Role in this phase:

.Use GenAI tools to develop a predictive model for identifying high-risk customers.

.Choose and justify the best approach—whether decision trees, logistic regression, or another technique.

.Define a plan to evaluate model performance while ensuring fairness and explainability.

1. Key metrics for model evaluation

Each metric provides a different perspective on model effectiveness. It's important to use multiple metrics together rather than relying on a single score:

- **Accuracy** – Measures the overall correctness of the model by dividing correct predictions by the total number of cases.
- **Precision (positive predictive value)** – Evaluates how many of the customers predicted to be delinquent actually are.
- **Recall (sensitivity)** – Measures how many actual delinquent customers were correctly identified by the model. High recall is important when missing a delinquent customer could result in financial loss.
- **F1 score** – A weighted balance between precision and recall. It is useful when both false positives and false negatives are costly.
- **AUC-ROC curve (area under the receiver operating characteristic curve)** – Assesses how well the model distinguishes between delinquent and non-delinquent customers. A score close to 1 means the model is highly effective at ranking risk levels, while a score near 0.5 suggests the model is no better than random guessing.
- **Confusion matrix** – A visual breakdown of actual vs. predicted classifications. It helps diagnose specific types of errors and determine whether the model is favoring one outcome over another.

2. What to do if model performance is poor

If your model is not performing well, there are several ways to improve it:

- **Feature engineering** – Adjust the dataset by adding or removing variables that may be impacting model predictions. For example, including customer tenure or past delinquency trends may enhance predictive power.
- **Rebalancing the dataset** – If the dataset is highly skewed (e.g., 95% non-delinquent, 5% delinquent), oversampling delinquent cases or undersampling non-delinquent cases can improve results.
- **Trying different models** – Some algorithms work better with certain data structures. If logistic regression is underperforming, a decision tree may provide better results.
- **Hyperparameter tuning** – Fine-tuning model parameters, such as adjusting the threshold for delinquency classification, can improve precision and recall scores.

Bias

Bias occurs when a model **systematically favors or disadvantages certain groups**, often due to historical inequalities or imbalanced data.

Common causes of bias:

- **Historical bias** – If past lending decisions were unfair, the model may replicate those patterns.
- **Selection bias** – If the dataset does not represent all customer demographics equally, predictions may be inaccurate for some groups.
- **Proxy bias** – Certain variables (e.g., ZIP code) may unintentionally act as proxies for protected characteristics like race or gender.

Explainability

Explainability ensures that decision-makers can understand and justify a model's predictions.

- **Decision trees and logistic regression** are more interpretable and show clear decision paths.
- **Neural networks** are highly complex and function as "black boxes," making explainability difficult.
- Analysts use tools like **SHAP (Shapley Additive Explanations)** to break down how different factors contribute to predictions.

Fairness

A fair model should:

- **Avoid systematic disadvantages** for certain demographic groups.
- **Be tested for disparate impact** to ensure fairness.
- **Use diverse and representative training data** to prevent reinforcing biases.

Utilizing GenAI tools to generate model code and refine predictions

1. How GenAI assists in model development

- **Generating a model framework** – A user can request a **logistic regression model for predicting delinquency**, and GenAI will provide an initial code structure. *However, it is essential to review, test, and refine the code to ensure correctness and efficiency*.
- **Feature selection assistance** – GenAI can recommend **which variables** to include based on the dataset. *However, analysts must verify that these selections do not introduce bias or proxy discrimination*.
- **Hyperparameter tuning** – Analysts can optimize model performance by asking for parameter adjustments. *While GenAI can suggest modifications, empirical testing and expert judgment are necessary to validate improvements*.

2. Refining and improving model predictions

After generating a model, it's crucial to **refine predictions** to ensure accuracy and fairness. GenAI tools can:

- Suggest modifications to improve precision and recall.
- Evaluate model outputs and identify overfitting or biases.
- Generate alternative models to compare performance.

Step1:

Predictive Modeling Pipeline for Credit Delinquency Forecasting

1. Model Selection:
2. • Logistic Regression as the primary model - interpretable, efficient, and well-suited for binary classification (delinquent/non-delinquent)
3. • Random Forest as an alternative - captures non-linear relationships and feature interactions while providing feature importance
4. • Gradient Boosting (XGBoost) for optimal performance - handles imbalanced data well and often achieves high accuracy
5. Key Input Features (Top 5):
6. Missed_Payments - Strongest predictor; directly reflects payment behavior
7. Credit_Utilization - High utilization indicates financial stress
8. Debt_to_Income_Ratio - Measures financial leverage and repayment capacity
9. Credit_Score - Reflects creditworthiness and financial history
10. Account_Tenure - Longer relationships may indicate lower delinquency risk
11. Modeling Workflow:

• Data preprocessing: Handle missing values (already done), encode categorical variables, normalize numerical features

• Feature engineering: Create derived features like payment-to-income ratio, utilization trends

• Train/validation/test split: 70%/15%/15% stratified split to maintain class distribution

• Model training: Fit models on training set with cross-validation for hyperparameter tuning

• Model evaluation: Assess using precision, recall, F1-score, and AUC-ROC (particularly focusing on recall for delinquent cases)

• Model interpretation: Analyze feature importance and partial dependence plots

• Deployment preparation: Save model, create prediction pipeline, establish monitoring procedures

12. Evaluation Metrics:

• Precision: Minimize false alarms (unnecessary interventions)

• Recall: Capture as many delinquent cases as possible

• F1-Score: Balance between precision and recall

• AUC-ROC: Overall model discriminative ability

13. Implementation Considerations:

• Handle class imbalance with SMOTE or class weights

• Regular model retraining with new data

• Monitor for feature drift and model performance degradation

Step 2:

Justification for Logistic Regression in Delinquency Prediction

1. Why Logistic Regression is Appropriate:

• Binary Classification Fit: Delinquency prediction is fundamentally a binary problem (delinquent/non-delinquent), which is the sweet spot for logistic regression

• Probabilistic Output: Provides probability scores rather than just classifications, enabling risk ranking of customers for prioritized interventions

• Linear Relationship Assumption: Financial indicators like credit score, debt-to-income ratio, and missed payments have approximately linear relationships with delinquency risk

2. Strengths for Geldium's Needs:

• High Interpretability: Each feature's coefficient directly indicates its impact on delinquency risk, crucial for explaining decisions to customers and regulators

• Regulatory Compliance: Meets regulatory requirements for explainable AI in financial services (e.g., GDPR, CCPA)

• Transparency: Easy to document and validate model behavior for internal audits and external examinations

• Computational Efficiency: Fast training and prediction times enable real-time risk scoring during customer interactions

• Stability: Less prone to overfitting with proper regularization, ensuring consistent performance on new data

3. Trade-offs and Considerations:

• Limited Non-linear Modeling: May miss complex interactions between features that tree-based models would capture

• Feature Engineering Dependency: Requires thoughtful feature engineering to capture non-linear relationships

• Assumption of Linearity: Assumes log-odds of delinquency change linearly with features, which may not always hold

4. Business Alignment:

• Risk Management: Probability outputs enable risk-based pricing and credit limit decisions

• Customer Experience: Interpretability allows customer service representatives to explain credit decisions

• Operational Efficiency: Simple model reduces IT complexity and maintenance costs

• Audit Trail: Clear coefficient values create an auditable trail of model decision factors

5. Deployment Advantages:

• Minimal Infrastructure: Doesn't require specialized hardware or complex deployment environments

• Easy Monitoring: Simple metrics and diagnostics facilitate ongoing performance monitoring

• Quick Updates: Model can be retrained quickly with new data as business conditions change

• Integration Friendly: Straightforward to integrate into existing credit decision systems

Step 3:

Model Evaluation Strategy for Delinquency Prediction

1. Accuracy Metrics:
2. • AUC-ROC: Overall discriminative ability (Target: >0.75)
3. • Precision: Proportion of predicted delinquents that are actually delinquent (Target: >0.60)
4. • Recall/Sensitivity: Proportion of actual delinquents correctly identified (Target: >0.70)
5. • F1-Score: Harmonic mean of precision and recall (Target: >0.65)
6. • Specificity: Proportion of non-delinquents correctly identified (Target: >0.70)
7. • Accuracy: Overall correct predictions (Target: >0.75)
8. • Brier Score: Calibration of probability estimates (Target: <0.25)
9. Fairness and Bias Metrics:

• Equal Opportunity Difference: (TPR_minority - TPR_majority) - Measures if true positive rates differ across groups
• Demographic Parity Difference: (PPV_minority - PPV_majority) - Measures if positive prediction rates differ
• Equalized Odds Difference: Average of differences in FPR and TPR across groups
• Calibration Gap: Difference in predicted vs actual delinquency rates across groups
• Individual Fairness: Similar individuals should receive similar risk scores

10. Group-Based Evaluation:

• Evaluate metrics across protected attributes: Age groups, Employment_Status, Location
• Check for disparate impact: No group should have significantly different outcomes
• Analyze performance across Credit_Card_Type and Account_Tenure segments

11. Interpretation Guidelines:

• AUC < 0.70: Model needs significant improvement
• AUC 0.70-0.75: Model is acceptable but could be improved
• AUC > 0.75: Model has good discriminative ability
• Precision < 0.50: Too many false alarms, adjust threshold
• Recall < 0.60: Missing too many actual delinquents, adjust threshold
• F1-Score < 0.60: Imbalance between precision and recall
• Fairness metrics > 0.10: Potential bias requiring investigation

12. Bias Detection Methods:

• Feature Importance Analysis: Check if protected attributes have undue influence
• Residual Analysis: Identify systematic errors across demographic groups
• Threshold Optimization: Ensure optimal thresholds for different groups
• Adversarial Testing: Test model with edge cases and boundary conditions

13. Monitoring and Improvement Triggers:

• Performance degradation: >10% drop in key metrics over 3 months
• Data drift: >20% change in feature distributions

- Bias emergence: >0.15 increase in fairness metric disparities
- Business impact: Significant increase in customer complaints or regulatory inquiries

14. Reporting Requirements:

- Monthly performance dashboards with all key metrics
- Quarterly fairness assessment reports
- Annual third-party model validation
- Ad-hoc analysis for significant business changes