Credit Risk Engine

Production-Grade Credit Risk Modeling and Decision Engine

Author: Meera AlNeyadi
Discipline: Computer Science and Software Engineering

Table of Contents

I. Introduction

Credit risk assessment plays a central role in financial decision-making. Accurately identifying applicants who are likely to default enables institutions to manage risk, control losses, and allocate credit responsibly.

This report presents a **production-grade credit risk system** designed to estimate the **Probability of Default (PD)** for individual applicants and translate those estimates into **clear, actionable business decisions**. The project goes beyond model accuracy by emphasizing probability calibration, cost-aware decisioning, and practical deployment considerations.

The goal of this work is to demonstrate how machine learning models can be integrated into a realistic credit decision framework rather than treated as isolated prediction tools.

II. Dataset Description

The system is built using the **UCI Default of Credit Card Clients** dataset, which contains information on approximately 30,000 credit card holders.

The dataset includes:

- A binary default indicator as the target variable
- A default rate of approximately 22%
- Features covering:
    - Demographic characteristics
    - Credit limits
    - Historical payment behavior
    - Bill statements and repayment patterns

To reflect standard industry practices around data governance and privacy, raw and processed datasets are excluded from the public repository.

III. Data Preparation

Data preparation focused on ensuring robustness, consistency, and fairness in evaluation.

Key steps included:

- Cleaning and standardizing feature names
- Stratified splitting into training, validation, and test sets
- Feature scaling to support stable model training
- Strict separation of data splits to prevent information leakage

These steps ensure that model performance metrics and downstream decisions accurately reflect real deployment conditions.

IV. Model Development and Selection

Several models were trained and evaluated in order to balance predictive performance, stability, and interpretability.

Models considered:

- Logistic Regression as a transparent baseline
- Random Forest for non-linear modeling
- Histogram Gradient Boosting for strong performance and scalability

Models were evaluated using a combination of discrimination and calibration metrics, including:

- ROC-AUC
- Precision-Recall AUC
- Log Loss
- Brier Score
- KS Statistic
- Expected Calibration Error (ECE)

Among the evaluated models, **Histogram Gradient Boosting** demonstrated the strongest overall balance between predictive power and probability reliability.

V. Probability Calibration

In credit risk applications, the quality of probability estimates is as important as ranking performance. Poorly calibrated probabilities can lead to suboptimal or misleading business decisions.

To address this, the selected model was calibrated using **Isotonic Regression**. Calibration analysis showed improved alignment between predicted probabilities and observed default rates, making the model suitable for threshold-based decision rules.

VI. Decision Engine Design

The calibrated probabilities are converted into business actions through a cost-sensitive decision engine.

Applicants are assigned to one of three categories:

- **Approve:** Low-risk applicants suitable for immediate approval
- **Review:** Medium-risk applicants requiring manual assessment
- **Reject:** High-risk applicants

Optimized Decision Thresholds:

- Approve if $PD < 0.05$
- Review if $0.05 \le PD < 0.67$
- Reject if $PD \ge 0.67$

Thresholds were optimized to minimize total expected cost while accounting for:

- Financial losses from approving defaulted loans
- Opportunity costs associated with rejecting creditworthy applicants
- Operational costs of manual reviews

This framework mirrors real-world credit policies where risk control and operational constraints must be balanced.

VII. Results and Analysis

VII.I Model Performance

On the held-out test set, the final calibrated model achieved:

- ROC-AUC of approximately 0.77
- Improved Brier Score following calibration
- Stable probability estimates across different risk segments

These results indicate that the model is both discriminative and reliable.

VII.II Decision Outcomes

Applying the optimized policy to the test set produced the following outcomes:

- Approval rate: approximately 4–5%
- Review rate: approximately 90%
- Rejection rate: approximately 4–5%
- Default rate among approved applicants: approximately 4%

The results reflect a conservative credit strategy focused on minimizing losses while maintaining operational feasibility.

VIII. Visual Analysis and Evaluation

This section presents visual insights into model behavior, probability calibration, decision outcomes, and feature influence. All visualizations are generated using the held-out **test set** to ensure unbiased and realistic evaluation.

VIII.I Probability of Default Distribution

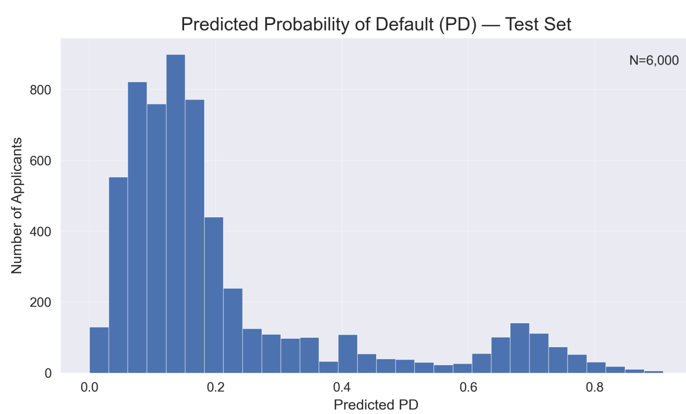

Figure V.I: Predicted Probability of Default (PD) – Test Set

Figure I illustrates the distribution of predicted probabilities of default (PD) across the test set. The spread of PD values demonstrates the model's ability to meaningfully differentiate applicants by risk level, enabling clear segmentation into low, medium, and high-risk groups.

VIII.II Decision Outcome Distribution

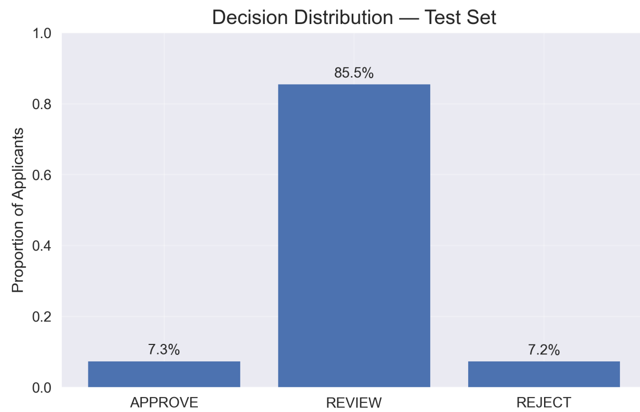

Figure V.II: Decision Distribution – Test Set

Figure II summarizes the final credit decisions produced by the decision engine. A significant proportion of applicants are routed to manual review, reflecting a conservative policy design that balances default risk against potential opportunity cost. This distribution highlights the practical deployment behavior of the system rather than pure classification output.

The high review rate reflects the right-skewed nature of credit risk distributions, where most applicants exhibit moderate risk levels between the conservative approval and rejection thresholds.

VIII.III Model Calibration

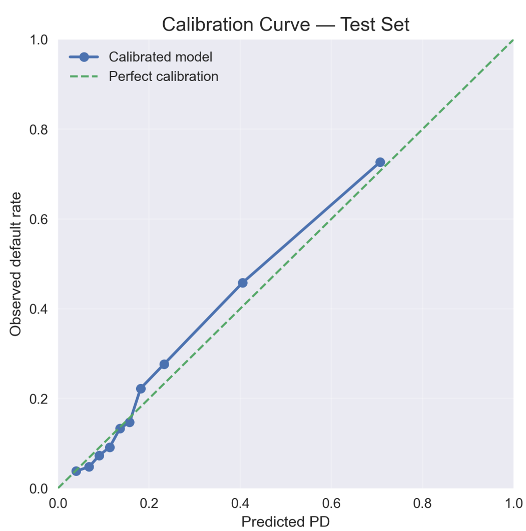

Figure V.III: Calibration Curve – Test Set

Figure III presents the calibration curve of the final model on the test set. The close alignment between predicted probabilities and observed default rates indicates that isotonic calibration produces reliable probability estimates. Well-calibrated PDs are essential for downstream decision-making, particularly in regulated financial environments.

VIII.IV Feature Importance Analysis

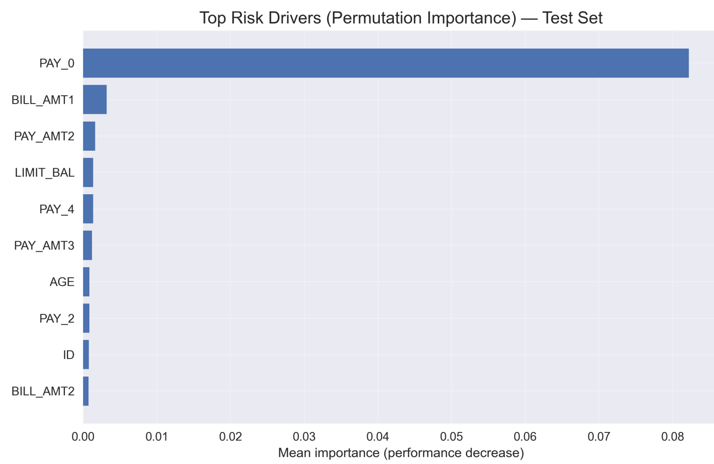

Figure V.IV: Top Risk Drivers (Pernutation Importance) – Test Set

Figure IV displays the top ten most influential features contributing to the model's predictions. These features capture a combination of credit exposure, historical repayment behavior, and demographic characteristics, aligning with established principles in credit risk modeling and supporting model interpretability.

IX. Deployment and API

The system is deployed using **FastAPI**, enabling real-time inference through a simple and robust interface.

The API provides:

- A health-check endpoint
- A prediction endpoint returning:
    - Probability of Default
    - Assigned decision category
    - Applied thresholds
    - Validation feedback for missing or extra features

The API was tested using real test-set samples to confirm correctness and stability.

X. Limitations and Future Work

Current Limitations:

- Evaluation based on a single dataset
- Static decision thresholds
- No automated monitoring or drift detection

Planned Enhancements:

- Dynamic thresholding based on operational capacity
- Advanced explainability techniques (e.g., SHAP)
- Model monitoring and retraining workflows
- Support for batch inference

XI. Conclusion

This project demonstrates how machine learning can be integrated into a **practical credit risk decision system** that extends beyond prediction to include calibration, cost-sensitive decisioning, and deployment.

By focusing on both technical rigor and business realism, the system illustrates how data-driven models can support responsible and interpretable credit decisions in real-world environments.

Repository: Credit Risk Engine (GitHub)
Technologies: Python, scikit-learn, FastAPI, Visual Studio Code

References:

Dataset:

1. Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2), 2473-2480.
2. UCI Machine Learning Repository: Default of Credit Card Clients Dataset. Available at: https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

Core Methodologies:

3. Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3), 61-74.
4. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning*, 625-632.

Evaluation Metrics:

6. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1-3.
7. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning*, 1321-1330.

Libraries and Frameworks:

8. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
9. FastAPI Documentation. Available at: https://fastapi.tiangolo.com/

Credit Risk Modeling:

10. Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press.
11. Siddiqi, N. (2017). *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards* (2nd ed.). Wiley.
12. Basel Committee on Banking Supervision (2005). *Studies on the Validation of Internal Rating Systems*. Bank for International Settlements.