DATA BOOTCAMP FINAL PROJECT

Credit Default Prediction and Policy-Oriented Threshold Selection

Objective: Predict borrower default risk and translate model outputs into credit decision policies

Focus: Trade-offs between accuracy, interpretability, and financial costs

Presented by Arnur Abdrakhman

Dataset & Predictive Task

Understanding the data context

Dataset	**UCI "Default of Credit Card Clients" (Taiwan)**
Observations	**30,000**
Features	**23**
Default Rate	**22.12%**
Task:	**Binary classification (default vs non-default)**

Modeling Approach

We used 3 models in assessing Predictive Performance

Logistic Regression

Interpretable Baseline used in real-world credit risk applications

Random Forest

A non-linear benchmark

HistGradientBoosting

Best-predictive performance

Evaluation Metrics:

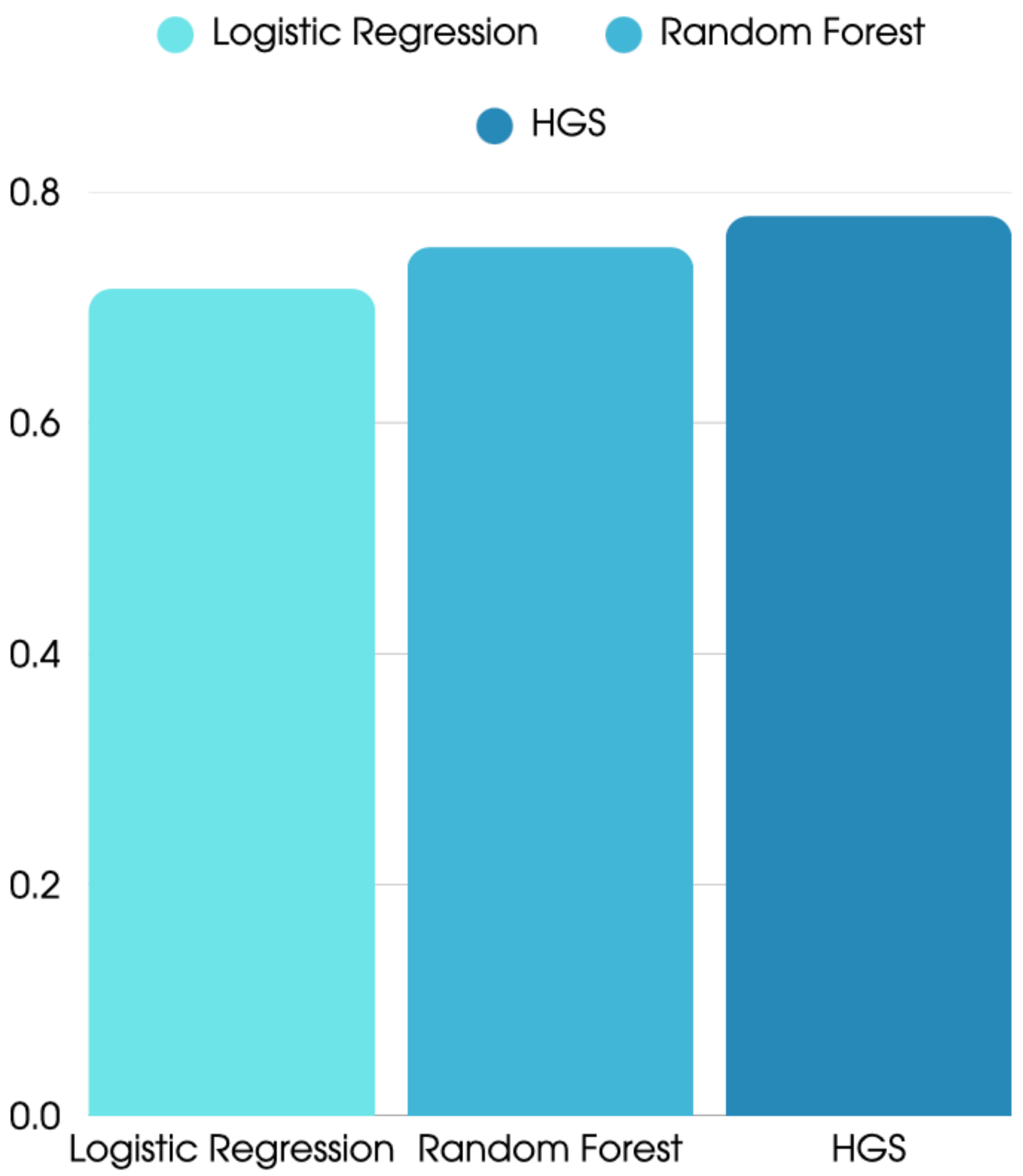
- ROC–AUC (ranking quality)
- Precision, Recall, F1 (decision quality)

Ranking performance (ROC-AUC)

Test-set ROC–AUC:

- **Logistic Regression: 0.716**
- **Random Forest: 0.752**
- **HistGradientBoosting: 0.779**

Tree-based models provide stronger risk separation, but accuracy alone does not determine deployment suitability.



High-Recall Thresholding

Risk-Averse Policy Approach

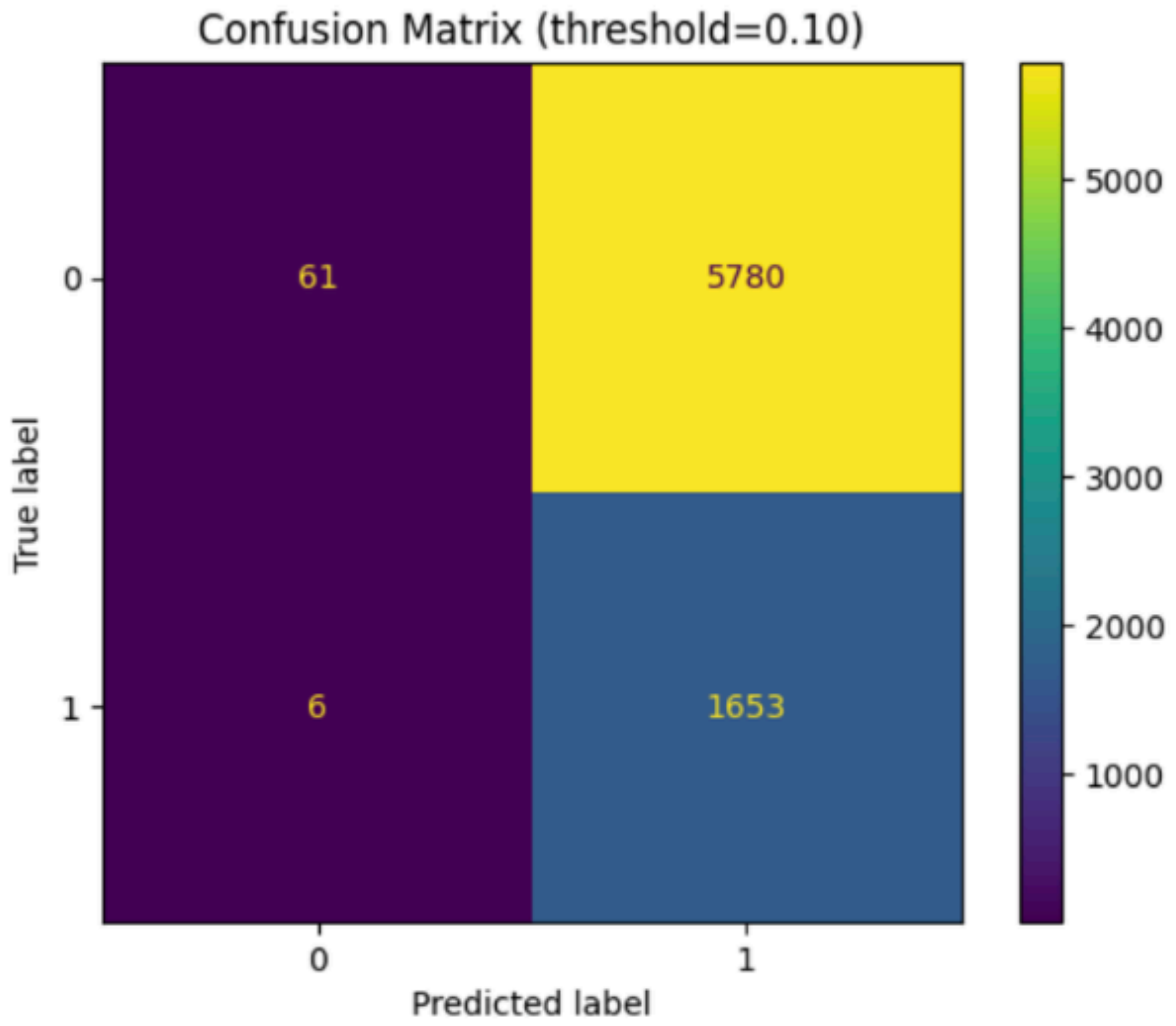
Extreme risk aversion leads to aggressive rejection and substantial lost business.

Motivation

- Missing a defaulter (FN) is often more costly than rejecting a good borrower (FP)

Logistic Regression (Recall-focused):

- Threshold ≈ 0.10
- Default recall ≈ 99.6%
- Default precision ≈ 22.2%
- False positives > 5,700



Balanced Threshold Selection (Max F1)

Balancing precision and recall produces more realistic decision policies.

Logistic Regression

- Threshold ≈ 0.60
- F1 ≈ 0.506

Random Forest

- Threshold ≈ 0.31
- F1 ≈ 0.525

HistGradientBoosting:

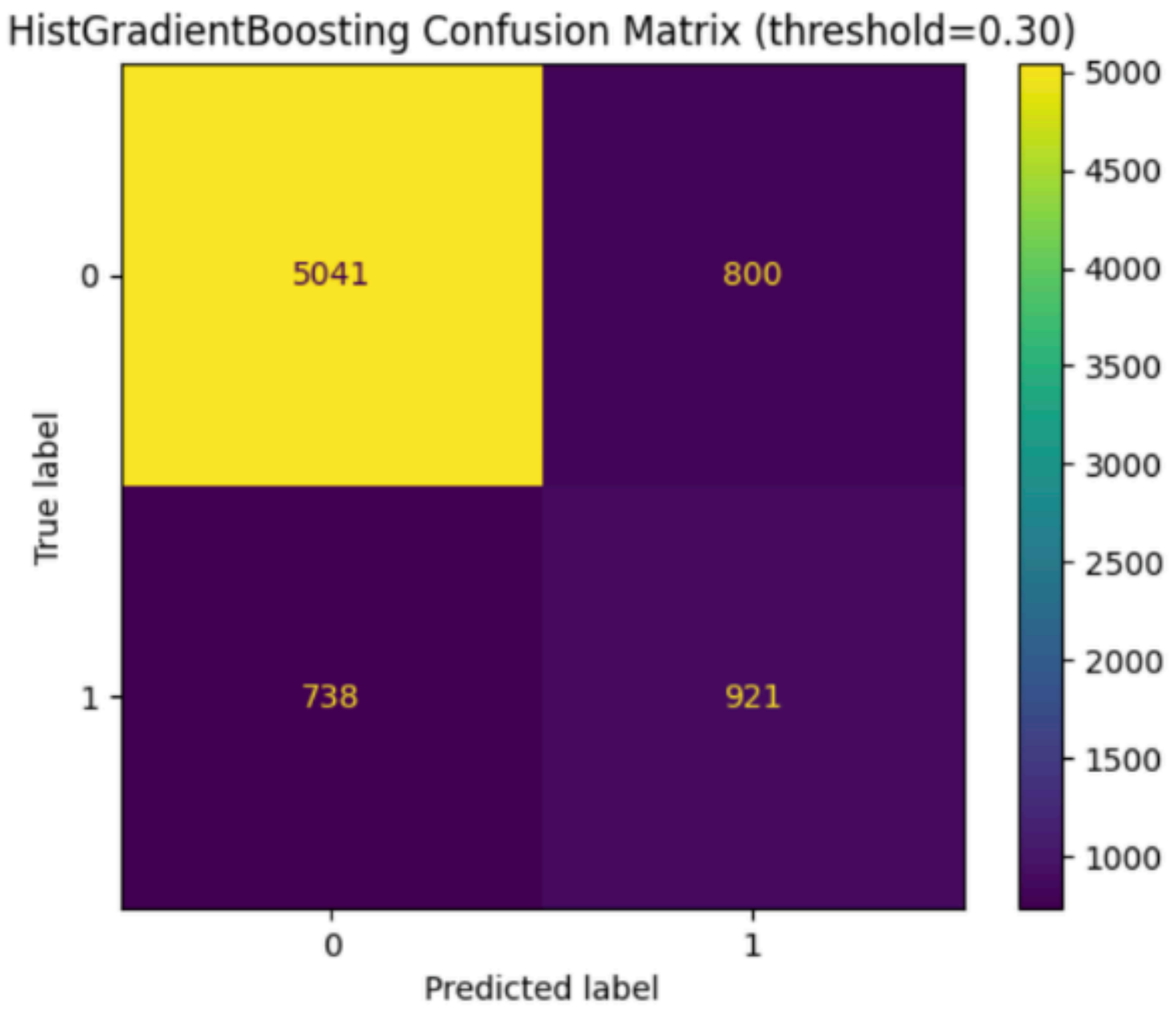
- Threshold ≈ 0.30
- F1 ≈ 0.545

Operational Performance Comparison

HistGradientBoosting @ max F1:

- Precision (default): 0.535
- Recall (default): 0.555
- Lowest combined FN + FP among models

Nonlinear models shift the precision–recall frontier outward, enabling better trade-offs.



Cost Based Threshold Optimization

Expected cost formulation:

$$Cost = C\_FN \times FN + C\_FP \times FP$$

Evaluated Cost Ratios:

- 5:1, 10:1, 20:1 (FN : FP)

Key Finding:

- Higher FN costs → lower thresholds → higher recall, more false positives

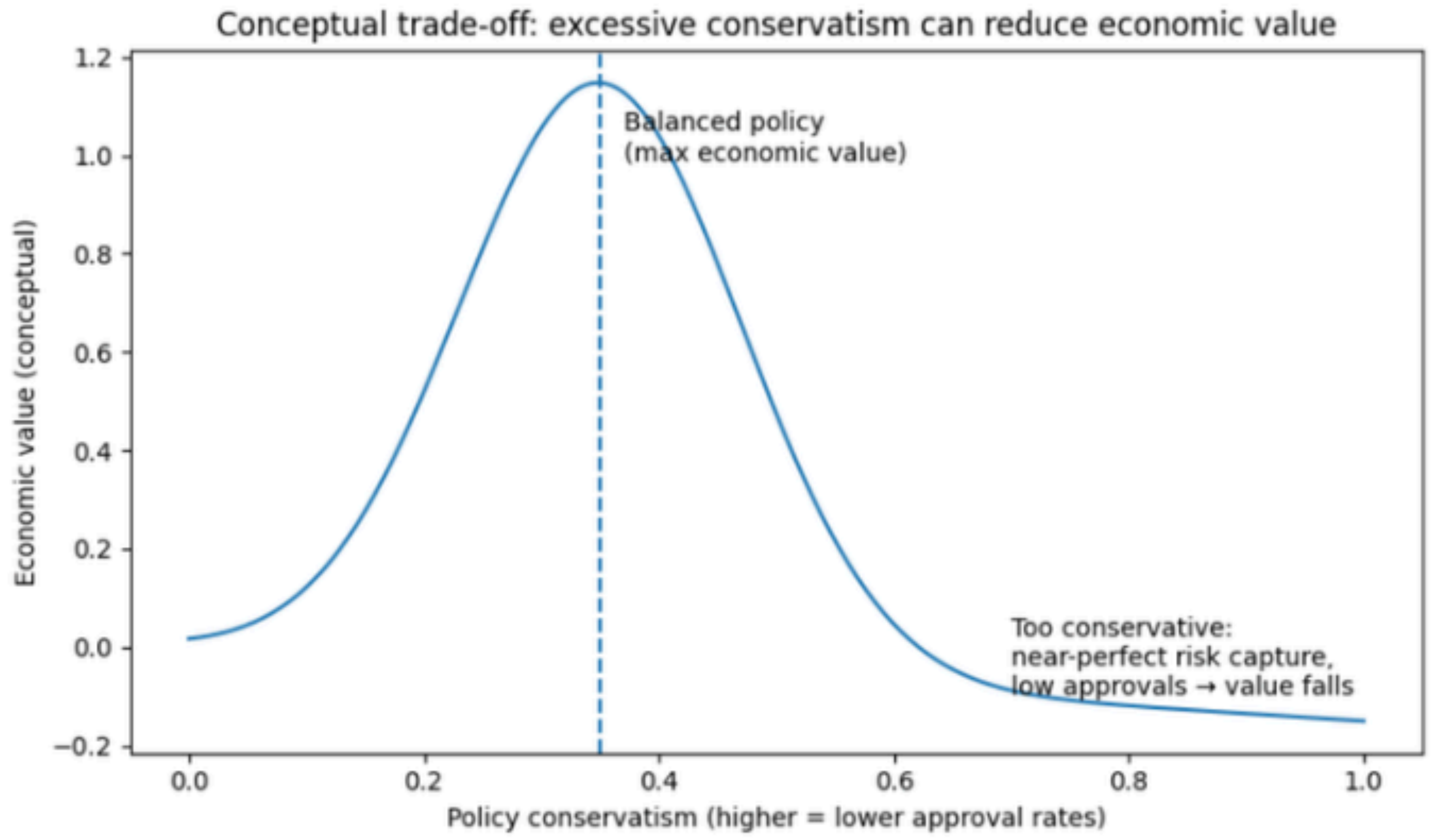
Policy Sensitivity to Risk Appetite

Risk-Averse Policy Approach

At 20:1 cost ratio (HistGradientBoosting):

- Threshold ≈ 0.06
- FN = 28
- FP = 5,150
- Recall ≈ 98%

Excessive conservatism eventually destroys economic value despite near-perfect risk capture.



Understanding Default Risk Drivers

Interpretation Focus is on Logistic Regression Only (due to transparent coefficients)

Key drivers are:

- Recent payment behavior variables (e.g., X6, X12, X19, X18)

Logistic regression provides stable, sign-interpretable relationships consistent with credit scorecard practice.

```
Top 15 coefficients (by absolute value):
   feature      coef
5        X6  0.578770
11      X12 -0.228464
18      X19 -0.181837
17      X18 -0.153287
0        X1 -0.124974
6        X7  0.107435
2        X3 -0.091156
7        X8  0.090387
4        X5  0.075561
3        X4 -0.068507
1        X2 -0.054164
13      X14  0.053457
20      X21 -0.047623
8        X9  0.046742
15      X16  0.030654
```

Conclusion & Next Steps

Recommendations:

- Use logistic regression for interpretability and governance
- Use HistGradientBoosting as a challenger or internal risk engine
- Select thresholds based on institutional risk appetite, not fixed metrics
- Treat recall, F1, and cost-based thresholds as policy levers

Next Steps:

- Probability calibration
- Time-based validation
- Capital- or profit-based objectives
- Fairness and stability analysis

Key Takeaway:

Credit modeling is a joint statistical and financial decision problem