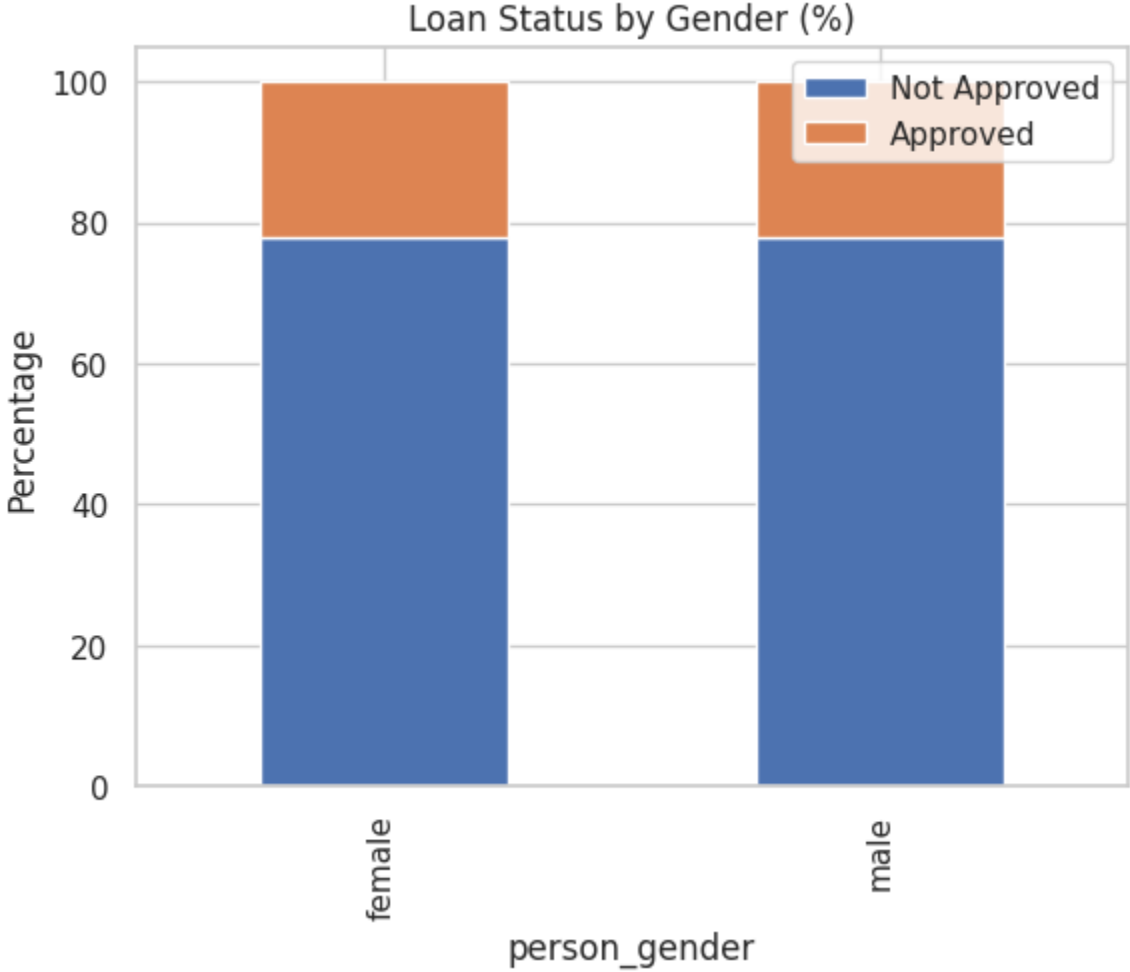
Imbalance also predicts there are potential patterns in the data that should not be neglected- e.g., is it possible that some applicant characteristics are systematically related to rejection (and thus the high rejection rates). This base-level means imbalance of the target variable meant resorting to class-balancing methods in later parts of the pipeline. In such analysis, the rebalancing of the training data using the Synthetic Minority Oversampling Technique, but on nominal and continuous data, was conducted after encoding the data and as such, the structure of categorical and numerical attributes was retained.

In addition to modeling issues, the imbalanced ratio of approvals also reflects on the risk attitude of the lending institution and its credit policies. A steep decline rate can be an indicator of conservative lending or a risk-averse approach, which may be caused by the fear of the level of income of the applicants or their credit history or loan purpose, which are further investigated in the subsequent visual analysis.

#### **4.2.2 Gender and Loan Approval Outcomes**

In a bid to examine possible gaps that may exist between loan approvals by gender and other demographic factors, the correlation between the gender of the applicant and the status of the loan was considered. Figure 6 exhibits the stacked bar chart, displaying the percentages of the loans approved and not approved for male and female applicants, respectively.

As illustrated in the visualization, there is a highly comparable distribution concerning the loan outcomes of the two gender groups. In both scenarios, a substantial percentage of the applications were never approved, with most of the cases constituting an inordinate amount of about 78 percent, with the remaining value being the approvals making up 22 percent. The bars representing both genders are nearly identical in height and composition, indicating that gender did not significantly influence the rate of loan approvals in the data overall.



*Figure 6 - Loan status by Gender*

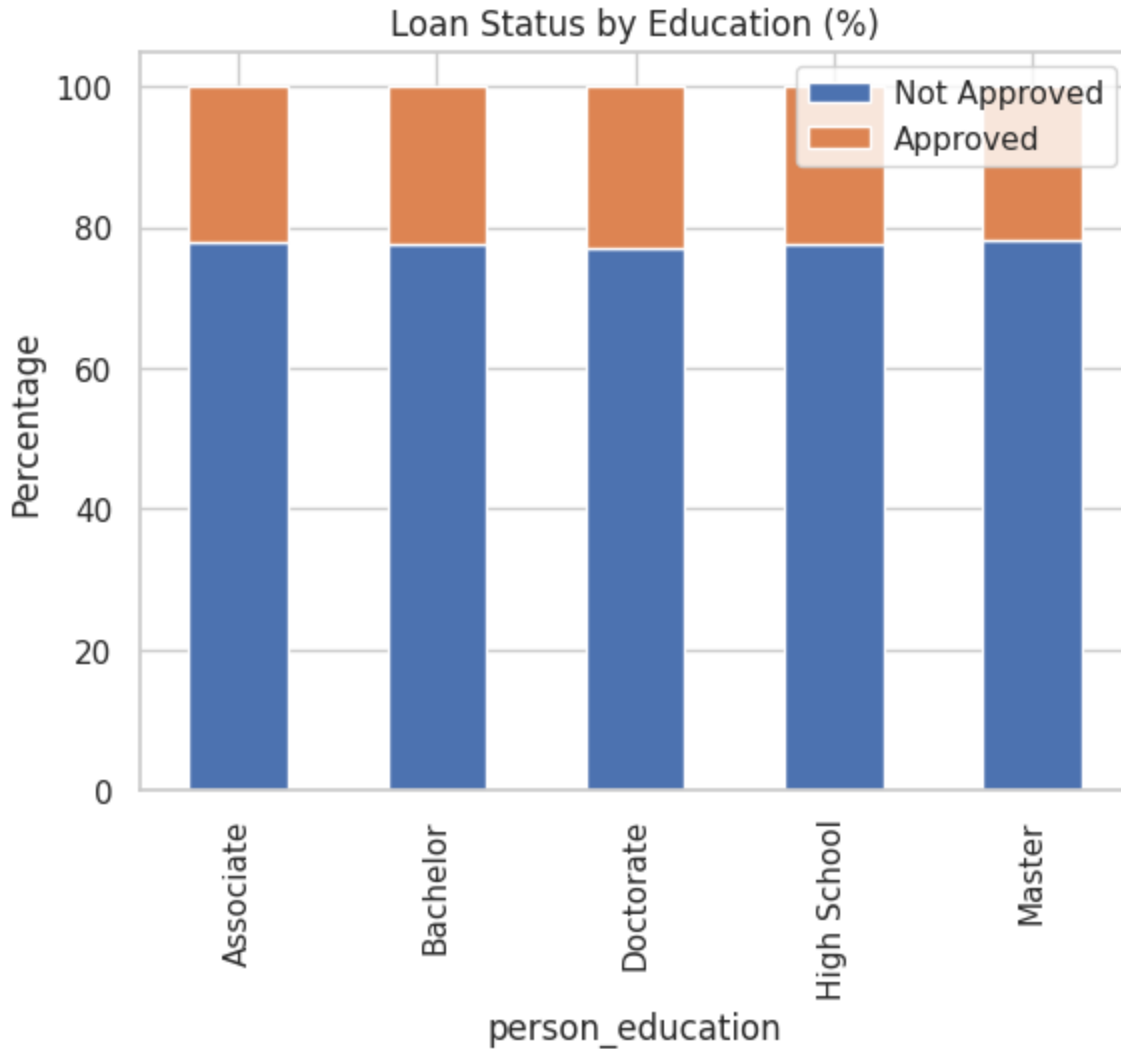
This finding is notable for two reasons. To begin with, it suggests a certain degree of equality in the evaluation of applications irrespective of gender. In past and in some present financial situations, women have been disadvantaged by the issue of not getting credit because they reported the low level of income, owning very little to no property, or serving as employees with lapses in the form of employment. However, the consistency in approval percentage of applicant

males and females in this study may indicate that such biases are not in the database or the logic of approval.

Second, it reconfirms the significance of applying the fairness-aware practices when developing the model. Referring to this project, gender value was originally included as one of the characteristics; however, explaining the model based on the SHAP values, it had no impact on its decision-making. In the financial system, this verification is critical since the requirements involved in regulation and moral responsibility require sensitive attributes not to be submitted in a manner that may result in discriminatory outputs.

Overall, this result provides encouraging evidence that the lending process, as captured in the data, maintained a relatively consistent standard of evaluation across genders.

#### **4.2.3 Education and Loan Approval Trends**



*Figure 7 - Loan Status by Education Percentage*

To reflect on whether there was any significant role that the educational attainment of an applicant would have on the approval of their loan application, this was also measured to determine the level of relationship between educational attainment and the outcome of a loan application. Figure 7 presents a stacked bar chart with every level of education being put separately with the proportion of applications approved and not approved.

This trend can be noted across all types of education, High School, Associate, Bachelor, Master and Doctorate. Most of the applications that were made irrespective of the level of education were declined, a factor that conforms on the trend exhibited by the whole system. Nevertheless,

one can see that a slightly higher approval rate can be observed among those applicants who have either a master's or doctorate degree, yet the difference is unrivaled.

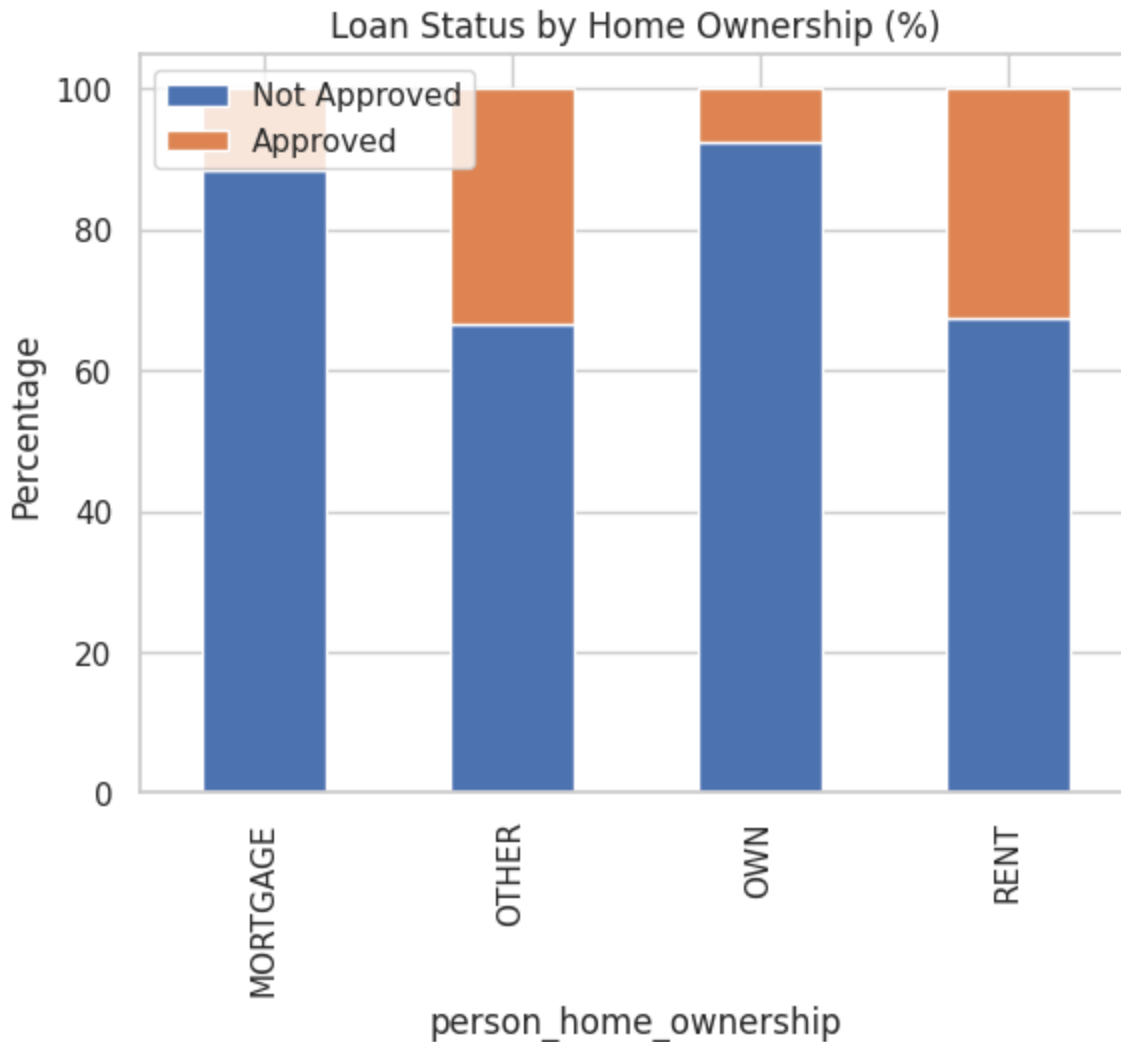
Such a small rise in approval percentages of those with advanced degrees might indicate that the lenders perceive higher education as leading to greater financial responsibility or earning power. Better qualified applicants may also have higher income levels, greater stability, and improved credit behavior, all of which could positively influence their credit ratings.

Interestingly, even those applicants who had a high school education had relatively similar rates of approval compared to the more educated. This indicates that education, while one of the factors considered, is not the primary determinant of creditworthiness. Other factors like credit score, availability of income, amount of the loan applied for, and debt-to-income are more likely to have more significance in the approval process.

One can also note that education level, despite its representation as a categorical feature in the modeling procedure, was not a leading predictor; lateral analysis of feature importance translated into SHAP values. This also contributes to the conclusion that, although educational attainment is relevant, it does not independently determine loan decisions in the dataset. Rather, it becomes the collective effort of several financial indicators to inform the model predictions.

#### **4.2.4 Home Ownership and Loan Status Patterns**

The other significant variable that was analyzed in the exploratory analysis was the ownership status of the applicant to the home. Owning a home may sometimes indicate money security, ownership of assets, or permanence of residence and can thus factor in the decision to lend. In Figure 8, loan approval results are indicated based on four ownership categories of home: mortgage, own, rent, and other.



*Figure 8 - Loan Status by Homeownership*

The chart shows that Own applicants have the highest loan rejection rates. More than 90 percent of the loans applied for by this group were denied, which is higher than the rejection rates for other homeownership guarantees. This finding is surprising because home ownership is usually associated with financial security, which is thought to help with creditworthiness. However, the data in this case does not support this conclusion, as it may be influenced by confounding variables such as income, predetermined debt requirements, or credit history that might be less favorable in this segment.

Conversely, those falling in the category of being designated as Other and Rent have higher rates of approval. The share of applications that have been approved is relatively lower in these These

groups account for one-third of the total population. This diversity also implies a slight improvement in the loan outcome. Applicants with "Mortgage" status fall between these extremes, with approval rates that appear moderate but closer to those who rent.

There is a possibility that the mortgage holders already have outstanding debt levels; thus, the loans would pose more risk as a lender. In the meantime, tenants can be a more varied pool and not necessarily have higher incomes and lower obligations on their assets, and this can at times help them be nicer loan prospects when the remaining financial signs are in good order.

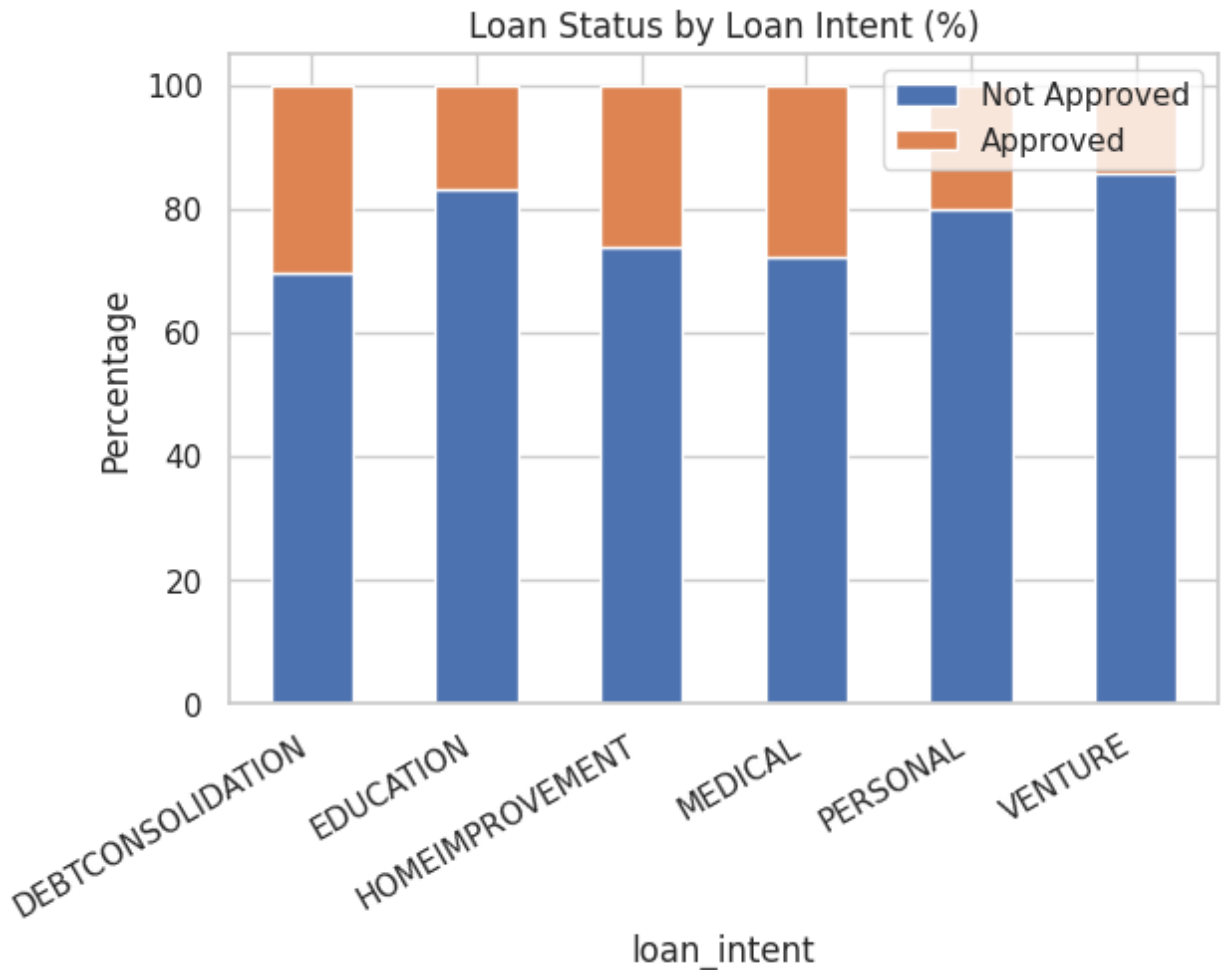
There is a possibility that the mortgage holders already have outstanding debt levels; thus, the loans would pose more risk as a lender. In the meantime, tenants can be a more varied pool and not necessarily have higher incomes and lower obligations on their assets, and this can at times help them be nicer loan prospects when the remaining financial signs are in satisfactory order.

#### **4.2.5 Loan Intent and Approval Outcomes**

The intended use of the loan or the purpose quoted at the time of the application is a fine contextual variable in lending. It does not only indicate the desires of the borrower but also gives the lender an idea about the type of risk he is taking. Figure 9 indicates the percentages of loan approvals and rejections of 6 categories of intents, namely, debt consolidation, education, home improvement, medical, personal, and venture.

Among them, the debt consolidation application was the most approved, with about one-third of the people who applied in this line showing a positive result. This indicates that lenders can be willing to lend money to customers seeking to control or reinstitute the current debt, particularly when the clients can show the possibility to pay back the money and have a good credit rating.

Conversely, the rates of approving education and venture loans were the lowest, as the percentage of those approved varied considerably less than other categories. This trend can be possibly affected by the perceived danger in such intents. The education loans are usually provided to the young customer, who may not receive a stable income or have well-built credit history, whereas loans connected to ventures are always viewed as high-risk loans, as there is no guarantee on returns and there is no developed stream of income.



*Figure 9 - Loan status by Loan Intent Percentage*

Applications belonging to the personal, medical, and home improvement intents were moderately approved with slight disparity between each. These categories are more neutral about risk and more heavily rely on personal financial statistics over consideration of the intent itself.

Overall, the chart indicates that loan intent contributes to differences in approval outcomes, but it is not the sole factor. In the machine learning models applied to this work, the loan intent was offered as a categorical variable and the later interpretation analysis of Shap values projected that in spite of the quantitative differences brought about by the intent, the factors of the finances, including the credit score, income, and loan amount, were more revealing than the intent itself.

These findings reinforce the idea that lenders do consider the purpose behind the loan, but only in conjunction with a broader set of financial features when making decisions.

### 4.2.6 Age Distribution and Its Association with Loan Status

Applicant age is a fundamental demographic variable often considered during credit evaluation processes. Age does not have a direct impact on being creditworthy but may be correlated with other economic aspects e.g. income stability, work history, credit behavior. Figure 10 presents a distribution plot comparing the age distribution of approved and non-approved loan applicants.

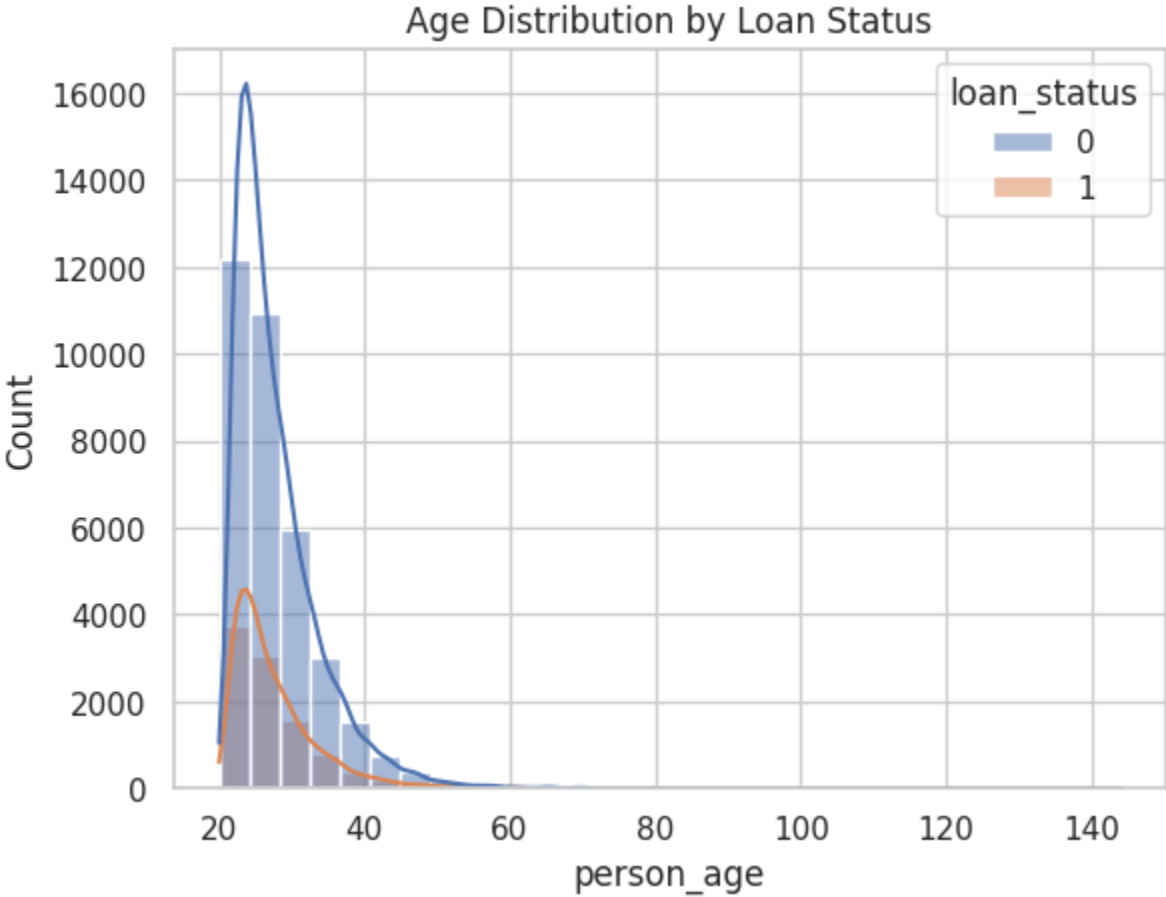


Figure 10 - Age Distribution by Loan Status

The chart indicates that most loan applications, whether positive or negative, are skewed towards the youthful age groups, who are mostly between 21 and 35 years old. Such a skewed distribution would generally be representative of the data set, as most applicants are in the early stages of adulthood or are young professionals.

Inspecting the 2 distributions, approved loans are distributed slightly younger on average, and approval density increases around the 26-age mark. In comparison, the spread of the non-approved applicants would be rather more than that, yet they lie in the same age range. This indicates that while age may have some influence, it is not acting as a clear dividing factor in loan approval outcomes.

Both the number of approvals and the number of rejections decline drastically past 40 years of age. This tendency can indicate the possibility that older people either borrow less often or constitute a smaller part of the sample. Another possibility is that older applicants may already have credit relationships and financial ties entered; thus, they do not have the need or space to avail themselves of newer loans.

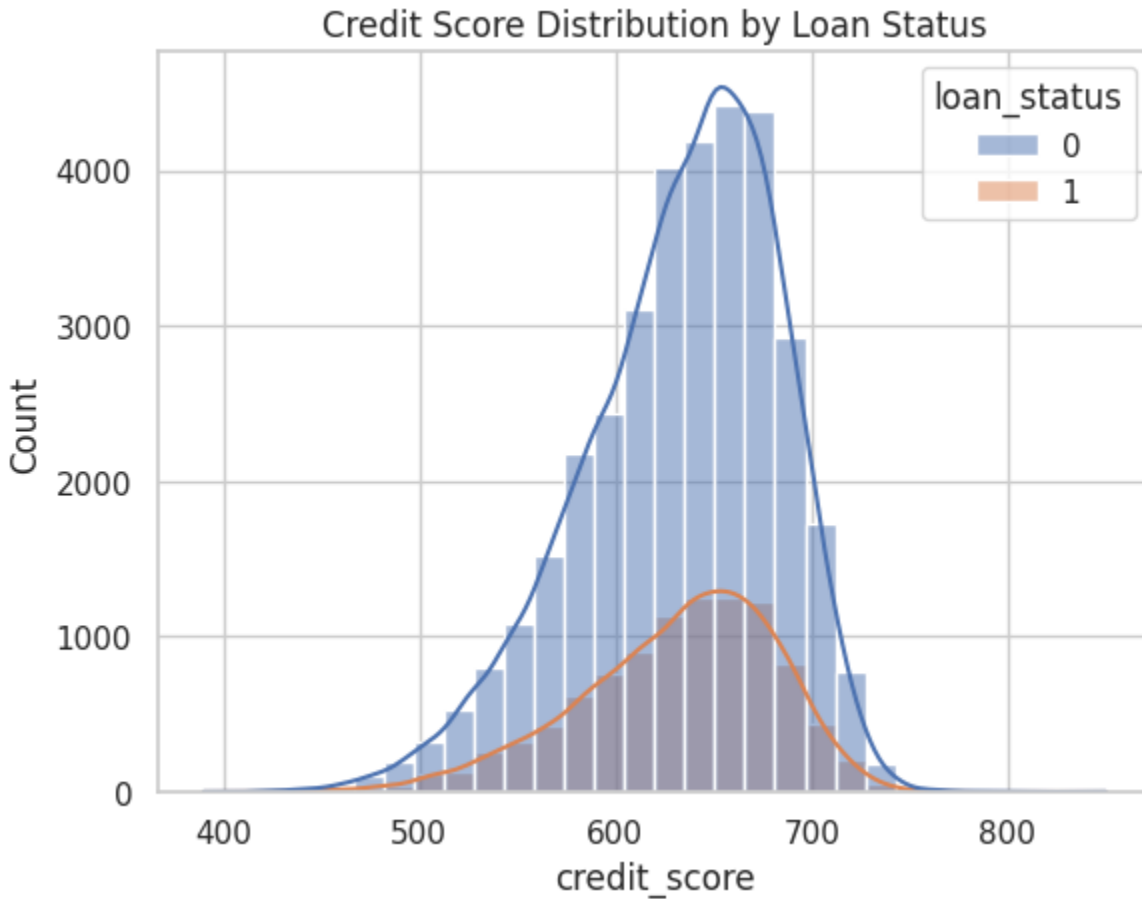
The age parameter that was used in the predictive modeling process was continuous and was in the figures. However, further analysis of the model helped to prove that age was rather uninfluential in relation to the rest of the parameters including credit score, loan amount, and income. This implies the conclusion that age can possibly correlate with such model predictions as the general credit profile, but it is not a core driver.

The findings reinforce that demographic characteristics like age must be interpreted in context, particularly when financial data offers more precise and actionable insights into risk.

#### **4.2.7 Credit Score Distribution and Approval Outcomes**

Credit score plays one of the leading roles when it comes to lending decisions, as it is a condensed version of the overall credit history of the person and their financial conduct. It summarizes the information about the history of previous debt, payment trends, and their total credit usage. Figure 10 demonstrates the allocation of credit scores of loan applicants who are and are not approved.

It is clear by the visual that the correlation between credit score and the probability of loan approval is high. Approvals are also concentrated closer to the higher limit of credit scores, and a clear peak is visible between 640 and 700. On the other hand, rejected applications have a wider and slightly left-skewed distribution with a high point around 620 mark, but extending further down the credit scores.



*Figure 11 - Credit score distribution by loan status*

Notably, very few accepted applicants had a credit score lower than 580, indicating some form of a cutoff or threshold effect at some stage of the credit decision process. On the one hand, however, the pool of rejected applicants runs the gamut of credit scores, including those at the middle or even high end. This dispersion shows that credit scores are not the only determinant of approval, being an important factor.

The central tendency of the qualified group indicates that lenders can favor the applicants that exist somewhere in the range between moderate and high credit thresholds although they do not really need very high scores. This is according to the traditional lending standards, which scores in the mid-600s can be defined as reasonable or near prime and allow access to funding under standard conditions.

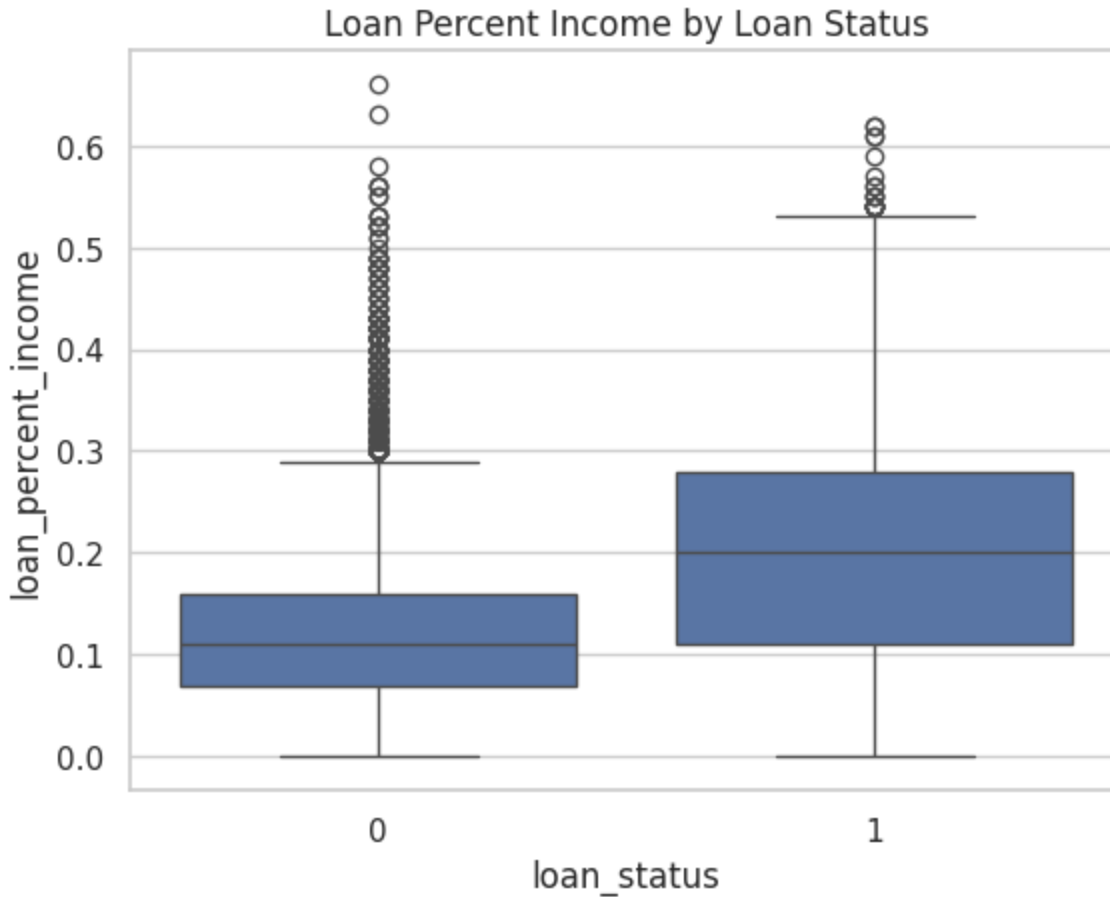
As far as the modeling aspect is concerned, credit scores turned out to be one of the most significant features in loan status prediction. Analysis with SHAP values affirmed that changes in credit scores influenced the predictions of the model substantially, especially when the score was outside the median of the dataset.

The chart thus adds to the conclusion that credit score is one of the central decision variables within the pipeline of loan approval. The predictive power of this indicator, however, rises on being interpreted in coalition with other financial indicators, including the level of income, debt ratio, and employment experience.

#### **4.2.8 Proportional Income Commitment and Loan Status**

loan\_percent\_income which is the ratio of loan amount to income, is a very important measure of the financial burden on an applicant. It entails the percentage of the monthly or annual income that an individual would pay to repay the loan. Figure 12 shows how this ratio is distributed through the approved applications and the non-approved ones with the box plots so that the central tendency, spread, and outlier behavior can be directly compared.

The chart demonstrates that the approved applicants are more likely to repay their loans with a larger percentage of their income than the ones who were not approved. The median of the approved individuals is clearly larger, and the interquartile range of the approved individuals is between 0.12 to 0.28. It implies that many approved borrowers are acquiring loans that comprise between 12 and 28 percent of their earnings, reflecting confidence on the part of the borrowers as well as laxity on the part of lenders.



*Figure 12 - Loan Percent Income by Loan status*

In contrast, non-approved applicants tend to be more conservative, exhibiting a median ratio of 0.10, and most of these applicants fall within the lower half of the ratio scale. Although this might suggest that it faces less risk, it might as well indicate the systemic filtering of those who do not pass through with high ratios of income to loans. Interestingly, even when the median values are more moderate, there still is a clump of outliers within the distribution of non-approved applicants at a higher level of ratios, which shows that they may be a warning sign that the debt-to-income ratio is too high.

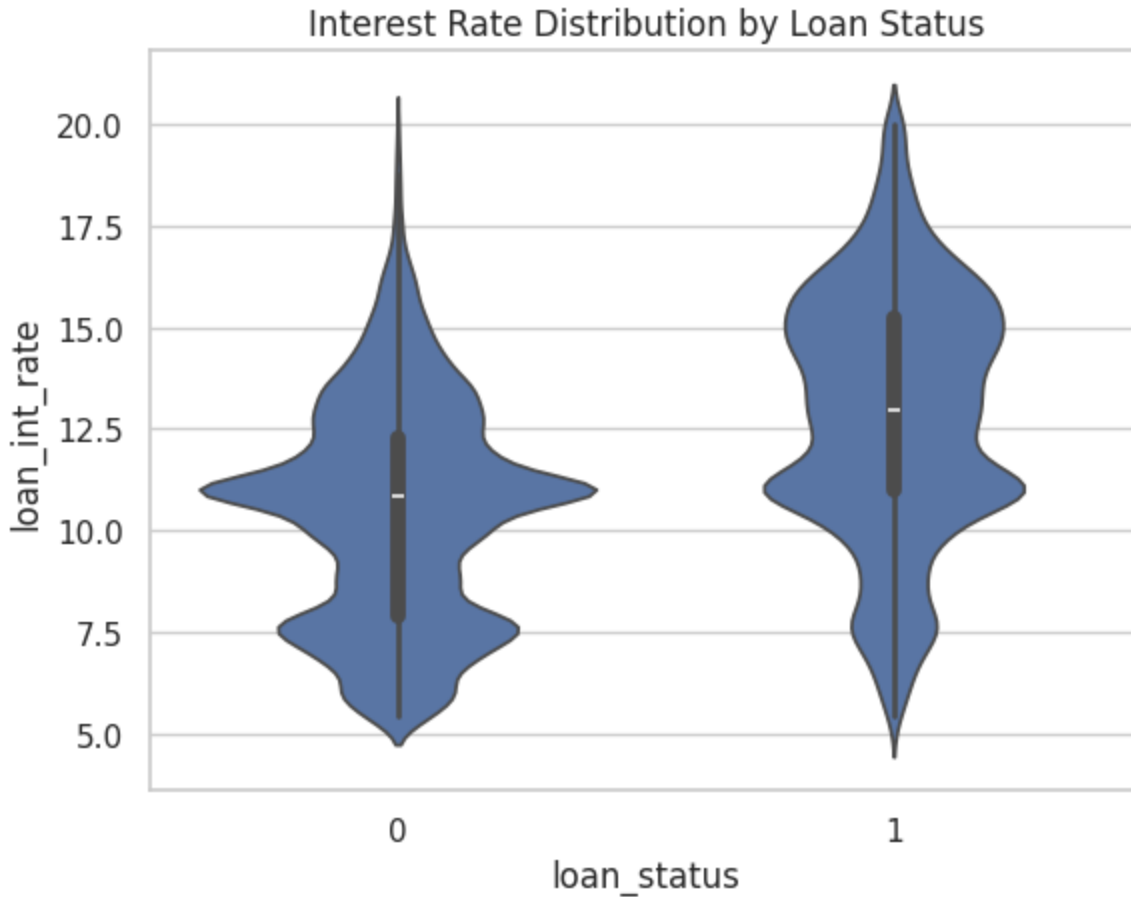
The availability of higher approval limits on higher loan\_percent\_income values may either suggest that the lenders have more confidence in the ability of the applicant to repay the loan or risk tolerance in some circumstances. But, from a risk management perspective, even large debt burden ratios could reflect an exposure to stress in the financial condition, particularly in a turbulent income environment.

This variable held moderate significance in classification within modeling, particularly when combined with absolute income and the loan amount. The SHAP summary plots showed that the values of `loan_percent_income` had a positive contribution toward prediction of approval to a certain level, then the contribution of the latter was reversed.

This graph further supports the idea that proportional loan burden is not the determinant of loan performance, but it provides substantial context when it is measured together with other financial indicators.

#### **4.2.9 Interest Rate Patterns and Approval Outcomes**

Interest rate, or `loan_int_rate` is an imperative factor determining the desire of the applicant to borrow money and expectations of the lender in terms of his risk-adjusted returns. In figure 3, violin plots are used to display interest rates of approved and non-approved loan applications that give a sense of density and centrality to each group.



*Figure 13 - Interest rate distribution by loan status*

The distributions reveal the fact that higher interest rates were more likely to be given to approved applicants than those that were not approved. The center of the violin plot of approved loans is concentrated around 13 to 15 percent, whereas the rejected group is focused on 10 to 12 percent. This difference suggests the perverse pattern that more interest-bearing loans were approved more often.

There are a few reasons that can be assumed behind this tendency. It could be that the loans that have been approved at a higher interest rate reflect a higher level of risk by the borrower and yet they were considered creditworthy maybe because of good collateral, steady job, or good credit record. It is possible that lenders have advanced such customers based on the presumption that the associated risk would be offset by the higher interest rates. On the other hand, those rejected can be the group of applicants that would have wanted a lower interest rate but did not have the requisite credentials to warrant the sought terms.

Market segmentation may be another reason. Lenders may design products to suit various risk bands, with higher interest rate loans being approved for near prime or subprime borrowers, and marginal applicants with poor credit still being denied loans at a lower rate. This may lead to a twofold effect, which is evident here, where both the low-risk and high-risk applicants exist, but only those that conform to the acceptable pricing models are approved.

In terms of modeling, interest rate was relatively significant but, in many cases, interacted with additional characteristics, including credit score and the size of a loan. SHAP values indicated that the results of predictions were affected by extreme values, both high and low, but the occurrence of a moderate high interest rate was usually consistent with approval.

Overall, the distribution indicates that the interest rate is not just an outcome of approval, but it is an element of the approval mechanism. This qualifies it to be an important attribute to consider in predictive modeling, especially in lender strategy and borrower segmentation.

**Random Forest Performance Evaluation**

```

Random Forest
Best Params: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Accuracy: 0.941
[[6520 480]
 [ 346 6654]]

```

	precision	recall	f1-score	support
0	0.95	0.93	0.94	7000
1	0.93	0.95	0.94	7000
accuracy			0.94	14000
macro avg	0.94	0.94	0.94	14000
weighted avg	0.94	0.94	0.94	14000

*Figure 14 - Random forest performance matrices*

Random Forest model demonstrated a very good reliability of performance showing an accuracy of 94.1% while n\_estimators=100, min\_samples\_leaf=1, min\_samples\_split=2 and max\_depth=None were identified as its hyperparameters. The model was very well balanced in both classes with f1-score 0.94 in both cases of loan-approved (class 1) and loan-not-approved (class 0). The precision and recall of the two classes ranged from 0.93 to 0.95, meaning that the model can reduce the false positives and false negatives. Notably, the confusion matrix

demonstrated fairly low numbers of misclassifications: 480 false positives and 346 false negatives with all predictions being of 14,000.

These satisfactory scores indicate that Random Forest performed the complexities within the dataset well. It considered procedures in both positive and negative cases without instituting significant bias. The similarity of the macro and weighted averages, both equal to 0.94, demonstrates stable results despite the class balancing. Considering such outcomes, this model is resistant to problems in real-life decision-making cases since fairness and predictive power are decisive.

### Logistic Regression Performance Evaluation

```

Logistic Regression
Best Params: {'C': 10, 'penalty': 'l2', 'solver': 'liblinear'}
Accuracy: 0.8486666666666667
[[5801 1199]
 [ 163 1837]]

```

	precision	recall	f1-score	support
0	0.97	0.83	0.89	7000
1	0.61	0.92	0.73	2000
accuracy			0.85	9000
macro avg	0.79	0.87	0.81	9000
weighted avg	0.89	0.85	0.86	9000

*Figure 15 - Logistic Regression performance matrices*

The logistic regression model, with C=10, penalty: l2, and solver: liblinear, achieved an overall accuracy of approximately 84.87%. There is a tremendous gap in performance measures between the two classes. In class 0 (not approved), a very high precision was indicated as 0.97 which meant that when the model indicates that a loan may not be approved, it is very reliable. But the recall of this class was 0.83, meaning the model was missing a proportion of non-trivial true cases that were not approved. Conversely, there was an opposite trend in class 1 (Approved). It had high recall (0.92) with most of the actual approved loans being identified as approved but very low precision (0.61) and therefore tended to misclassify a lot of the non-approved loans as approved.

The F1 scores of the two classes also prove this: 0.89 of class 0 and 0.73 of class 1, which evidences the higher accuracy in terms of predicting that approval is not going to take place. The performance is skewed towards classes, as seen in the macro-averaged metrics (precision: 0.79, recall: 0.87, f1-score: 0.81), but weighted averages (precision: 0.89, recall: 0.85, f1-score: 0.86) look more promising. These results suggest that the model is overall satisfactory, but it is over-predicting loan acceptance, which may be dangerous in high-stakes financial use. It may be possible to increase reliability by fine-tuning thresholds or by using class-balancing methods, with a particular effect on the minority (approved) class.

**XGBoost – Performance Evaluation**

```
XGBoost
Best Params: {'learning_rate': 0.1, 'max_depth': 8, 'n_estimators': 150, 'subsample': 0.8}
Accuracy: 0.9194444444444444
Confusion Matrix:
[[6481 519]
 [ 206 1794]]
Classification Report:
              precision    recall  f1-score   support

     0         0.97         0.93         0.95         7000
     1         0.78         0.90         0.83         2000

 accuracy                   0.92         9000
 macro avg          0.87         0.91         0.89         9000
 weighted avg       0.93         0.92         0.92         9000
```

*Figure 16 - Logistic Regression performance matrices*

XGBoost classifier was tuned using these hyperparameters: learning\_rate = 0.1, max\_depth = 8, n\_estimators = 150, and subsample = 0.8. This config gave a total accuracy of about 91.94% on the testing set, a high generalization. The confusion matrix shows that the model correctly predicted majority class (label 0 - not approved) 6481 as true negatives and 519 as false positives and minority class (label 1 - approved) 1794 as true positives and 206 as false negatives. In the classification report, the performance is strong with F1 values of 0.95 and 0.83 on the majority and minority classes, respectively.

Class-wise precision was high, 0.97 in class 0 and 0.78 in class 1, demonstrating the model's capability to ensure that the false positive is minimized without an excessive loss in true positive in the minority class. The recalls were 0.93 and 0.90 in both classes 0 and 1, showing a high

sensitivity to the results of approval and rejection. The macro average F1-score was 0.89, and the weighted average F1-score was 0.92, with the predictive power being well balanced between both categories of loan statuses. All these findings indicate that the XGBoost model is an efficient way of addressing the class imbalance problem without too much accuracy and overall fairness penalty.

### ROC Curve

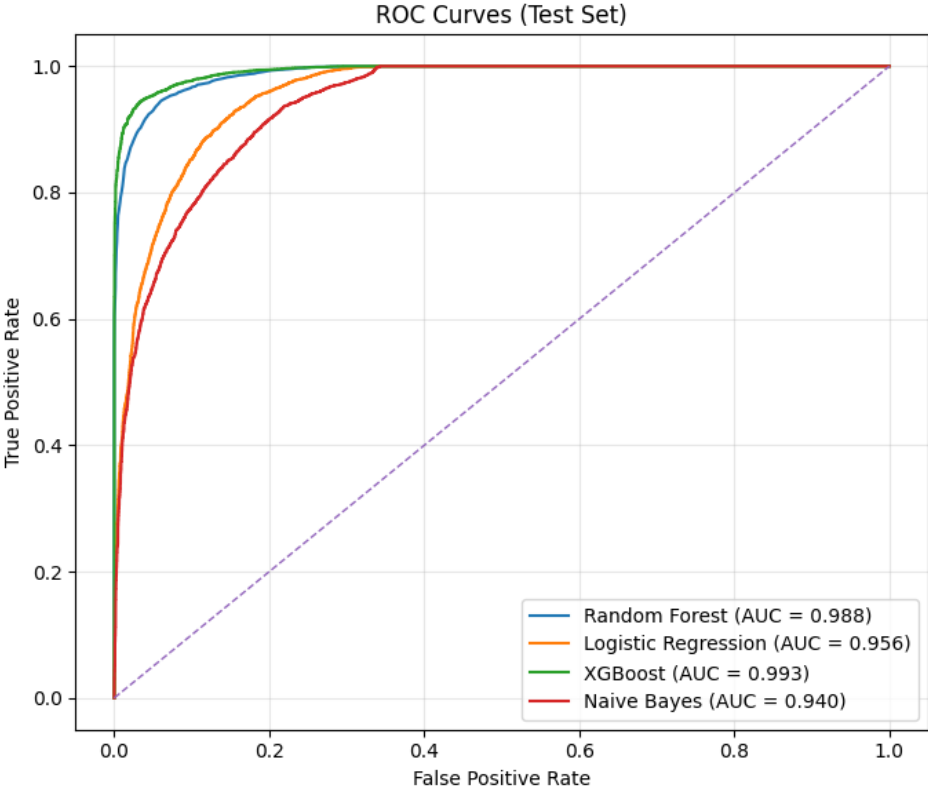


Figure 17 - Consolidated ROC Curve

The ROC curve (Figure X) demonstrates how various models can classify loans between accepted and rejected loan applications. The model that outperformed the others was the XGBoost model, with its score at 0.993, and it has almost perfect classification performance and thus the best choice to predict loans that will be approved. Random Forest also illustrated exemplary value with an AUC of 0.988 indicating a great strength to capture non-linear ties in the features of applicants. Logistic regression (AUC = 0.956) was chosen as describing the best baseline since the results are also straightforwardly interpretable, with Naive Bayes (AUC =

0.940) being the poorest performer yet still showing a respectable ability to distinguish between cases and controls. Comprehensively, this evaluation indicates that ensemble algorithms, especially the XGBoost, will provide excellent predictive performance on the loan approval scenario, thereby being more reliable and fair to judge individuals in terms of loan acceptance or refusal.

### Feature Importance

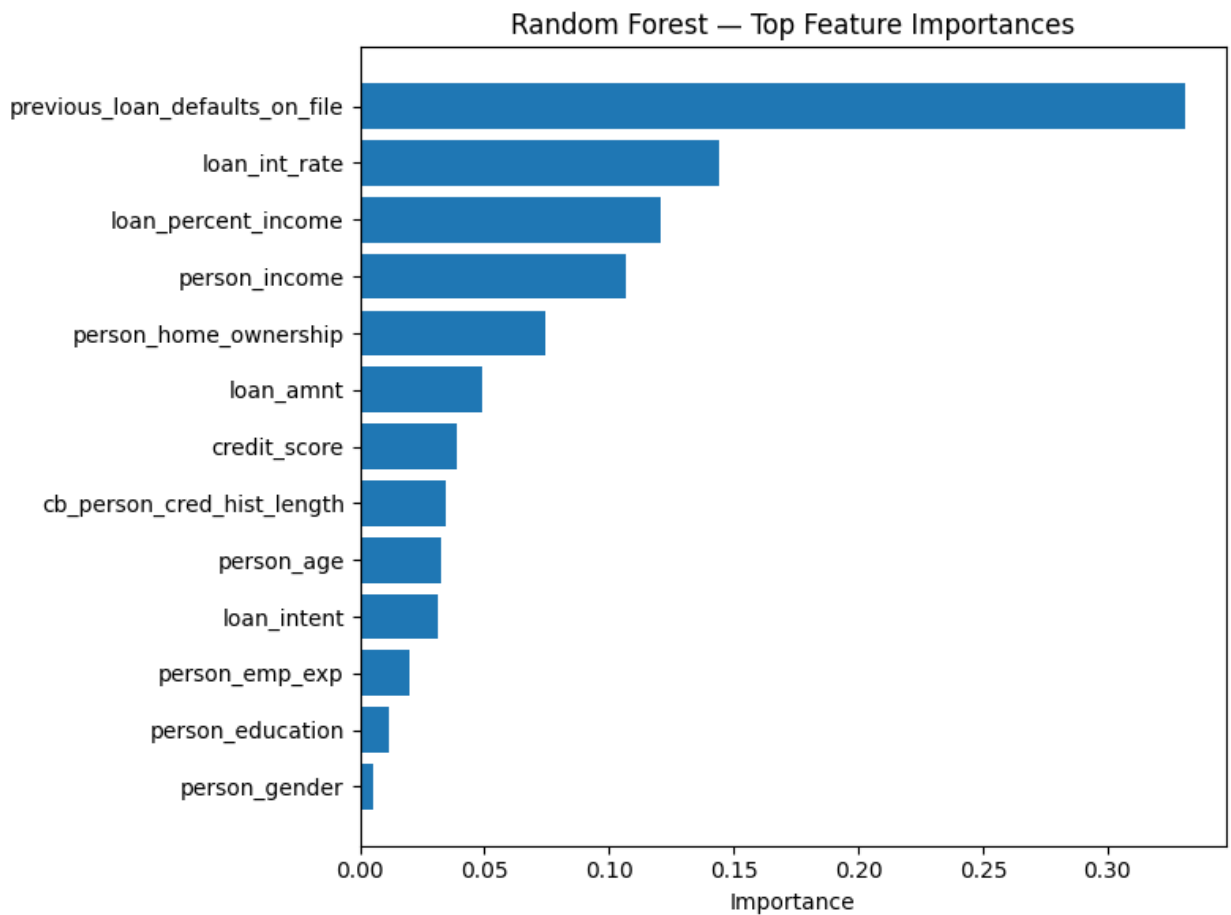
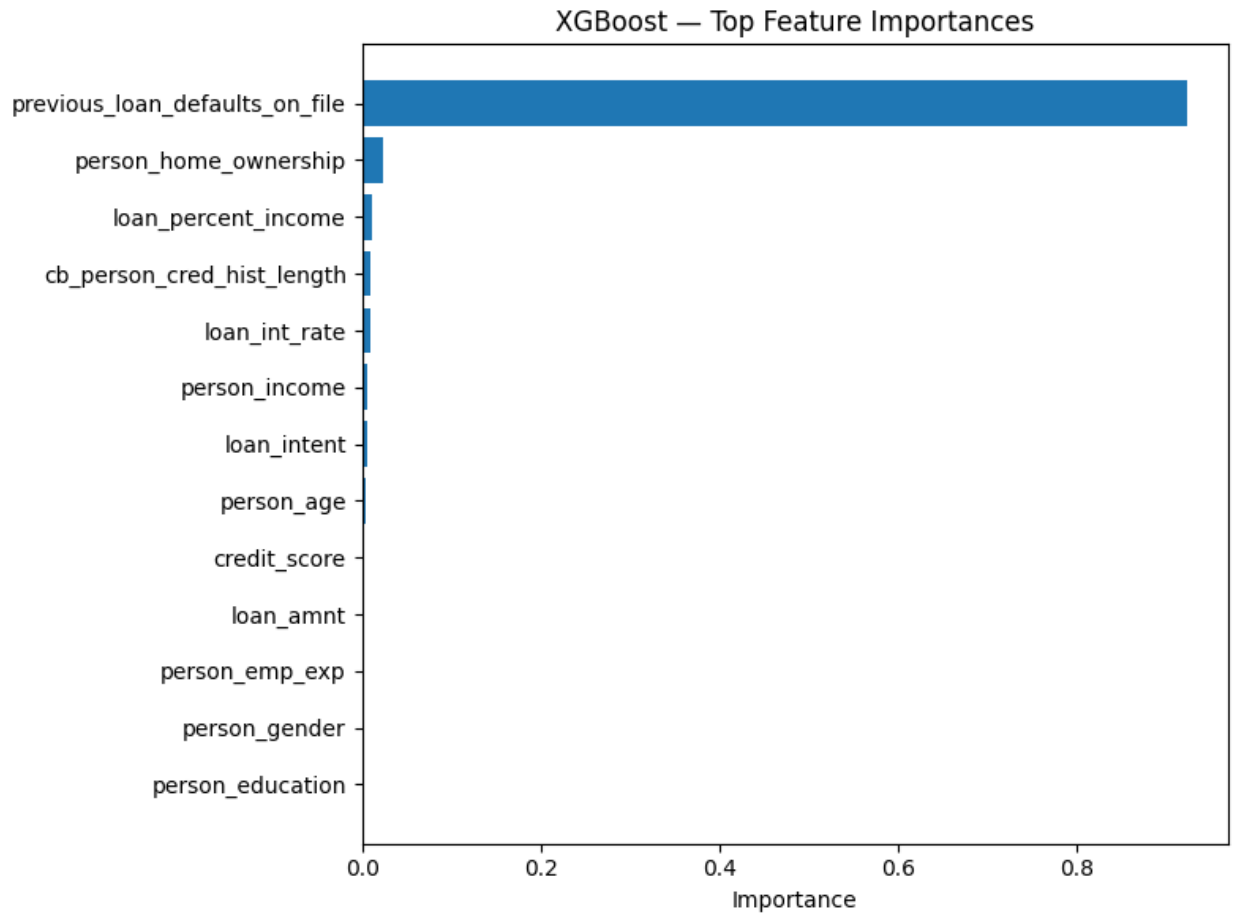
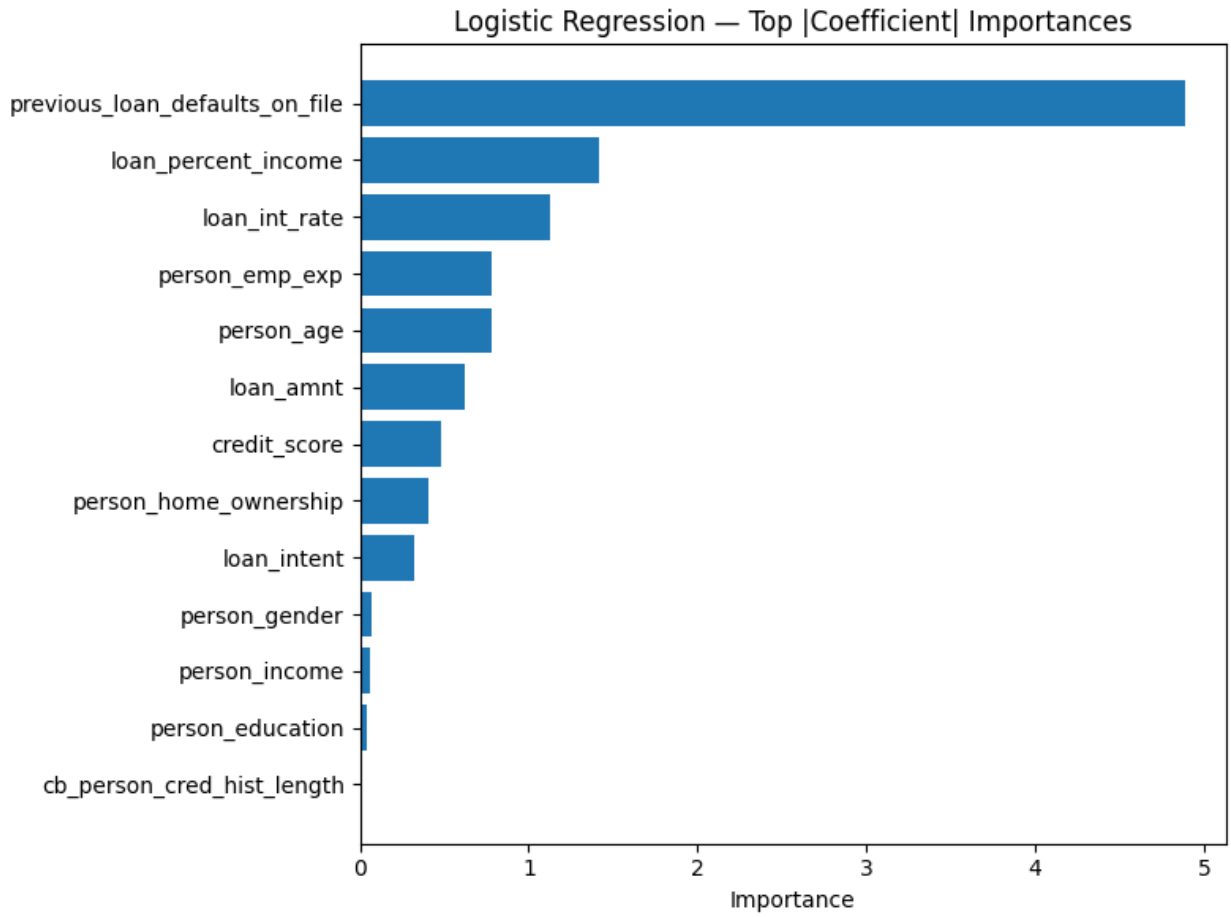


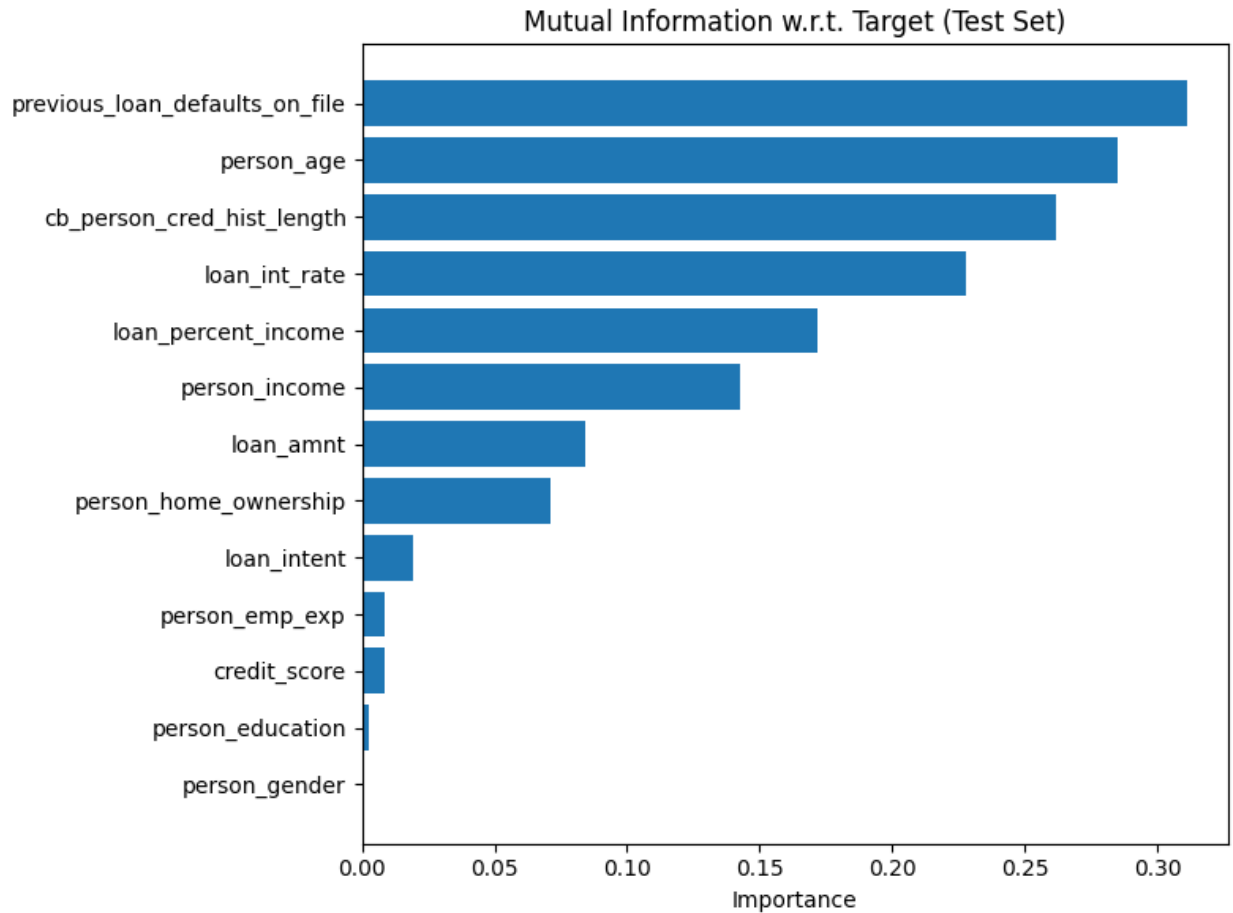
Figure 18 - Random Forest Feature Importance



*Figure 19 - XG Boost Feature Importance*



*Figure 20 - Logistic Regression Importance*



*Figure 21 - Mutual Information Naive Bayes*

The cross-model commonality observed was that past loan defaults on file were the strongest predictors of whether the loan would be approved or not across all models, which is consistent with the industry norms and the intuitiveness that the prior behavior of credit is the best indicator of repayment intentions. Random Forest assigned a less important distribution to features such as the interest rate on a loan, the percentage of income that individuals plan to spend on loans, and the income of loan applicants, which demonstrates the model's superiority in identifying a broader range of financial key points. Logistic regression highlighted more or less the same characteristics, having high weights on defaults, loan to income and interest rates, making their results readable and consistent with credit scoring rationale.

Meanwhile, XGBoost was much more focused on assigning extreme significance to loan defaults prominently, which implies that its boosting architecture preferred it to dominate the feature

model. But the analysis of mutual information gave a slightly different picture where defaults, age and credit history length are at the top of the list with financial ratios and income noted. In sum, these results confirm that constellation and boosting models can optimize predictive performance, but that when it comes to fairness and transparency in loan enhancements, a blend of statistical (mutual information) and understandable (logistic regression) decision models can only offer a more balanced explanation of what led to those decisions.

### SHAP Force Plot for a Single Prediction



Figure 22 - Force Plot

The above SHAP force plot provides a concise visual representation of the contribution that various input features had on a particular loan approval prediction. Beginning with a starting value of about 0.499, the model at last forecasted a sizeable likelihood of borrower approval, and the output was assertively nudged toward the value of 1.00. The latter change was largely caused by several robust positive contributors. The fact that there were no loan defaults in the past and that the applicant enjoyed favorable house ownership status contributed to the likelihood of approval especially. Moreover, the amount of the loan and relative proportion of income reserved to the loan (loan\_percent\_income = 0.4) also had a positive contribution, so being within the training distribution of the model, this combination could have been frequently coupled with approved applications.

Conversely, high income (82,440) of the applicant was the only aspect that marginally dragged the prediction down. In spite of the general correlation between the strength of repayment capacity and the level of income, higher income in the given case provided very slight counterbalancing effects. This could be attributed to the interrelations with other variables such as the high loan percent income which could be an indicator of possible over-leveraging though a high salary. Conclusively, the SHAP visualization is very helpful in terms of interpretability because they put a quantitative value to the contributions of each feature, so the stakeholders are

not only shown the result of the model, but also how the model came to that result. Such explainability is particularly handy in sensitive areas like lending decision-making which involves trust and accountability aspects.

### Integration into Streamlit app

After the development of models, the best-performing one was exported and is then converted into a Streamlit application to make it useful for public testing. Figure 18 shows the interface of this application

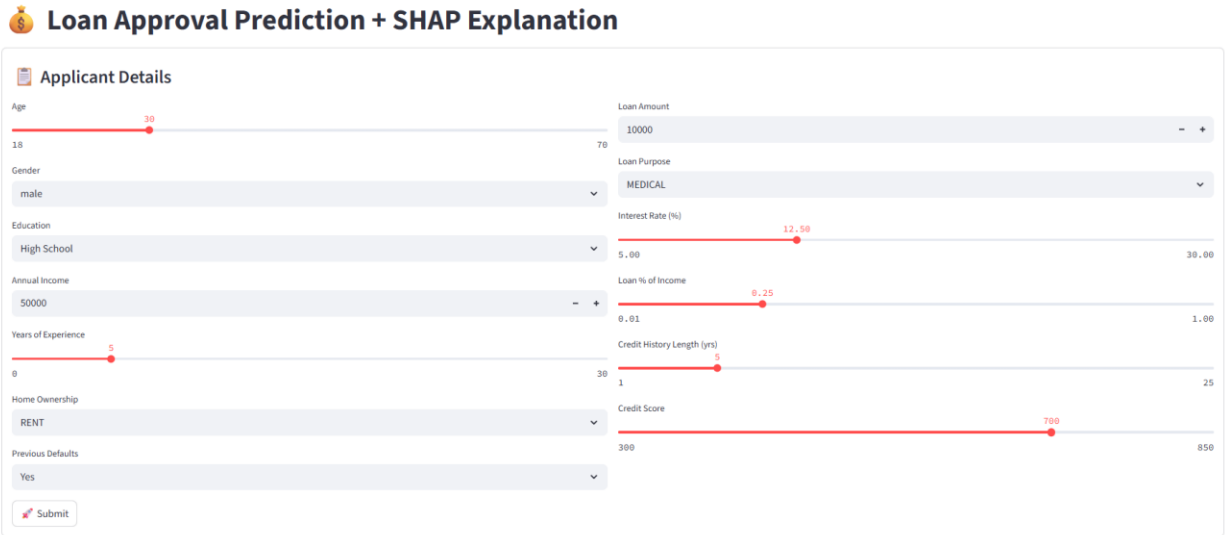


Figure 23 - Streamlit Application

## Chapter 5 - Conclusion and Future Work

This study aimed at forecasting loan approval based on a diverse set of applicant-related characteristics, such as income, credit score, home ownership, employment type and loan intent. A number of machine learning models were developed and compared, such as Logistic Regression, Random Forest, and XGBoost. Among these, the best performance was XGBoost which obtained an accuracy of more than 91 percent with balanced precision and recall in the approved and not approved classes. Application of SHAP values also introduced additional interpretability, such as the identification of features that would most impact individual decisions. The two major influence factors identified to have a high impact on the results of approval were loan-to-income ratio, prior loan default, and the composition of home ownership.

The findings would indicate that machine learning offers an effective solution to automating the process of loan approvals and creating a sense of transparency too. This system can assist in more equal and uniform lending activities by clearly defining the attributes of the applicants that differentially influence the decision-making process. Such models may help financial institutions to better assess risk, streamline operations, and make data-driven explanations to build trust among applicants.

### Future Work

Despite the current model's strong performance, there are numerous areas for improvement. A potential idea is to add new dynamic data like past financial activity or changing credit trends. This would enhance the model's ability to capture patterns that the static features would overlook. Another area of focus could be improving the approach to class imbalance in order to reduce minor biases in predictions and enhance equity among different groups of applicants.

The potential direction of future work can also be related to fairness-aware modeling approaches that do not depend on such demographic features as gender, age, or educational level to make consistent predictions. They may monitor their models to assess performance over time and identify changes in data trends. Lastly, adding an element of feedback on the model based on real-world results of approved and rejected loans would allow the model to be improved over

time, becoming more dynamic and adaptive to changes in borrower behavior and economic conditions.

## References

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. MIT Press.  
<http://fairmlbook.org/>

Breiman, L. (2001). Random Forests. *Machine Learning*. 45. 5-32. 10.1023/A:1010950718922.

Brown, Iain & Mues, Christophe. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst. Appl.*. 39. 3446-3453.  
10.1016/j.eswa.2011.09.033.

Buda, Mateusz & Maki, Atsuto & Mazurowski, Maciej. (2017). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. 106.  
10.1016/j.neunet.2018.07.011.

Chawla, Nitesh & Bowyer, Kevin & Hall, Lawrence & Kegelmeyer, W.. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)*. 16. 321-357.  
10.1613/jair.953.

Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785-794.  
10.1145/2939672.2939785.

Crook, Jonathan & Edelman, David & Thomas, Lyn. (2007). Recent developments in Consumer Credit Risk assessment. *European Journal of Operational Research*. 183. 1447-1465.  
10.1016/j.ejor.2006.09.100.

Krawczyk, Bartosz. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*. 5. 10.1007/s13748-016-0094-0.

Hall, P., Gill, N., Kurka, M., & Phan, W. (2019). *Machine learning interpretability with H2O driverless AI*. Mountain View, CA: H2O.ai, Inc.

Hand, D.J. and Henley, W.E. (1997) Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160, 523-541. <https://doi.org/10.1111/j.1467-985X.1997.00078.x>

Khandani, Amir & Kim, Adlar & Lo, Andrew. (2010). Consumer Credit-Risk Models Via Machine-Learning Algorithms. *Journal of Banking & Finance*. 34. 2767–2787. 10.1016/j.jbankfin.2010.06.001.

Lessmann, Stefan & Baesens, Bart & Seow, Hsin-Vonn & Thomas, Lyn. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*. (doi:10.1016/j.ejor.2015.05.030). 10.1016/j.ejor.2015.05.030.

Lundberg, Scott & Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions. 10.48550/arXiv.1705.07874.

Malekipirbazari, Milad & Aksakalli, Vural. (2015). Risk Assessment in Social Lending via Random Forests. *Expert Systems with Applications*. 10.1016/j.eswa.2015.02.001.

Mehrabi, Ninareh & Morstatter, Fred & Saxena, Nripsuta & Lerman, Kristina & Galstyan, Aram. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*. 54. 1-35. 10.1145/3457607.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627–635. <https://doi.org/10.1057/palgrave.jors.2601545>

Doshi-Velez, Finale & Kim, Been. (2017). A Roadmap for a Rigorous Science of Interpretability. 10.48550/arXiv.1702.08608.

H. He and E. A. Garcia, "Learning from Imbalanced Data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, Sept. 2009, doi: 10.1109/TKDE.2008.239.

Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition (Springer Series in Statistics).

Jain, A., Kumar, V., & Singh, A. (2019). Preprocessing techniques for classification using machine learning. *Materials Today: Proceedings*, 11, 572–581.

Kearns, Michael & Roth, Aaron. (2021). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. *Perspectives on Science and Christian Faith*. 73. 55-56. 10.56315/PSCF3-21Kearns.

Kotsiantis, Sotiris & Zaharakis, I. & Pintelas, P.. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*. 26. 159-190. 10.1007/s10462-007-9052-3.

Kumar, Abhishek & Kumar, Abhijeet & Singh, Aditya & Nikita, Ms. (2024). Credit Scoring System Using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 10. 376-380. 10.32628/CSEIT2410334.

Lipton, Zachary. (2016). The Mythos of Model Interpretability. *Communications of the ACM*. 61. 10.1145/3233231.

Lund Chawla, Nitesh & Bowyer, Kevin & Hall, Lawrence & Kegelmeyer, W.. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)*. 16. 321-357. 10.1613/jair.953.

Martens, David & Baesens, Bart & Van Gestel, Tony & Vanthienen, Jan, 2007.

"Comprehensible credit scoring models using rule extraction from support vector machines," *European Journal of Operational Research*, Elsevier, vol. 183(3), pages 1466-1476, December.

Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C, Crandall JW, Christakis NA, Couzin ID, Jackson MO, Jennings NR, Kamar E, Kloumann IM, Larochelle H, Lazer D, McElreath R, Mislove A, Parkes DC, Pentland A', Roberts ME, Shariff A, Tenenbaum JB, Wellman M. Machine behaviour. *Nature*. 2019 Apr;568(7753):477-486. doi:

10.1038/s41586-019-1138-y. Epub 2019 Apr 24. PMID: 31019318.

Zhang, Runchi and Ligu, Xue and Qin, Wang, An Ensemble Credit Scoring Model Based on Logistic Regression with Heterogeneous Balancing and Weighting Effects. Available at SSRN: <https://ssrn.com/abstract=4167821> or <http://dx.doi.org/10.2139/ssrn.4167821>

Baylor, Denis & Koc, Levent & Koo, Chiu & Lew, Lukasz & Mewald, Clemens & Modi, Akshay & Polyzotis, Neoklis & Ramesh, Sukriti & Roy, Sudip & Whang, Steven & Wicke, Martin & Breck, Eric & Wilkiewicz, Jarek & Zhang, Xin & Zinkevich, Martin & Cheng, Heng-Tze & Fiedel, Noah & Foo, Chuan & Haque, Zakaria & Jain, Vihan. (2017). TFX: A TensorFlow-Based Production-Scale Machine Learning Platform. 1387-1395. 10.1145/3097983.3098021.

Binns, Reuben & Van Kleek, Max & Veale, Michael & Lyngs, Ulrik & Zhao, Jun & Shadbolt, Nigel. (2018). 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. 10.1145/3173574.3173951.

Jagtiani, Julapa & Lemieux, Catharine. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*. 48. 10.1111/fima.12295.

Liu, Gao, Jing & Sun, Wenjun & Sui, Xin. (2021). Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model. *Discrete Dynamics in Nature and Society*. 2021. 1-13. 10.1155/2021/5080472.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. <https://doi.org/10.1145/3287560.3287596>

Raji, Inioluwa & Smart, Andrew & White, Rebecca & Mitchell, Margaret & Gebru, Timnit & Hutchinson, Ben & Smith-Loud, Jamila & Theron, Daniel & Barnes, Parker. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. 33-44. 10.1145/3351095.3372873.

Reserve Bank of India. (2021). *Report of the Working Group on Digital Lending including Lending through Online Platforms and Mobile Apps*.

Rudin, Cynthia. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*. 1. 206-215. 10.1038/s42256-019-0048-x.

Sirignano, Justin & Sadhwani, Apaar & Giesecke, Kay. (2016). Deep Learning for Mortgage Risk. 10.48550/arXiv.1607.02470.

Toit, Hendrik & Schutte, Willem & Raubenheimer, Helgard. (2023). Shapley values as an interpretability technique in credit scoring. *Journal of Risk Model Validation*. 10.21314/JRMV.2023.010.

Wang, Weiyu & Siau, Keng. (2019). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management*. 30. 61-79. 10.4018/JDM.2019010104.

Wang, Kui & Li, Meixuan & Cheng, Jingyi & Zhou, Xiaomeng & Li, Gang. (2022). Research on personal credit risk evaluation based on XGBoost. *Procedia Computer Science*. 199. 1128-1135. 10.1016/j.procs.2022.01.143.

Zaharia, Matei A. et al. "Accelerating the Machine Learning Lifecycle with MLflow." *IEEE Data Eng. Bull.* 41 (2018): 39-45.

Zhang, Y., Tang, J., & Yang, Z. (2022). A decentralized ML-based credit scoring system with blockchain support. *Information Processing & Management*, 59(2), 102815.

Zliobaite, Indre. (2017). Measuring discrimination in algorithmic decision making. *Data Mining and Knowledge Discovery*. 31. 1-30. 10.1007/s10618-017-0506-1.

Zheng, A. and Casari, A. (2018) *Feature Engineering for Machine Learning: Principles and techniques for Data Scientists*. O'Reilly Media, Inc., Sebastopol.