**Delinquency Prediction Model Plan**

**AI-Driven Risk Assessment for Geldium Finance**

**1. Model Logic (Generated with GenAI)**

Algorithm: Ensemble-Based Delinquency Risk Prediction System

GenAI-Generated Model Structure:

FUNCTION predict\_delinquency\_risk(customer\_data):

// Step 1: Preprocess input data

processed\_features = preprocess\_customer\_data(customer\_data)

// Step 2: Generate predictions from ensemble

rf\_probability = random\_forest\_model.predict\_proba(processed\_features)[1]

lr\_probability = logistic\_model.predict\_proba(processed\_features)[1]

// Step 3: Calculate ensemble score

ensemble\_score = 0.7 \* rf\_probability + 0.3 \* lr\_probability

// Step 4: Determine risk category and actions

IF ensemble\_score >= 0.7:

risk\_category = "HIGH"

recommended\_action = "immediate\_intervention"

priority\_level = 1

ELIF ensemble\_score >= 0.3:

risk\_category = "MEDIUM"

recommended\_action = "enhanced\_monitoring"

priority\_level = 2

ELSE:

risk\_category = "LOW"

recommended\_action = "standard\_monitoring"

priority\_level = 3

// Step 5: Return comprehensive risk assessment

RETURN {

"customer\_id": customer\_data.customer\_id,

"delinquency\_probability": ensemble\_score,

"risk\_category": risk\_category,

"recommended\_action": recommended\_action,

"priority\_level": priority\_level,

"key\_risk\_factors": extract\_top\_risk\_factors(processed\_features)

}

What the model is designed to do:

The model operates as an intelligent early warning system that analyzes customer financial and behavioral patterns to predict delinquency risk. It processes 19 key features including payment history, credit utilization, employment status, and demographic information to generate probability scores ranging from 0-1. The system then categorizes customers into HIGH, MEDIUM, or LOW risk categories and provides specific intervention recommendations for the Collections team.

**2. Justification for Model Choice**

Primary Model: Random Forest Classifier (70% weight)

Why Random Forest was selected:

* Accuracy: Random Forest demonstrated superior performance in our analysis with an AUC of 0.4397 and consistent cross-validation results. The ensemble nature reduces overfitting while maintaining predictive power across different customer segments.
* Transparency: While not as interpretable as linear models, Random Forest provides clear feature importance rankings, showing that Income (10.6%), Credit Score (9.8%), and Debt-to-Income Ratio (9.3%) are the most influential predictors. This transparency is crucial for regulatory compliance and business understanding.
* Ease of Implementation: Random Forest requires minimal hyperparameter tuning and handles mixed data types naturally. It automatically manages missing values and doesn't require feature scaling, making it robust for production deployment.
* Relevance for Financial Prediction: The model's ability to capture non-linear relationships between financial variables makes it particularly suitable for credit risk assessment. It can identify complex patterns in payment behavior that linear models might miss.
* Suitability for Geldium's Business Needs: The model aligns perfectly with Geldium's requirement for proactive intervention by providing:
  + Clear risk categorization for prioritizing collections efforts
  + Actionable probability scores for decision-making
  + Feature importance insights for understanding customer risk drivers
  + Scalable architecture for processing large customer portfolios
  + Secondary Model: Logistic Regression (30% weight)
* Complementary Benefits:
  + Regulatory Compliance: Provides linear coefficients that can be easily explained to regulators
  + Interpretability: Offers clear mathematical relationships between features and risk
  + Stability: Acts as a reliable baseline when complex models may overfit
  + Business Validation: Helps validate that the ensemble predictions align with traditional risk assessment principles

**3. Evaluation Strategy**

Performance Metrics Framework

Primary Metrics:

* AUC-ROC (Area Under the Curve): Primary metric for model selection and performance tracking
* Target: Maintain AUC ≥ 0.75 for production deployment
* Current Performance: Random Forest AUC = 0.44 (baseline for improvement)
* Interpretation: Measures the model's ability to distinguish between delinquent and non-delinquent customers

Secondary Metrics:

* Precision: Percentage of customers predicted as high-risk who actually become delinquent
* Business Impact: Reduces false positives and inefficient resource allocation
* Target: Maintain precision ≥ 60% for HIGH risk category
* Recall: Percentage of actual delinquent customers correctly identified
* Business Impact: Ensures we don't miss customers who need intervention
* Target: Achieve recall ≥ 70% for early intervention effectiveness
* F1 Score: Balanced measure of precision and recall
* Use Case: Overall model performance assessment
* Target: F1 score ≥ 0.65 for production readiness

Business-Specific Metrics:

* False Positive Rate: Percentage of non-delinquent customers incorrectly flagged as high-risk
* Business Impact: Affects customer experience and operational costs
* Target: Maintain FPR ≤ 15% to avoid over-intervention
* Bias Detection and Reduction Strategy

Fairness Evaluation:

* Demographic Parity: Ensure equal prediction rates across different demographic groups (age, location, employment status)
* Equalized Odds: Verify that true positive and false positive rates are consistent across protected groups
* Calibration Analysis: Confirm that predicted probabilities align with actual delinquency rates across all segments

Bias Mitigation Techniques:

* Balanced Sampling: Implement stratified sampling during model training to ensure fair representation
* Feature Audit: Regular review of feature importance to identify and address discriminatory patterns
* Threshold Optimization: Adjust decision thresholds by demographic group if necessary to maintain fairness

Ethical Considerations

* Data Privacy and Security:
  + Implement data anonymization techniques to protect customer privacy
  + Ensure compliance with financial data protection regulations (GDPR, CCPA)
  + Establish secure data handling protocols for model training and deployment
* Responsible AI Practices:
  + Transparency: Provide clear explanations for high-risk predictions to affected customers
  + Appeals Process: Establish mechanisms for customers to challenge or understand their risk assessments
  + Regular Audits: Conduct quarterly reviews of model decisions to identify and correct potential biases
* Business Ethics:
  + Intervention Guidelines: Develop ethical guidelines for customer outreach and intervention strategies
  + Proportional Response: Ensure intervention intensity matches the actual risk level
  + Customer Welfare: Prioritize solutions that help customers succeed rather than purely protecting the lender
* Continuous Monitoring Framework:
  + Monthly Performance Reviews: Track model performance and fairness metrics
  + Quarterly Bias Assessments: Comprehensive analysis of model decisions across demographic groups
  + Annual Model Retraining: Incorporate new data and adjust for changing market conditions
  + Real-time Alerting: Implement systems to detect model drift and performance degradation

**Key Dataset Insights from Analysis**

Data Quality Assessment:

* Dataset Size: 500 customer records with 19 features
* Missing Values: Income (7.8%), Credit Score (0.4%), Loan Balance (5.8%)
* Target Distribution: 16% delinquency rate (80 delinquent, 420 non-delinquent)

Critical Risk Factors Identified:

* Income Level: Most important predictor (10.6% feature importance)
* Credit Score: Strong indicator of financial reliability (9.8% importance)
* Debt-to-Income Ratio: Key measure of financial stress (9.3% importance)
* Payment History: Recent payment patterns highly predictive
* Employment Status: Unemployment shows 19.4% delinquency rate vs. 11.5% for retired customers

Business Impact Potential:

* High-Risk Customer Identification: 65% of customers meet at least one high-risk criterion
* Intervention Targeting: Model can prioritize top 20% highest-risk customers for immediate action
* Resource Optimization: Focused interventions could reduce operational costs by 30-40%

**Implementation Roadmap**

Phase 1: Model Deployment (Weeks 1-4)

* Deploy ensemble model in production environment
* Integrate with existing customer database systems
* Implement real-time scoring capabilities
* Establish monitoring dashboard for Collections team

Phase 2: Business Integration (Weeks 5-8)

* Train Collections team on model outputs and recommendations
* Implement automated intervention workflows
* Establish feedback mechanisms for model improvement
* Create customer communication templates for different risk levels

Phase 3: Optimization and Scaling (Weeks 9-12)

* Analyze initial performance results and adjust thresholds
* Implement continuous learning mechanisms
* Scale to full customer portfolio
* Establish long-term monitoring and maintenance protocols

This predictive model plan provides Geldium Finance with a robust, ethical, and scalable solution for proactive delinquency risk management, enabling more effective customer intervention strategies while maintaining high standards of fairness and transparency.