



Predicting Lender Approval Decisions and Optimizing Broker Routing Using Machine Learning

Group 3 – Case Study 4 (Car Finance) Technical Report – MS984

Ebrima Khan, Muhammad Qureshi, Daayum Mohsin, Shivangi Sinha

Contents

1. Executive Summary.....	2
2. Exploratory Data Analysis	2
1. Target Distribution	2
2. Feature Distributions	2
3. Feature Relationships	3
3. Predictive Modelling	3
1. Logistic Regression Baseline	3
2. Boosted Model.....	4
4. Lender Decision Rules	4
1. The prime lender.....	4
2. The mid-prime lender	4
3. The stability-focused lender	4
5. Broker Routing Decision Strategy	5
6. GDPR and Ethical Considerations.....	5
7. Limitations and Future Improvements	5
8. Conclusion.....	6
9. Appendix	6

1. Executive Summary

Financial brokers must decide which lender is most likely to approve a borrower's loan application. Because lenders apply different risk policies, sending applications to unsuitable lenders can lead to high rejection rates, unnecessary credit checks, and reduced broker efficiency. This project explores how machine learning can support broker decision-making by predicting lender approval outcomes and improving the routing of loan applications.

Using a dataset of simulated loan applications, predictive models were developed to estimate approval probabilities for three lenders: a prime lender, a mid-prime lender, and a stability-focused lender. Exploratory analysis identified key borrower risk indicators influencing approval decisions, particularly debt-to-income ratio, credit utilization, and missed payment history. Two models were developed: a logistic regression baseline and a boosted ensemble model. The boosted model achieved the strongest performance, with ROC-AUC scores ranging from 0.921 to 0.975.

The analysis also revealed clear differences in lender risk tolerance. The prime lender applies the strictest criteria, while the mid-prime and stability lenders demonstrate greater flexibility. Based on these insights, an intelligent routing strategy is proposed in which brokers use model predictions to submit applications to the lender with the highest probability of approval. This approach can increase first-time approvals, reduce unnecessary credit checks, and improve broker efficiency.

2. Exploratory Data Analysis

Exploratory data analysis was conducted to understand borrower characteristics and identify variables that influence lender approval decisions.

1. Target Distribution

Analysis of lender approval outcomes shows clear differences in risk tolerance across lenders. The prime lender demonstrates the strictest approval policy, approving approximately 19% of applications. In contrast, the mid-prime lender approves around 44% of applications, while the stability-focused lender approves approximately 38%.

These differences suggest that lenders operate with distinct risk appetites. The prime lender appears to prioritize low-risk borrowers, while the mid-prime and stability lenders are more flexible in accepting applicants with moderate financial risk. The distribution of approval rates is illustrated in **Appendix Figure A1**.

2. Feature Distributions

The distribution of borrower financial characteristics reveals several important patterns. Applicant income exhibits a right-skewed distribution, with most borrowers earning between £30,000 and £70,000. **Employment tenure** is concentrated between five and ten years, indicating moderate employment stability for the majority of applicants. **Credit history length** also displays a right-skewed distribution, suggesting that many borrowers have relatively short credit histories.

Credit utilization ratios cluster between 0.4 and 0.7, indicating moderate reliance on credit among many borrowers. **Missed payment** counts remain relatively low, with most borrowers having between zero and two missed payments in the previous twelve months. **Loan requests** are similarly right-skewed, with most borrowers requesting loans between £5,000 and £20,000.

These patterns indicate that while demographic and income variables provide useful context, lenders likely rely more heavily on behavioral risk indicators such as leverage and repayment history when making approval decisions. These feature distributions are illustrated in **Appendix Figures A1–A2**.

3. Feature Relationships

Further analysis of relationships between borrower variables and lender approval outcomes highlights several important risk indicators. Higher debt-to-income ratios and higher credit utilization levels are strongly associated with lower approval probabilities. Additionally, borrowers with multiple missed payments exhibit significantly lower approval rates, particularly for the prime lender.

These findings suggest that three variables play a particularly important role in lender decisions:

- Debt-to-income ratio
- Credit utilization ratio
- Missed payment history

The relationships between these variables and approval outcomes are visualized in **Appendix Figures A2**.

3. Predictive Modelling

To predict lender approval outcomes, two classification models were developed: a logistic regression baseline model and a boosted model using the HistGradientBoosting algorithm.

1. Logistic Regression Baseline

Logistic regression was selected as the baseline model due to its interpretability and ability to estimate approval probabilities directly. The dataset was divided into training and testing sets using an 80/20 split with stratified sampling to maintain class balance. Model performance

was evaluated using the receiver operating characteristic area under the curve (ROC-AUC), which measures the model's ability to distinguish between approved and rejected applications.

The logistic regression model demonstrated strong predictive performance across all lenders. The ROC-AUC scores were 0.954 for the prime lender, 0.895 for the mid-prime lender, and 0.889 for the stability-focused lender. These results indicate that borrower financial characteristics provide strong predictive signals for lender approval decisions.

The ROC curves for these models are presented in **Appendix Figure A4 & A5**.

2. Boosted Model

To improve predictive performance, a boosted model using HistGradientBoostingClassifier was implemented. Boosting algorithms construct an ensemble of **decision trees** sequentially, enabling the model to capture nonlinear relationships and interactions between borrower characteristics.

The boosted model achieved higher predictive performance for all lenders. The ROC-AUC scores increased to 0.975 for the prime lender, 0.926 for the mid-prime lender, and 0.921 for the stability lender. This improvement suggests that lender approval decisions involve nonlinear relationships between borrower risk indicators that cannot be fully captured by a linear model.

A comparison of model performance is illustrated in **Appendix Figure A3, A3a & A5**

4. Lender Decision Rules

By analyzing approval probabilities across borrower risk indicators, several decision rules that characterize lender behavior were identified.

1. The prime lender

Exhibits the strictest approval policy. Borrowers with debt-to-income ratios exceedingly approximately 0.9 are consistently rejected. Additionally, applicants with two or more missed payments within the previous year receive no approvals, indicating a strict repayment reliability threshold.

2. The mid-prime lender

Demonstrates greater flexibility. Borrowers with moderate leverage and up to two missed payments may still receive approval. However, applicants with extremely high debt levels or three or more missed payments are consistently rejected.

3. The stability-focused lender

Appears to have the most flexible lending policy. While this lender accepts borrowers with higher leverage levels, it still enforces minimum repayment reliability requirements, rejecting borrowers with three or more missed payments.

These approval patterns are visualized in **Appendix Figures A6**.

5. Broker Routing Decision Strategy

The predictive models developed in this study allow brokers to estimate the probability that each lender will approve a loan application. These predictions can be used to guide loan routing decisions.

Instead of submitting applications randomly, brokers can calculate approval probabilities for each lender and submit the application to the lender with the highest predicted probability of approval. This approach enables more effective matching between borrowers and lenders.

An effective routing strategy involves submitting applications sequentially. Borrowers with strong financial profiles and low risk indicators are best suited to the prime lender. Applicants with moderate leverage may be routed to the mid-prime lender, while borrowers with higher leverage but stable repayment behavior may be directed to the stability-focused lender.

This strategy can increase first-time approval rates, reduce unnecessary credit checks, and improve broker operational efficiency. A conceptual routing framework is illustrated in **Appendix Figure A7**.

6. GDPR and Ethical Considerations

Although the dataset used in this study is synthetic, real-world implementations of such models must comply with data protection regulations such as the General Data Protection Regulation (GDPR).

Loan approval models rely on personal financial data, which is considered sensitive information under GDPR. Organizations using such models must ensure that borrower data is processed lawfully, transparently, and only for legitimate purposes. Data minimization principles require that only necessary borrower information be collected and processed.

In addition, algorithmic decision systems must remain explainable and auditable. Financial institutions are required to provide meaningful explanations for automated decisions affecting individuals. Therefore, interpretable models and transparent decision rules remain important considerations when deploying machine learning systems in lending environments.

7. Limitations and Future Improvements

Several limitations should be acknowledged. The dataset used in this analysis is synthetic and may not fully represent real-world borrower behavior or lender policies. Additionally, important variables such as borrower savings, loan purpose, and asset ownership were not included.

Future work could improve model performance by incorporating additional borrower attributes and performing further feature engineering. Model performance could also be improved through hyperparameter tuning and cross-validation. Finally, real-world deployment would require

continuous monitoring to ensure that model predictions remain accurate as lending policies and borrower behavior evolve.

8. Conclusion

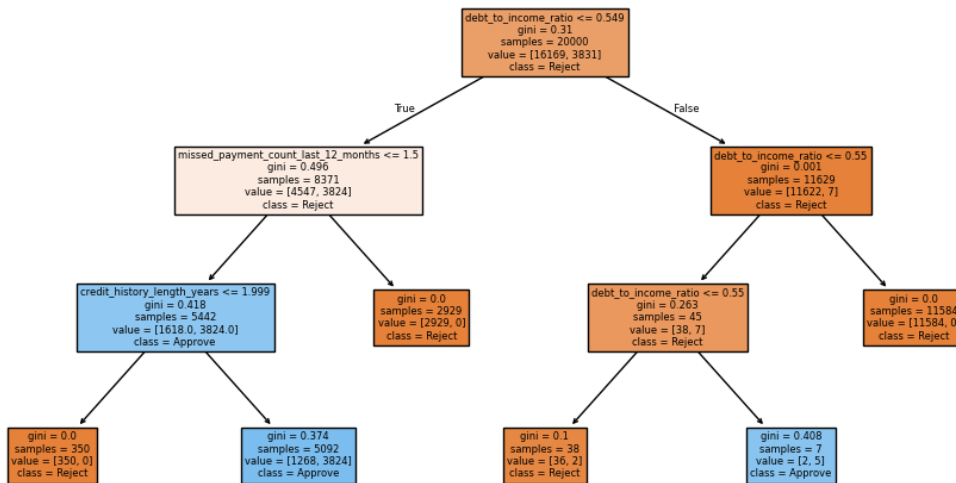
This study demonstrates how machine learning can support broker decision-making in loan application routing. By analyzing borrower financial characteristics and modelling lender approval behavior, the study identifies key risk indicators that influence lender decisions.

Both logistic regression and boosted models achieved strong predictive performance, with the boosted model providing the highest accuracy. The analysis also reveals clear differences in risk tolerance between lenders, enabling the identification of decision rules that characterize their approval behavior.

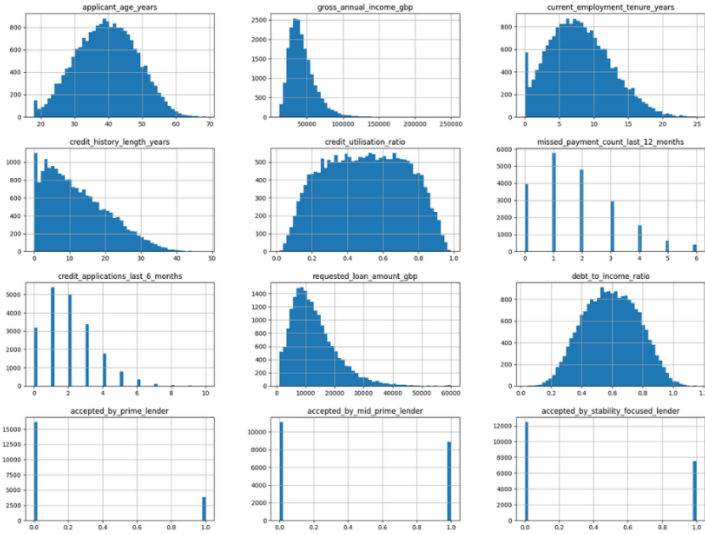
These insights allow brokers to implement intelligent routing strategies that match borrowers with lenders whose risk preferences align with their financial profiles. Such an approach can increase first-time approvals, reduce failed applications, and improve the efficiency of loan brokerage operations

9. Appendix

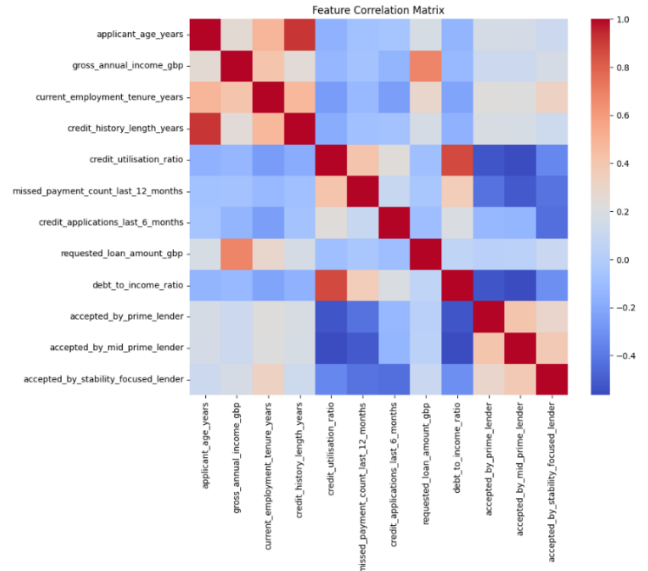
Appendix A3a



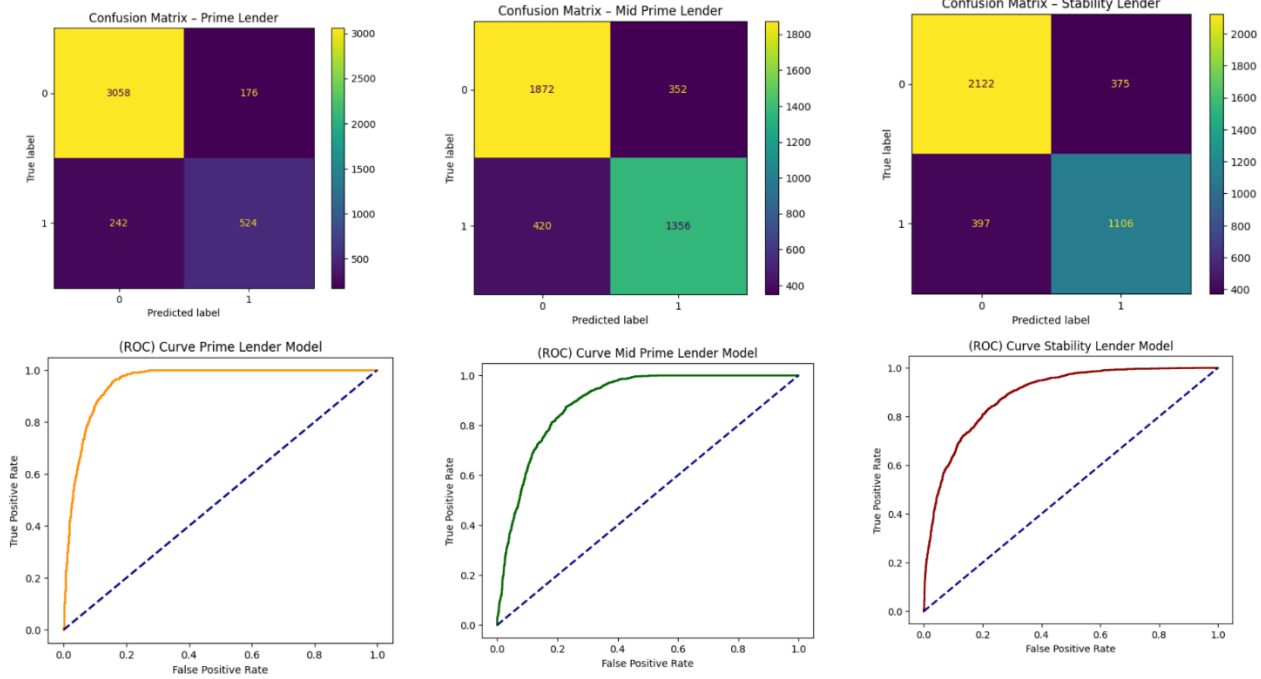
Appendix A1



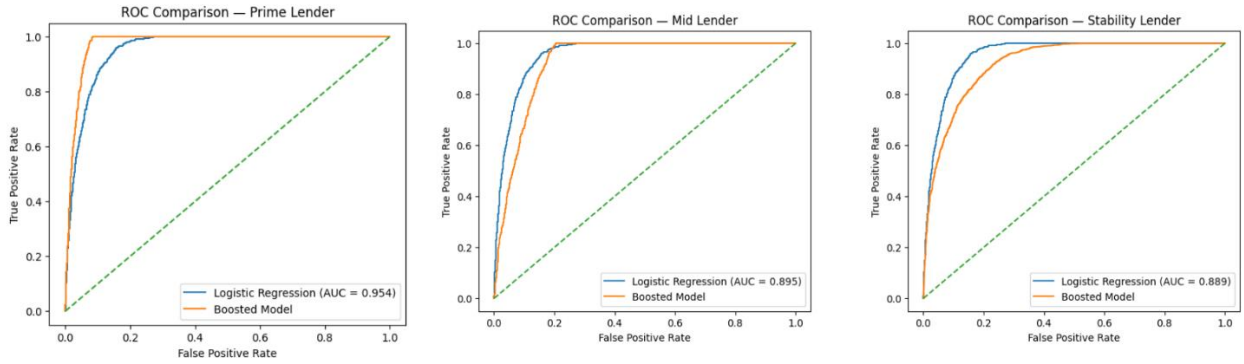
Appendix A2



Appendix A4



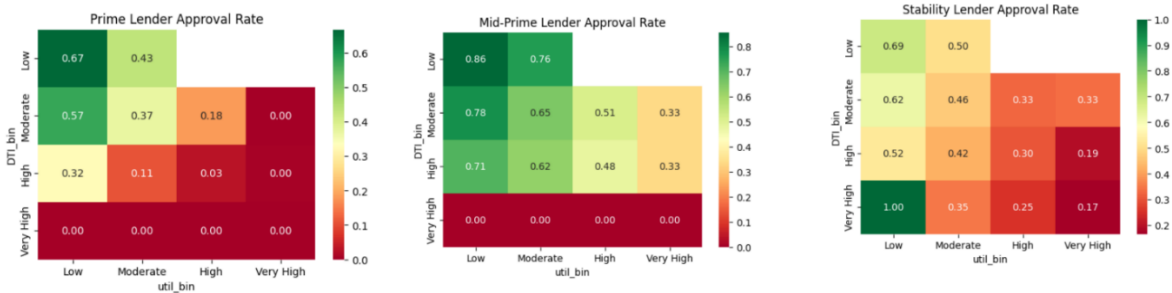
Appendix A3



Appendix A5

	Lender	Logistic ROC-AUC	Boosted ROC-AUC	Improvement
0	Prime	0.954	0.975	0.021
1	Mid	0.895	0.926	0.031
2	Stability	0.889	0.921	0.032

Appendix A6



Appendix A7

Borrower Profile	Best Lender
Low DTI + low utilisation + clean payment history	Prime lender
Moderate DTI or moderate utilisation	Mid-prime lender
High DTI but stable repayment behaviour	Stability lender

Appendix 8

https://github.com/KhanJallow/Broker-Routing-Using-Machine-Learning/blob/main/Lender_Decision.ipynb