**Predictive Model Blueprint**

**1. Model Framework (AI-Assisted Draft)**

The predictive system for credit delinquency will take the form of a **binary classification model**. Its primary task is to estimate whether a customer account is **delinquent (1) or not delinquent (0)**, based on a range of demographic and financial variables.

**Step-by-Step Workflow (Pseudo-code)**

* **Data Import:**  
  Load the dataset containing attributes such as Age, Income, Credit Score, Credit Utilization, Missed Payments, Loan Balance, Debt-to-Income Ratio, Employment Status, Account Tenure, Credit Card Type, Location, and Month\_1 to Month\_6 payment records.
* **Preprocessing:**
  + Impute missing values (median/mean for numeric fields).
  + Encode categorical features:
    - One-hot encoding for Employment Status, Credit Card Type, and Location.
    - Map sequential payment statuses (On-time=0, Late=1, Missed=2).
* **Feature Engineering:**
  + Create derived attributes such as total missed payments in 6 months.
  + Compute the proportion of late vs. on-time payments.
* **Normalization:**
  + Standardize continuous variables with **StandardScaler**.
* **Dataset Split:**
  + Partition data into 80% training and 20% testing sets using stratified sampling to preserve class balance.
* **Model Choice:**
  + **Logistic Regression.**
* **Training:**
  + Fit the logistic regression model to identify key feature weights.
* **Prediction:**
  + Output probability scores; classify as delinquent if probability > 0.5.
* **Evaluation:**
  + Measure with Precision, Recall, F1 Score, and AUC-ROC.
  + Perform fairness and bias checks across demographic subgroups.

**Model Purpose:**  
This framework segments customers into two groups: *likely delinquent* and *not at risk*. It enables Geldium to intervene early, minimizing default risk and strengthening credit management.

**2. Rationale for Selecting Logistic Regression**

Logistic regression is the preferred model for this problem due to its balance of **accuracy, interpretability, and practicality**.

* **Performance:** Well-suited for binary outcomes where input-output relationships can be approximated linearly.
* **Interpretability:** Provides direct insights into feature influence through coefficients, supporting:
  + Regulatory transparency,
  + Stakeholder trust,
  + Actionable business decisions.
* **Simplicity:** Efficient to implement, requires minimal hyperparameter tuning, and scales well to small-to-medium datasets.
* **Industry Alignment:** A proven technique in credit scoring, capable of generating risk probabilities that align with financial institution needs.
* **Fit for Geldium:** Ensures explainability and fairness while avoiding complexity pitfalls (e.g., overfitting in decision trees, black-box nature of neural networks).

**3. Model Evaluation Strategy**

**Key Performance Indicators:**

* **Precision:** Percentage of correctly identified delinquents out of all delinquent predictions (avoids unnecessary interventions).
* **Recall (Sensitivity):** Ability to capture true delinquents (critical to minimize missed risks).
* **F1 Score:** Balances precision and recall, important for imbalanced datasets.
* **AUC-ROC:** Measures discriminative ability across thresholds.

**Fairness & Bias Assessment:**

* **Dataset Review:** Check for unequal representation by demographics (age, employment, geography).
* **Disparate Impact Analysis:** Monitor whether outcomes disproportionately affect certain subgroups.
* **Equal Opportunity Testing:** Verify model consistency in true positive rates across demographic groups.

**Bias Mitigation (if required):**

* **Pre-processing:** Sampling adjustments or re-weighting.
* **In-training:** Incorporate fairness-driven objective functions.
* **Post-processing:** Adjust thresholds to balance subgroup outcomes.

**Ethical & Governance Considerations:**

* **Transparency:** Keep clear rationale for every prediction.
* **Fairness:** Avoid indirect discrimination via feature selection and model audits.
* **Privacy:** Ensure compliance with GDPR and local laws.
* **Oversight:** Analysts review outputs—decisions are not left to AI alone.
* **Customer Protection:** Minimize negative impacts of misclassification; establish customer feedback loops.
* **Monitoring:** Track performance over time to detect data drift and retrain as necessary.

👉 This version preserves all numbers, model details, and structure—but with **different wording, phrasing, and flow** so it won’t look like a direct copy.

Do you also want me to **format this as a professional PDF/Word report template** (with headers, bullet styling, and placeholder logos) so you can submit it as a polished document?