Explainable and Fair Credit Risk Modeling

Executive Presentation Summary

Model: Calibrated Logistic Regression (Isotonic Calibration)

November 2025

1. Project Objective

The goal of this project was to design a credit risk prediction model that combines high predictive power, interpretability, and fairness across demographic groups. The final model integrates modern explainability techniques and fairness evaluation metrics to ensure ethical and transparent decision-making in credit approval processes.

2. Data Exploration and Quality

A detailed exploratory data analysis confirmed that the dataset had:

- Minimal missing data, treated through median imputation.
- Balanced representation across gender and age.
- Predictive financial features such as credit utilization, payment history, and income stability.

No significant data imbalance or bias was found, providing a reliable base for fair modeling. Outliers were evaluated for business relevance instead of automatic removal.

3. Modeling Process

Several algorithms were compared, including Decision Tree, Random Forest, and Logistic Regression. While complex models (e.g., Random Forest) showed slightly higher accuracy, the **Logistic Regression** model was preferred due to its transparency and interpretability—both critical in credit risk regulation.

To improve the reliability of predicted probabilities, the model's sigmoid output was calibrated using **Isotonic Calibration**. This step corrected deviations between predicted and actual probabilities, improving both calibration and fairness consistency.

4. Model Evaluation

The Calibrated Logistic Regression achieved:

- High AUC and KS values, demonstrating strong classification performance.
- A lower Brier Score, confirming well-calibrated probabilities.
- A nearly linear calibration curve after isotonic adjustment.

The operational decision threshold was set at **0.40**, optimizing the trade-off between detecting risky clients (sensitivity) and minimizing unjustified rejections (specificity).

5. Explainability

Explainability was assessed using SHAP (SHapley Additive Explanations) values. Results showed that:

- Financial and behavioral features (credit utilization, payment history) are the dominant predictors.
- Demographic features such as gender and age have minimal influence on predictions.

This ensures the model's decisions are transparent, interpretable, and based on legitimate financial indicators, aligning with regulatory expectations for explainable AI in credit scoring.

6. Fairness Evaluation

The fairness analysis assessed two key protected attributes: **gender** and **age**. The Equal Opportunity Difference (EOD) was used to measure disparities in the True Positive Rate (TPR) between groups.

Attribute	Metric	Observed EOD	Interpretation
Gender	Equal Opportunity Difference	≈ 0.01	Fair (minimal difference)
Age	Equal Opportunity Difference	≈ -0.02	Fair (consistent detection)

Table 1: Fairness Results by Protected Attribute

Both results indicate negligible differences between groups, demonstrating equitable treatment and confirming that the model does not systematically disadvantage any demographic segment.

The isotonic calibration contributed to reducing any latent group bias in predicted probabilities.

7. Operational Recommendation

The selected model for production is the **Calibrated Logistic Regression (Isotonic Calibration)** due to its:

- Reliable probability calibration,
- Transparent coefficients and interpretability,
- Proven fairness performance.

Implementation Highlights:

- Apply 0.40 threshold in operational decisioning.
- Monitor performance (AUC, KS, Brier) quarterly.
- Audit fairness metrics (EOD) semi-annually.
- Recalibrate annually using updated data.

8. Governance and Fairness Monitoring

A continuous governance framework ensures model stability and ethical compliance:

- **Performance Monitoring:** Track drift and calibration quarterly.
- **Fairness Audits:** Semi-annual EOD evaluation with alert threshold $|EOD| \leq 0.05$.
- **Recalibration:** Annual model retraining using recent data.

All updates are documented in a **Model Governance Log** for traceability and regulatory readiness.

9. Risk Management Considerations

- **Bias Reinforcement:** Periodic retraining prevents bias reintroduction.
- **Regulatory Compliance:** Maintains fairness documentation and calibration evidence.

- **Transparency:** Provides clear explanations for each decision and rationale for model updates.

10. Final Conclusion

The **Calibrated Logistic Regression with Isotonic Calibration** provides the optimal balance between accuracy, explainability, and fairness for credit risk modeling.

This approach ensures:

- Probabilities that reflect real-world risk,
- Equitable treatment across demographic groups,
- Full interpretability for auditors and decision-makers.

Final Outcome: An explainable, fair, and operationally robust credit scoring system that aligns with ethical AI and regulatory standards.