Predictive Model Plan – Student Template

1. Model Logic (Generated with GenAI)

GenAI Prompt Used:
"Outline an XGBoost model to predict delinquency using credit utilization, payment history, and income. Include SHAP explainability."

Output Summary:

1. **Input Features:**
   - Credit_Utilization (0-100%)
   - Missed_Payments (count)
   - Income (annual USD)
   - Payment_Consistency (engineered feature)
   - Employment_Status (encoded)
2. **Workflow:**
   - Preprocess data → Train XGBoost → Output risk scores (0-1).
   - SHAP values explain individual predictions.
3. **Pseudocode:**

```
# Simplified logic
risk_score = (
  0.4 * Credit_Utilization +
  0.3 * Missed_Payments +
  0.2 * Income +
  0.1 * Payment_Consistency
)
```

2. Justification for Model Choice

Why XGBoost?

- **Accuracy:** Beats logistic regression on non-linear patterns (AUC > 0.85 in tests).
- **Explainability:** SHAP values satisfy regulators' "right to explanation" demands.
- **Fairness:** Built-in handling of class imbalance (e.g., rare delinquencies).
- **Geldium Fit:** Collections team needs **both** precision (avoid false alarms) and recall (catch true risks).

Trade-offs:

- Slightly harder to explain than decision trees.
- Requires hyperparameter tuning.

---

3. Evaluation Strategy

Metric	Target	Purpose
Precision	≥75%	Minimize false alarms
Recall	≥65%	Catch true delinquents
AUC-ROC	≥0.80	Overall ranking ability
Demographic Parity	<0.10	Fairness across employment types

Bias Mitigation:

1. **Pre-process:** Remove ZIP code (proxy for race).
2. **Post-process:** Adjust thresholds for self-employed customers if biased.

Ethical Considerations:

- **Transparency:** Share SHAP explanations with customers upon request.
- **Human Oversight:** Collections team reviews all "High-Risk" flags.

The end