**GELDIUM DELINQUENCY RISK PREDICTION – STRUCTURED MODEL PLAN**

**✅ Step 1: Predictive Model Logic**

**Chosen Model Type:**

We recommend using **Logistic Regression** as a baseline model due to its simplicity, transparency, and suitability for financial applications. For improved performance and the ability to capture complex patterns, we also propose a **Random Forest Classifier**.

**Top 5 Input Features:**

1. **Credit\_Score** – Strong inverse relationship with delinquency; a key financial indicator.
2. **Credit\_Utilization** – Higher utilization often reflects financial stress.
3. **Missed\_Payments** – Historical payment behavior is a direct predictor of delinquency.
4. **Income** – Lower income levels are linked to increased risk of non-payment.
5. **Debt\_to\_Income\_Ratio** – Measures the financial burden of debt relative to income.

**Modeling Workflow:**

1. **Data Ingestion**: Load the cleaned dataset.
2. **Preprocessing**:
   * Impute missing values using median values.
   * One-hot encode categorical features (Employment\_Status, Credit\_Card\_Type, Location).
   * Scale numeric features.
3. **Model Training**:
   * Train Logistic Regression for interpretability.
   * Train Random Forest for higher accuracy.
4. **Prediction Output**:
   * Predict delinquency probability.
   * Apply threshold (e.g., 0.5) for binary classification.

**✅ Step 2: Justification for Model Choice**

Logistic Regression is widely used in credit risk modeling due to its transparency and explainability, which are critical for regulatory compliance and customer communication. It provides clear coefficients that link customer behavior to risk. However, to capture non-linear relationships and interactions between features, **Random Forest** is proposed as a more robust alternative. It provides better predictive performance and includes built-in feature importance scoring, helping the team identify top risk drivers. This dual-model approach aligns with Geldium's need for both operational efficiency and regulatory accountability.

**✅ Step 3: Evaluation Strategy**

**Metrics to Assess Model Performance:**

| **Metric** | **Purpose** |
| --- | --- |
| **Accuracy** | General correctness of predictions |
| **Precision** | Avoid misclassifying non-risk customers as high-risk |
| **Recall** | Ensure real high-risk customers are identified |
| **F1 Score** | Balances false positives and false negatives |
| **AUC-ROC** | Measures model's ability to distinguish between risk and non-risk |
| **Fairness Checks** | Detect bias across demographic groups (e.g., Location, Employment\_Status) |

**Bias & Fairness Auditing Plan:**

* Compare recall and precision across groups (e.g., urban vs rural, full-time vs unemployed).
* Conduct disparity analysis using tools like Fairlearn.
* If bias is detected, explore:
  + Reweighting training data.
  + Threshold adjustments.
  + Fairness-aware models.

**✅ Conclusion**

This predictive modeling plan balances **interpretability** and **performance**. It is designed to help Geldium transition from reactive strategies to **proactive risk management**. The dual-model approach (Logistic Regression + Random Forest) supports transparency, scalability, and compliance while enabling data-driven decisions to reduce delinquency rates and improve customer targeting.